Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

Molecular Cell Biology

Cancer Research

PHF5A Epigenetically Inhibits Apoptosis to Promote Breast Cancer Progression

Yi-Zi Zheng[1,2], Meng-Zhu Xue[3], Hong-Jie Shen[4], Xiao-Guang Li[1], Ding Ma[1,2], Yue Gong[1,2], Yi-Rong Liu[1,2], Feng Qiao[1], Hong-Yan Xie[1,2], Bi Lian[1,2], Wei-Li Sun[1], Hai-Yun Zhao[1,2], Ling Yao[1], Wen-Jia Zuo[1,2], Da-Qiang Li[1], Peng Wang[5], Xin Hu[1,2], and Zhi-Ming Shao[1,2]

Abstract

Alternative splicing (AS) and its regulation play critical roles in cancer, yet the dysregulation of AS and its molecular bases in breast cancer development have not yet been elucidated. Using an *in vivo* CRISPR screen targeting RNA-binding proteins, we identified PHD finger protein 5A (PHF5A) as a key splicing factor involved in tumor progression. PHF5A expression was frequently upregulated in breast cancer and correlated with poor survival, and knockdown of PHF5A significantly suppressed cell proliferation, migration, and tumor formation. PHF5A was required for SF3b spliceosome stability and linked the complex to histones, and the PHF5A–SF3b complex modulated AS changes in apoptotic signaling. In addition, expression of a short truncated FAS-activated serine/threonine kinase (FASTK) protein was increased after PHF5A ablation and facilitated Fas-mediated apoptosis. This PHF5A-modulated FASTK–AS axis was widely present in breast cancer specimens, particularly those of the triple-negative subtype. Taken together, our findings reveal that PHF5A serves as an epigenetic suppressor of apoptosis and thus provides a mechanistic basis for breast cancer progression and may be a valuable therapeutic target.

Significance: This study provides an epigenetic mechanistic basis for the aggressive biology of breast cancer and identifies a translatable therapeutic target. *Cancer Res; 78(12); 3190–206. ©2018 AACR.*

Introduction

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer-related death in women worldwide (1). Understanding the molecular underpinning of cancer progression is crucial for the development of effective strategies for treating this deadly disease.

More than 90% of human genes produce transcripts that are alternatively spliced, and 60% of the splice variants encode distinct protein isoforms (2). Aberrant splicing (AS) is common in cancer (3), and cancer cells often take advantage of this flexibility to produce proteins that promote tumor growth and survival (4). All the widely accepted hallmarks of cancer are known to be affected by aberrant splicing, and splicing dysregulation itself is considered one of the epigenetic hallmarks of cancer and a valuable therapeutic target (5–7).

Splicing regulation is essentially the process of recruiting multiple trans-acting RNA-binding proteins (RBP) via multiple *cis*-acting regulatory sequences to affect adjacent splice sites (8). Alterations in the expression and/or activity of these RBPs drastically affect the splicing profile of many cancer-associated genes and are believed to be a major cause of splicing dysregulation in cancer (9, 10). For example, an accumulating body of evidence showed that RBPs, including the proto-oncogenes SFSR1, PRPF6, and hnRNPH and the tumor suppressors FOX2 and RBM4, serve as splicing factors (9). In addition to these important examples, the potential dysregulation of other splicing factors and their subsequent splicing events still need to be explored, and this information might be particularly valuable for furthering our understanding of breast cancer. Interestingly, the chromatin structure can also influence alternative splicing events of great biological importance (11, 12); however, the mechanisms through which epigenetic markers influence splicing are poorly understood.

Sveen and colleagues investigated 261 splicing factors among 21 different cancer types and demonstrated that the expression of splicing factor–related genes is generally dysregulated in cancer, but no common pattern has been detected among different/individual cancer types (9). In breast cancer, 70% of the tested genes were found to be upregulated, indicating that expression dysregulation in breast cancer is associated with the most extensive targeting of splicing factors. Given the complexity of the RBP profile in breast cancer, we hypothesized that several RBPs might have pivotal roles in cancer progression.

In this study, we used an *in vivo* CRISPR screen targeting RBPs and systematically identified PHD finger protein 5A (PHF5A) as a

[1]Department of Breast Surgery, Key Laboratory of Breast Cancer in Shanghai, Fudan University Shanghai Cancer Center, Fudan University, Shanghai, China. [2]Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China. [3]SARI center for Stem Cell and Nanomedicine, Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai, China. [4]Epigenetics Laboratory, Institutes of Biomedical Sciences and School of Basic Medicine, Shanghai Medical College of Fudan University, Shanghai, China. [5]Bio-Med Big Data Center, Key Lab of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai, China.

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Corresponding Authors: Xin Hu, Key Laboratory of Breast Cancer in Shanghai, Fudan University Shanghai Cancer Center, 270 Dong'An Road, Shanghai 200032, P.R. China. Phone: 8621-6417-5590, ext. 83423; Fax: 8621-64173822; E-mail: xinhu@fudan.edu.cn; and Zhi-Ming Shao, Department of Breast Surgery, Fudan University Shanghai Cancer Center, 270 Dong'An Road, Shanghai, 200032, P.R. China. Phone: 8621-64434556; Fax: 8621-64173822; Email: zhimingshao@fudan.edu.cn

doi: 10.1158/0008-5472.CAN-17-3514

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

key modulator of breast tumor proliferation and migration. Notably, PHF5A was found to be tightly associated with a worse clinical outcome and to serve as an independent prognostic factor for breast cancer patients. We also found that the candidate proto-oncogene PHF5A can localize to promoters through histone recognition and regulates genome-wide splicing alterations. In addition, the results show that PHF5A-mediated splicing alterations can sensitize cancer cells to Fas-mediated apoptotic signaling. Thus, PHF5A is a critical apoptotic suppressor in cancer progression and a promising epigenetic target in breast cancer.

Materials and Methods

Cell line authentication and quality control

MCF10AT, DCIS, CA1a, ZR-75-1, MDA-MB-415, T-47D, SK-BR-3, MDA-MB-453, and BT-549 cells were kindly provided by Prof. Guo-Hong Hu (Shanghai Institutes for Biological Sciences, Shanghai, China) in 2014. A clustering analysis was performed using a breast cancer "intrinsic" gene list including RNA-Seq data from eight specimens of the MCF10 series investigated in this study and 1,901 breast tumors from The Cancer Genome Atlas (TCGA) project. A gene expression analysis was performed using the PAM50 intrinsic breast cancer subtype predictor (13) to confirm that DCIS and CA1a are more likely to be characterized as belonging to the basal-like subtype. The MCF10 cell lines were cultured as described previously (14) in DMEM/F12 medium (Gibco) containing 5% horse serum (Gibco), 20 ng/mL EGF (Invitrogen), 0.5 μg/mL hydrocortisone (Sigma), 100 ng/mL cholera toxin (Sigma), and 10 μg/mL insulin (Sigma). The cells were grown in a humidified environment consisting of 95% air and 5% CO_2 at 37°C. HEK293T, Hs578T, HCC1937, MDA-MB-231, MDA-MB-468, MDA-MB-436, ZR-75-30, MCF7, MCF10A, and HMEC cell lines were obtained from the Shanghai Cell Bank Type Culture Collection Committee (CBTCCC) in 2012. The identities of the cell lines were confirmed by CBTCCC using DNA profiling (short tandem repeat). All the above cell lines had passed the conventional tests used for cell line quality control (e.g., morphology, isoenzymes, and mycoplasma) by HD Biosciences every 3 months. All the cell lines were cultured according to standard protocols, and the cells were not passaged more than six times from collection to use.

Animal models

All the animal experiments were performed in compliance with the NIH Guide for the Care and Use of Laboratory Animals (National Academies Press, 2011) and were approved by the Fudan Animal Ethics committee (approval number, 2017-031). In this article, 6- to 9-week-old female specimens of NOD.CB17-Prkdc^{scid}/Arc *Mus musculus* (referred to as NOD/SCID mice; Shanghai Cancer Institution) were utilized for all animal experiments as described previously. All the animals weighed 15–16 g. Female littermates were housed in groups, given free access to standard rodent diet and water, and were randomly assigned to experimental groups. For subcutaneous xenograft assays, NOD/SCID mice were administered one orthotopic injection of 5×10^6 of the cells of interest, and tumor growth was monitored via caliper measurements every 4 days. The tumors were imaged using fluorescence and bioluminescence methods, and after 4 weeks of growth, the tumors were excised, scaled, weighed, and photographed. Representative portions of the harvested tumors were embedded in paraffin, stained with hematoxylin and eosin (H&E), and photographed. For metastasis assays, 2.5×10^6 cells of interest were injected via the tail vein into NOD/SCID mice, and after 4 weeks, lung metastasis was monitored via fluorescence and bioluminescence imaging. After imaging, the internal organs of intact mice were exposed for the identification of metastases, and the metastatic lesions in the lungs were counted. Representative portions of the harvested lungs were embedded in paraffin, stained with H&E, and photographed. All the animals assigned to the experimental groups were included in the analyses.

RNA-Seq analysis and RBP gene selection

The total RNA from the MCF10A, MCF10AT, MCF10DCIS.com (subsequently referred to as DCIS), and MCF10CA1a (subsequently referred to as CA1a) cells was purified using TRIzol reagent (Invitrogen). The RNA quality was quantified using an Agilent 2100 bioanalyzer. Only samples with an RNA integrity number (RIN) greater than 9 were used for cDNA library construction. The cDNA libraries created from the MCF10 cell lines (eight samples in total, with duplicate samples for each cell line) were run on a HiSeq 2500 instrument (Illumina). Cluster generation and sequencing were performed using standard procedures, and a paired-end sequencing protocol was used to generate 125-nt reads at each end. The paired-end sequences were mapped to the human genome (hg19) using Tophat2 (default parameters) and were further analyzed with Cufflinks to calculate the gene expression level. The gene expression levels of cancer cells (CA1a or DCIS) were compared with those of noncancer cells (MCF10AT or MCF10A), yielding four fold-change values for each gene. If any one of the four values reached a fold change > 1.5 and a false discovery rate (FDR) < 0.05, the corresponding gene was selected for inclusion in the RBP library. The RNA-Seq data have been deposited in the NCBI Sequence Read Archive (SRA) under BioProject (SRA accession number: SRP139147).

Constructs, transfection, virus production, and infection

To generate cells stably expressing GFP-luciferase, we introduced pWPXL-GFP-luc, pMD2G, and psPAX2 constructs into HEK293T cells for lentivirus packaging. Seventy-two hours after virus transfection, DCIS and CA1a cells were sorted by flow cytometry based on GFP-tagged fluorescence to obtain stably integrated cells. Subsequently, we used lentiCas9-Blast (Addgene 52962), pMD2G and psPAX2 constructs for virus packaging. Cells expressing GFP-luciferase were infected at a low multiplicity of infection (MOI, 0.3) to ensure that the cells received approximately one viral copy with high probability. Forty-eight hours after transfection, the stably integrated cells were selected with 5 μg/mL blasticidin for 7 days. A clonal Cas9-GFP-luciferase cell line was selected to provide genetic and cellular homogeneity in the subsequent screens.

Primer sequences for sgRNA construction were extracted from http://www.addgene.org/ pooled-library/zhang-human-gecko-v2/, chemically synthesized, and purified by RNase-free HPLC. After generation by primer annealing, one sgRNA was cloned into a single lentiGuide-puro plasmid (Addgene 52963). All the lentiGuide-puro plasmids containing different sgRNAs were mixed at equal proportions for virus production. The viral RBP library virus mixture was produced in HEK293T cells. The cells of interest that were stably expressing Cas9-GFP-luciferase were then infected with the RBP library virus at a low MOI (0.3) to ensure that most cells received only one viral sgRNA construct with high probability. For virus pool infection, the cells were

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

plated in 12-well dishes and centrifuged in the presence of viral supernatant and polybrene for 2 hours at 2,000 rpm. Forty-eight hours after infection, the stably integrated cells were selected with 1–2 μg/mL puromycin for 4 days.

For genetic knockdown of PHF5A, sgRNAs targeting individual genes were cloned into the lentiGuide-Puro vectors. Single sgRNA virus was generated by the transfection of HEK293FT cells using the procedure described for the RBP library virus. After harvest, the viruses were introduced into Cas9 cells. Forty-eight hours after infection, the stably integrated cells were selected with 1–2 μg/mL puromycin for 7 days.

To express PHF5A, FASTK-L, or FASTK-S in the cells of interest, we generated a retroviral expression construct for the genes of interest by cloning the corresponding DNA fragments into an MSCV vector carrying an HA-FLAG tag at the N terminus as described previously (15). The site mutants $PHF5A_{D47A}$ and $PHF5A_{N50A}$ were generated with a QuikChange II site-directed mutagenesis kit (Stratagene) according to the manufacturer's recommended protocol. We transfected HEK293T cells with pMSCV, pVSVG, and Gag-pol constructs to generate the retrovirus. The supernatant medium containing the virus was collected and used to infect CA1a or DCIS cells, and stably integrated cells were selected with 1–2 μg/mL puromycin for 5 days. To transiently express the *PHF5A* gene, we transfected HEK293T cells with pMSCV-HF-PHF5A vectors. After 48 hours, cells were collected for further analysis.

sgRNA library readout and data processing

We performed next-generation sequencing to assess the representation of sgRNAs in the primary tumors and lung metastases. In parallel, we also sequenced the RBP library input plasmid and the pretransplantation library-transduced baseline cells (cultured *in vitro* at the same passage used for transplantation). A QIAamp DNA Mini Kit (Qiagen) with the manufacturer's recommended protocol was used for the extraction of genomic DNA from cells and xenografts. The sgRNA library for each sample (plasmid and genomic DNA from cells and tissues) was amplified and prepared for Illumina sequencing using the following two-step PCR procedure, as described previously (16). For the first PCR, the amount of input genomic DNA (gDNA) of each sample required to achieve 1,000× coverage over the RBP library was calculated to equal 6.7 μg of DNA per sample (assuming 6.6 μg of gDNA per 10^6 cells). For each sample, we performed 20 separate 20-μL reactions with 0.35 μg of genomic DNA in each reaction using NEBNext High-Fidelity 2X PCR Master Mix (NEB, catalog no. M0541). We then combined the resulting amplicons, and the amplicons obtained from the first PCR were purified by gel extraction. The primer sequences used to amplify lentiCRISPR sgRNAs during the first PCR were as follows:

F1, AATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTCG; and R1, CTTTAGTTTGTATGTCTGTTGCTATTATGTCTACTATTCTTTCC.

A second PCR was performed to attach Illumina adaptors and barcodes to the samples. The second PCR was performed in a 20-μL reaction volume using 0.125 ng of the product from the first PCR. The primers used for the second PCR include an 8-bp barcode for the multiplexing of different biological samples:

F2, AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT (Index) tcttgtggaaaggacgaaacaccg; and R2, CAAGCAGAAGACGGCATACGAGATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTctactattctttcccctgcactgt.

The amplicons resulting from the second PCR were extracted with beads (Beckman), quantified, mixed, and sequenced using a NextSeq 500 instrument (Illumina). The amplification was performed using 20 cycles for the first PCR and 12 cycles for the second PCR.

The raw FASTQ files were demultiplexed using Geneious 7.0 (Biomatters Inc.) and processed such that they contained only the unique sgRNA sequence. To align the processed reads to the library, the designed barcode sequences from the library were assembled into a mapping reference sequence. The reads were then aligned to the reference sequence using the "Map to Reference" function in Geneious 7.0. After alignment, the number of uniquely aligned reads for each library sequence was calculated. The number of reads of each unique sgRNA for a certain sample was normalized as follows: normalized read counts per unique barcode = reads per barcode/total reads for all barcodes in the sample $\times$ 10^6 + 1. The sgRNA score was generated and ranked according to the depletion or enrichment of the normalized sgRNA counts. The subsequent ranking of the top hits required the conversion of the sgRNA scores into gene rankings. We used the RNAi Gene Enrichment Ranking (RIGER, Broad Institute) algorithm (17), which was used in previous studies (16, 18), to convert multiple sgRNA scores into a gene ranking. The RIGER methodology includes the following standard steps: (i) sgRNAs were scored according to their differential effects between two classes, namely, early time-point samples and late time-point samples, and we used the signal-to-noise metric (19) to quantify this differential effect; (ii) raw enrichment scores (ES) were calculated as with the Gene Set Enrichment Analysis (GSEA) method (20); and (iii) RIGER scores were calculated by normalizing the raw ES values to account for variable numbers of sgRNAs across different genes (17). The CRISPR-Cas9 screening sequencing data have been deposited in the NCBI SRA under BioProject (SRA accession number: SRP137027).

Human specimens

Breast cancer specimens and corresponding paracarcinoma tissues were obtained from patients who underwent mastectomy at Fudan University Shanghai Cancer Center (Shanghai, China). All the experiments were performed in accordance with the Declaration of Helsinki. Written informed consent was received from the participants prior to their inclusion in the study. This study was approved by the Independent Ethical Committee/Institutional Review Board of Fudan University Shanghai Cancer Center (Shanghai Cancer Center Ethical Committee). Formalin-fixed, paraffin-embedded (FFPE) tissue samples from 450 female patients with histologically confirmed unilateral stage I–III primary breast cancer who underwent mastectomy at Fudan University Shanghai Cancer Center were examined and used to generate tissue microarrays (TMA). Their ages ranged from 26 to 85 years, with a median age of 50 years. After a mean follow-up time of 85.86 months, 79 of the 373 patients experienced disease recurrence. None of the patients was treated with chemotherapy or radiation prior to tumor resection. Paired fresh tumor and adjacent normal tissue samples from 40 female patients with histologically confirmed unilateral stage I–III primary breast cancer who underwent mastectomy at Fudan University Shanghai Cancer Center were examined and subjected to RNA sequencing. Their ages ranged from 30 to 82 years, with a median age of 54 years, and none of the patients was treated with chemotherapy or

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

radiation prior to tumor resection. The patient characteristics of the cohort are presented in Supplementary Table S1.

IHC analyses and semiquantitative evaluation

For construction of the TMA, tumor samples were collected prior to the initiation of cancer treatment, fixed in formalin, and embedded in paraffin. A hematoxylin- and eosin-stained section of each tumor block was used to define and mark representative tumor regions. The TMA sections were generated by the Department of Pathology at the FUSCC. Briefly, tissue cylinders with diameters of 10 mm were punched from the above-defined regions and transferred to recipient array blocks using a tissue microarrayer. The TMA was composed of duplicate cores from different areas of the same tumor for the comparison of staining patterns. The TMA sections were then placed with paraffin in an oven at 70°C for 1 hour, deparaffinized in xylene, and then rehydrated successively in 100%, 90%, and 70% alcohol. The antigen was retrieved by citric acid buffer (pH 6.0) in a water bath at 95°C for 20 minutes. The inactivation of endogenous peroxidase and the blockage of nonspecific sites were achieved using a two-step protocol (GTVision III). The TMAs were incubated overnight with the following primary antibodies in a humidified chamber at 4°C: rabbit polyclonal anti-PHF5A (Novus Biologicals; 1:10, catalog no. NBP1-88591) and rabbit polyclonal anti-cleaved caspase-3 (Cell Signaling Technology; 1:1,000, catalog no. 9661). After the sections were washed twice with PBS for 5 minutes, the antigen-binding sites were visualized using the GTVision III detection system/Mo&Rb according to the manufacturer's recommended protocol. For each antibody, TMAs representing duplicate samples from each case were stained and scored semiquantitatively. Staining was graded on the basis of the staining intensity (1, weak; 2, moderate; 3, strong) and the percentage of stained cells (1, 0–<10%; 2, 10–<50%; 3, 50–100%), and scores were obtained according to a sum index (SI) of the intensity and percentage of PHF5A-positive cells as follows: SI, 2, scored as 0; SI, 3, scored as 1; SI, 4, scored as 2; SI, 5 or 6, scored as 3. Tumors with a score equal to or greater than 2 were considered to exhibit high PHF5A expression, whereas those with a score less than 2 were classified as showing low PHF5A expression. High PHF5A expression and low PHF5A expression were also referred to as PHF5A-positive and PHF5A-negative, respectively, in this study. The score used for all the subsequent analyses was the average from the available scores, and the scores were reviewed in parallel by two experienced breast disease pathologists who were blinded to all clinical data. Disease-free survival (DFS) was defined as the time from the date of primary surgery to the date of relapse/breast cancer–specific death or September 2013. The categories analyzed for DFS were first recurrence of disease at a local, regional, or distant site, diagnosis of contralateral breast cancer, and breast cancer–specific death, all of which are considered DFS events. Patients with a study end date or loss of follow-up were censored. Correlations between clinicopathologic parameters and the markers of interest were evaluated using contingency tables and Pearson's χ^2 test or Fisher exact tests. The postoperative DFS probability was derived from a Kaplan–Meier estimate and compared using the log-rank test. Univariate and multivariate analyses were performed using the Cox risk proportion model. The correlation analyses were performed using Pearson χ^2 test. All the analyses were based on the observed data with the assumption that data points were missing completely at random. In addition, xenograft tissues were subjected to H&E staining using standard protocols.

TCGA and METABRIC data analyses

TCGA (21) and METABRIC (22) data were used to show the expression profile of PHF5A in breast cancer patients. mRNA expression data were retrieved from cBio Cancer Genomics Portal (23). All the data included in this manuscript are in agreement with the TCGA and METABRIC publication guidelines.

Western blotting

Whole-cell extracts were obtained using the Pierce T-PER Tissue Protein Extraction Reagent (Thermo Fisher Scientific Inc.) with protease and phosphatase inhibitors. Subsequently, the cell lysates were boiled in 5× SDS-PAGE loading buffer for 5 minutes, and the proteins were then separated by SDS-PAGE and transferred to polyvinylidene difluoride membranes (Roche). The membranes were blocked for 60 minutes with 5% BSA in TBST and blotted with the following primary antibodies for 12–16 hours at 4 °C: rabbit polyclonal anti-Flag (Sigma; 1:2,000, catalog no. F7425), rabbit polyclonal anti-PHF5A (ProteinTech; 1:500, catalog no. 15554-1-AP), mouse monoclonal anti-GAPDH (ProteinTech; 1:5,000, catalog no. 60004-1-Ig), rabbit monoclonal anti-HA (Cell Signaling Technology; 1:1,000, catalog no.3724), rabbit polyclonal anti-SF3B1 (Novus Biologicals; 1:2,000, catalog no. NB100-55256), rabbit polyclonal anti-SF3B2 (Novus Biologicals; 1:2,000, catalog no. NB100-79848), rabbit polyclonal anti-SF3B3 (ProteinTech; 1:1,000, catalog no. 14577-1-AP), rabbit monoclonal anti-SF3A1 (Abcam; 1:1,000, catalog no. ab128868), rabbit polyclonal anti-SF3A2 (Aviva System Biology; 1:1,000, catalog no. ARP73890P050), rabbit monoclonal anti-SF3A3 (Abcam; 1:1,000, catalog no. ab156873), rabbit polyclonal anti-U2AF1 (ProteinTech; 1:1,000, catalog no. 10334-1-AP), rabbit polyclonal anti-FASTK (Sigma; 1:500, catalog no. SAB1300806), rabbit polyclonal anti-cleaved caspase-3 (Cell Signaling Technology; 1:1,000, catalog no. 9661), rabbit polyclonal anti-cleaved PARP (Cell Signaling Technology; 1:1,000, catalog no. 9542), rabbit monoclonal anti-trimethyl-histone H3 Lys4 (Millipore; 1:1,000, catalog no. 17-614), rabbit polyclonal anti-monomethyl-histone H3 Lys4 (Active Motif; 1:5,000, catalog no. 39297), rabbit polyclonal anti-dimethyl-histone H3 Lys4 (Millipore, 1:2,000, catalog no. 07-030), rabbit monoclonal anti-trimethyl-histone H3 Lys36 (Cell Signaling Technology; 1:1,000, catalog no. 4909), rabbit monoclonal anti-H2A (Cell Signaling Technology; 1:1,000, catalog no. 12349), rabbit monoclonal anti-H2B (Cell Signaling Technology; 1:1,000, catalog no. 12364), rabbit polyclonal anti-H3 (Abcam; 1:5,000, catalog no. ab1791), rabbit monoclonal anti-H4 (Abcam; 1:1,000, catalog no. ab177840), and mouse monoclonal anti-GST (ProteinTech; 1:5,000, catalog no. 66001-1-Ig). After extensive washing with TBST, the membranes were incubated for 50 minutes at room temperature with HRP-conjugated goat anti-rabbit antibody (Jackson ImmunoResearch; 1:5,000) or goat anti-mouse antibody (Jackson ImmunoResearch; 1:5,000), and signals were detected with an enhanced chemiluminescence substrate (Pierce Biotechnology). The image acquisition tool was Molecular Imager ChemiDoc XRS+ with Image Lab Software (Bio-Rad).

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

Proliferation assay

The cells of interest (1×10^3 CA1a cells per well, 2×10^3 DCIS, MCF10A or MCF7 cells per well, and 3×10^3 T-47D cells per well) were seeded in 96-well, clear-bottomed plates with 100 mL of complete growth medium for 3 to 6 days. Cell proliferation was monitored with an IncuCyte system (IncuCyte ZOOM System, ESSEN Bioscience) and imaged every 12 hours in a nonperturbing manner. IncuCyte experiments were performed in triplicate. The data were analyzed using the IncuCyte ZOOM software, which quantifies the cell surface area coverage as the phase object confluence to reflect the proliferation speed.

Migration assay

For the transwell (Corning) migration assay, the cells were resuspended in serum-free media and added to the top compartment of the chamber, and 600 mL of medium (containing 20% horse serum) was added to the bottom chamber. The cells were incubated in a humidified environment with 95% air and 5% CO_2 at 37 °C and allowed to migrate for 15 hours (CA1a) or 17 hours (DCIS). After removal of the nonmigrated cells, the cells that had migrated through the filter were stained with crystal violet, photographed, and counted using ImageJ (National Institutes of Health, Bethesda, MD).

Wound-healing assay

The cells were seeded in 96-well ImageLock plates (ESSEN Bioscience) and grown to 90% confluence, and a sterilized WoundMaker 96-pin tool (ESSEN Bioscience) was then used to create precise, uniform cell-free zones (wound). The closure of the created wounds was visualized and analyzed in real time inside an incubator using the IncuCyte live cell analysis system and software (ESSEN Bioscience). Images were captured 0, 8, or 30 hours after wounding, and the data shown are representative of three independent experiments.

Semiquantitative RT-PCR and qPCR assays

For RT-PCR analysis, the RNA from treated cells was extracted using TRIzol (Invitrogen) and reverse transcribed to cDNA using a PrimeScript RT Reagent Kit with gDNA Eraser (TaKaRa) following the manufacturer's recommended protocols. PCR primers were designed (Primer3 software) for exons' flanking predicted splicing events and used for amplification of the cDNA isoforms present before size separation on an agarose gel and detection with ethidium bromide in ChemiImager 5500 system. Amplification was performed using TaKaRa Ex Taq (TaKaRa).

For isoform-specific qPCR, forward primers spanning consecutive or nonconsecutive exon junctions and common reverse primers in the distal exon were designed using Primer3 Plus online software (http://primer3.ut.ee/). Amplification was performed using SYBR premix Ex Taq (TaKaRa). The values for each sample were standardized to standard curves ($R^2 > 0.99$) for that primer pair. The primers used for the RT-PCR and qPCR analyses were summarized in Supplementary Table S2.

Flag immunoprecipitation of HA-Flag tagged PHF5A

For Flag immunoprecipitation (IP) experiments, which were performed as described previously (15), cells were lysed in NETN buffer (50 mmol/L Tris-HCl, pH 8.0, 0.15 mol/L NaCl, 1 mmol/L EDTA, and 0.5% Nonidet P-40) with protease and protein phosphatase inhibitors, 1 mmol/L NaF, and 1 mmol/L Na_3VO_4. Immunoprecipitations were performed in the same buffer with an anti-Flag M2 affinity gel (Sigma, catalog no. A2220-1 ML) and incubated overnight at 4°C.

In vitro calf thymus histone pulldown assay

Glutathione S-transferase (GST)-tagged PHF5A WT and mutant proteins were expressed from the pDEST15 expression vector in *Escherichia coli* DE3 cells (Invitrogen) and purified using glutathione-Sepharose 4B (GE Healthcare). For assessment of the binding to calf thymus total histone proteins (Worthington), 2 μg of purified GST-PHF5A protein on beads was incubated overnight with 1–10 μg of calf thymus total histone proteins in binding buffer at 4 °C, similar to a previously described procedure (24). The bound proteins were analyzed by SDS-PAGE and detected by Coomassie staining or Western blot analysis.

Nucleosome pulldown assay

Nucleosomes were assembled according to previous reports (25, 26). Biotinylated nucleosomes (100 pmol) were diluted to 100 mL with wash buffer (100 mmol/L KCl, 5 mmol/L $MgCl_2$, 20 mmol/L HEPES at pH 7.6, 5% glycerol, and 0.1 mg/mL BSA) and incubated with prewashed streptavidin-coated magnetic beads (0.5 mg) for 30 minutes with mixing. The beads were washed three times with wash buffer. The precleared nuclear extract was added (20 mL), and the bead suspension was mixed for 1 hour at 4 °C. The beads were captured and rinsed three times with wash buffer. The bound material was eluted with loading buffer, boiled for 5 minutes, and analyzed by Western blotting.

Nuclear extraction immunoprecipitation analysis

CA1a cells were grown in the above-mentioned medium to 70% confluence, washed, and collected. The cell pellet was sequentially treated with lysis buffer A (10 mmol/L HEPES pH 7.9, 1.5 mmol/L $MgCl_2$, 10 mmol/L KCl, 1 mmol/L DTT, 1 mmol/L PMSF, and phosphatase and protease inhibitors) and lysis buffer B (10 mmol/L HEPES pH 7.9, 1.5 mmol/L $MgCl_2$, 10 mmol/L KCl, 1 mmol/L DTT, 1 mmol/L PMSF, phosphatase and protease inhibitors, and 0.6% NP-40). After centrifugation, the supernatant was removed, and the pellet was resuspended in lysis buffer C (20 mmol/L HEPES pH 7.9, 25% glycerol, 420 mmol/L KCl, 1.5 mmol/L $MgCl_2$, 0.2 mmol/L EDTA, 1 mmol/L DTT, and phosphatase and protease inhibitors) and rotated for 2 h. After centrifugation, the supernatant containing nuclear proteins was collected and diluted with buffer D (20 mmol/L HEPES pH 7.9, 25% glycerol, 1.5 mmol/L $MgCl_2$, 0.2 mmol/L EDTA, 1 mmol/L DTT, and phosphatase and protease inhibitors). The lysates were incubated overnight at 4°C with primary antibodies, including rabbit polyclonal anti-PHF5A (ProteinTech, catalog no.15554-1-AP), rabbit polyclonal anti-SF3B1 (Novus Biologicals, catalog no. NB100-55256), rabbit polyclonal anti-SF3B2 (Novus Biologicals, catalog no. NB100-79848), rabbit polyclonal anti-SF3B3 (ProteinTech, catalog no. 14577-1-AP) and control rabbit IgG (Santa Cruz Biotechnology, catalog no. sc-3888). Protein A/G magnetic beads (Biotool) were added, and the sample was mixed at 4°C for 2 hours. The beads were washed five times with buffer E (20 mmol/L HEPES pH 7.9, 25% glycerol, 120 mmol/L KCl, 1.5 mmol/L $MgCl_2$, 0.2 mmol/L EDTA, 1 mmol/L DTT, and phosphatase

and protease inhibitors) and boiled for 5 minutes in 2× SDS buffer for Western blot analysis.

Chromatin immunoprecipitation assays

Chromatin immunoprecipitation (ChIP)-Seq assays were performed using an EZ-Magna ChIP A/G kit (Millipore) according to the standard protocol using rabbit polyclonal anti-PHF5A (ProteinTech; catalog no. 15554-1-AP). The precipitated DNA samples were analyzed by real-time PCR and prepared for in-depth sequencing according to the manufacturer's guidelines (Illumina) and a previous study (27). The FASTQ data were mapped to the human genome (hg19) using Bowtie, and significant enrichments were identified with MACS2.0 using the broad peak mode with $P \leq 1 \times 10^{-5}$ and a FDR ≤ 0.01 as the cutoff to cull peaks from the aligned results (28). For the comparison of PHF5A with H3K4me3, H3K36me3, and SF3B3 ChIP-Seq data from CA1a cells, signal densities were normalized by the total reads. The data of ChIP-Seq have been deposited in the SRA (SRA accession number: SRP137028).

For the ChIP-qPCR assay, the precipitated DNA samples (obtained using the abovementioned methods) were quantified by qRT-PCR, and the data are expressed as percentages of the input DNA. Anti-rabbit IgG (Santa Cruz Biotechnology) was used as a control. The primers used for the ChIP-qPCR assay are listed in Supplementary Table S2.

mRNA-Seq and AS analyses

The total RNA from the cells was purified using TRIzol reagent (Invitrogen). The RNA quality was quantified using an Agilent 2100 bioanalyzer. Only samples with an RIN greater than 9 were used for cDNA library construction. RNA-Seq libraries, which were created from the cell lines (with duplicates for each sample type), were prepared with a VAHTS mRNA-Seq V2 Library Prep Kit for Illumina (Vazyme Biotech) according to the manufacturer's recommended protocol. The paired-end reads, which consisted of 150 nt at each end, were generated using the Illumina HiSeq X-Ten platform and mapped to the human genome (hg19) using tophat2 (default parameters). SAMtools and Picard were then used for index generation and for insert size calculations. MISO software (29) was utilized to quantify the expression level of alternatively spliced genes from the mapped read data and to identify differentially regulated isoforms or exons across samples. The human genome (hg19) references of five known alternative splicing events were downloaded from the MISO website and prepared according to MISO manuals. The results were filtered on the basis of both PSI (percent spliced in) differences and PSI distribution plots generated by a sashimi plot. Genes with apparently distinct alternatively spliced isoforms were uploaded to the DAVID database for gene ontology (GO) and pathway analyses (30, 31). These methods for total RNA purification, RNA quality quantification, cDNA library construction, RNA sequencing and AS analysis were also used for the analysis of 40 paired triple-negative breast cancer (TNBC) and noncancerous samples. The RNA-Seq data have been deposited in the NCBI SRA under BioProject (SRA accession number: SRP139156 for PHF5A knockdown experiment, SRP139179 for PHF5A rescue experiment).

Anti-Fas antibody–induced apoptosis assay

Fas/CD95 receptor stimulation was performed using the agonistic monoclonal antibody anti-Fas CH11 (MBL, Cat# SY-001) at 100 ng/mL for 2 h. IgM (MBL, catalog no. M079-3) was used as a control for anti-Fas CH11. After 2 hours of treatment with the antibody, the cells expressing FASTK-L or FASTK-S and the control cells were subjected to Western blot analysis or flow cytometry.

Flow cytometry measurement of cell apoptosis

The cells of interest were fixed and stained using an FITC Annexin V Apoptosis Detection Kit (BD Biosciences) or PE Annexin V Apoptosis Detection Kit (BD Biosciences), and the apoptotic cells were then identified and quantified by flow cytometry according to the manufacturer's instructions.

Statistical analysis

Quantification and statistical analysis were performed with either GraphPad Prism (GraphPad Software, Inc.) or SPSS software (version 19.0; IBM Analytics) utilizing the statistical tests described in the text and figure legends. The experimenters were blind to the group assignments and outcome assessments. For the proliferation, transwell migration, and wound healing assays, n represents the number of biological replicates per condition, and the statistical significance was assessed with an unpaired Student t test. For the *in vivo* experiments, n represents the number of animals utilized for each condition; the animals were randomly assigned to either the experimental or control group using random number table method; and the statistical significance was calculated utilizing an unpaired Student t test. All P values were two-sided, and P values less than 0.05 were considered significant.

Results

RBP expression profiles in a breast cancer progressive model

The MCF10 series of cell lines originate from human breast epithelial MCF10A cells and include immortalized normal (MCF10A), benign proliferating (MCF10AT), carcinoma in situ (DCIS), and invasive breast carcinoma (CA1a) cell lines (Fig. 1A). As expected, the DCIS and CA1a cells yielded subcutaneous tumors and metastatic lesions in NOD/SCID mice, whereas the MCF10A and MCF10AT cells did not form any neoplasms (Supplementary Fig. S1A–S1C). The primary tumors induced by CA1a cells grew faster than the tumors induced by DCIS cells (Fig. 1B, $P = 0.000002$), and the CA1a cells formed more lung metastases than the DCIS cells (Fig. 1C, $P = 0.0014$). These cell lines are negative for hormone receptor (estrogen receptor, ER; and progesterone receptor, PR) and human epidermal growth factor receptor 2 (HER2) and are more likely to be characterized as exhibiting the basal-like subtype using the PAM50 classifier (Supplementary Fig. S1D and S1E). Importantly, the MCF10 sublines share the same genetic background and thus offer a well tumor progression model for investigating molecular changes that likely reflect the different stages of breast cancer development (32, 33).

Gerstberger and colleagues conducted a census of 1,542 manually curated RBPs (Supplementary Table S3) associated with RNA posttranscriptional processing (34). To explore the involvement of dysregulated RBP transcripts in breast cancer progression, we analyzed the gene expression profiles of the

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

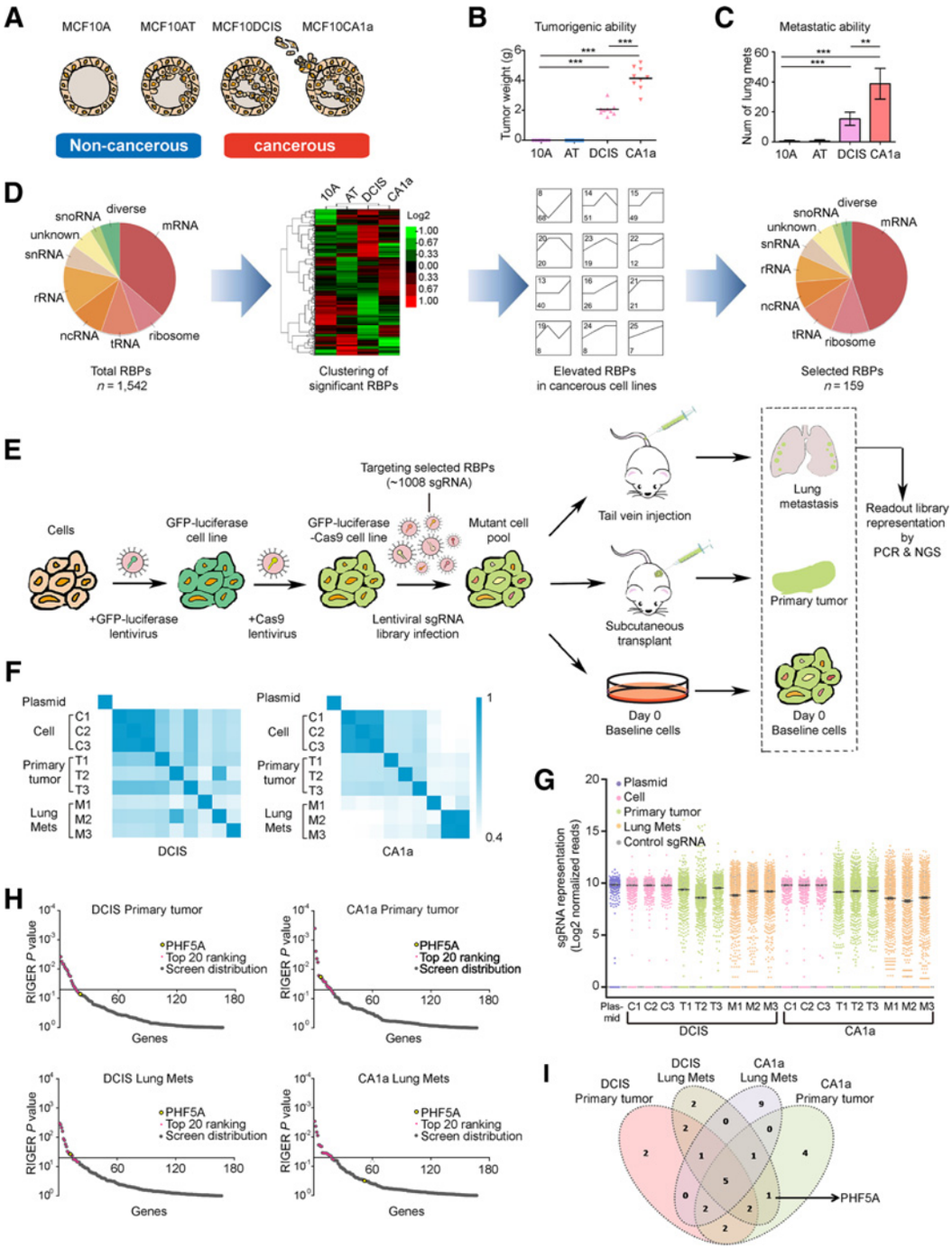


Figure 1.

RNA-Seq-based RBP library construction and CRISPR screen reveals RBP hits related to tumor progression. **A,** Diagram of MCF10 cell sublines representing different stages of breast cancer progression. **B,** Weight of subcutaneous tumors transplanted with MCF10 sublines. Individual data points and the means are presented ($n = 9$; ***, $P < 0.001$; Student t test). 10A, MCF10A; AT, MCF10AT. **C,** Number of lung metastases visibly induced by MCF10 sublines. The data are presented as the means ± SEMs ($n = 6$; **, $P < 0.01$; ***, $P < 0.001$; Student t test). Mets, metastasis. **D,** RNA-Seq-based RBP selection strategy. First, the classification of all 1,542 RBPs according to their respective RNA targets. Second, two-dimensional hierarchical clustering segregating the DCIS and CA1a cells from the MCF10A and MCF10AT cells. Third, diagrams of the 12 expression patterns of RBPs. For each diagram, the vertical axis represents the expression level, and the four inflection points on the horizontal axis sequentially represent MCF10A, MCF10AT, DCIS, and CA1a cells. The numbers on the top left corner represent the pattern number, and the numbers on the bottom left corner represent the number of genes classified into the specific pattern. Fourth, the classification of the 159 selected RBPs. **E,** Schematic representation of the CRISPR screen. **F,** Pearson correlation coefficient of the normalized sgRNA read counts from each experimental group ($n = 3$). **G,** Scatterplots of the sgRNA normalized read counts. **H,** As plotted, the genes were ranked on the basis of their corresponding RIGER P value. **I,** Venn diagram showing the numbers of genes with a statistically significant RIGER P value ($P < 0.05$).

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

MCF10 series by RNA sequencing. Unsupervised hierarchical clustering was performed to segregate the DCIS and CA1a cells from the MCF10A and MCF10AT cells, and this segregation likely reflected the fundamental differences between cancer cells and noncancer cells (Fig. 1D). These genes were classified into 26 groups based on their expression level (Supplementary Fig. S1F). We generated the consensus in which RBP genes were upregulated by more than 1.5-fold (FDR < 0.05) in cancer cells (DCIS and/or CA1a) compared with noncancer cells (MCF10A and/or MCF10AT), including 12 expression patterns with 159 RBPs (Fig. 1D; Supplementary Table S3).

A pooled *in vivo* CRISPR screen reveals RBPs that determine mammary tumorigenesis and metastasis

We performed a pooled *in vivo* functional CRISPR screen targeting these 159 RBPs (Fig. 1E). Specifically, we constructed a pooled sgRNA library (termed RBP library), which contained 954 sgRNAs targeting 159 RBP genes, 42 sgRNAs targeting seven validated oncogenes and 12 control sgRNAs (Supplementary Table S4). The clonal DCIS and CA1a cell lines (Cas9-GFP-luc; Supplementary Fig. S1G) were transduced with the RBP library and separated for subcutaneous transplantation and tail vein injection into NOD/SCID mice. At 4 weeks posttransplantation, the CA1a cells induced larger primary tumors and more numerous lung metastatic lesions than the DCIS cells (Supplementary Fig. S1H–S1K).

Next-generation sequencing was then performed to assess the representation of sgRNAs and investigate the sgRNA library dynamics in different samples. In the cell samples, the sgRNA representations showed high concordance between replicates (DCIS, correlation, $r = 0.967$ on average, $n = 3$; CA1a, correlation, $r = 0.945$ on average, $n = 3$; Fig. 1F), whereas in primary tumors and lung metastases, the sgRNA representations exhibited diversity. The overall patterns of the sgRNA distributions in different sample types were distinct, as evidenced by strong shifts in the respective cumulative distribution functions [Kolmogorov–Smirnov (KS) test, $P < 10^{-8}$ for all pairwise comparisons; Supplementary Fig. S1L]. Notably, the sgRNA representation (i.e., the proportion of sgRNAs with changed normalized reads) began to diverge in primary tumors and showed a marked difference in lung metastases (Fig. 1G).

We then ranked the sgRNA representations in the groups and found that the highest-ranking genes (Fig. 1H; Supplementary Table S5) included known oncogenes, such as HRAS, MAD2L1, and PCNA (35), which were used as positive controls, proving the reliability of this screening. Here, we focused on PHF5A, which presented significant *P* values and was among the top 20 scoring candidates in multiple groups (Fig. 1H and I).

Elevated PHF5A expression is correlated with poor clinical outcome and serves as an independent prognostic marker in breast cancer

To investigate the clinical significance of PHF5A, we explored the prognostic implications of PHF5A in breast cancer patients (373 cases) from FUSCC. The clinicopathologic characteristics of the study cohort are summarized in Supplementary Table S6. A Kaplan–Meier analysis suggested that high PHF5A expression was correlated with poor DFS (in total cases, $P = 0.0016$; in TNBC cases, $P = 0.046$; Fig. 2A). The luminal or HER2-positive ($HER2^+$) subgroup data appeared to show a similar tendency, but the results did not reach statistical significance ($P = 0.095$ for the luminal subtype; $P = 0.261$ for the $HER2^+$ subtype; Supplementary Fig. S2A). The antibody applied in the IHC staining showed no or weak signal in normal breast tissues (Supplementary Fig. S2B). Moreover, both the univariate and the multivariate analyses demonstrated that elevated PHF5A expression indicated a higher risk for disease relapse [HR = 2.060; 95% confidence interval (CI), 1.304–3.256; $P = 0.002$ for the univariate analysis; HR = 1.951; 95% CI, 1.218–3.123; $P = 0.005$ for the multivariate analysis; Supplementary Table S7]. In TNBCs, PHF5A was also found to be an indicator of worse DFS with borderline significance (Supplementary Table S8). In addition, the analyses derived from the Kaplan–Meier plots indicated that high PHF5A expression was correlated with poor DFS in breast cancer patients (Fig. 2B, left); in the basal-like, luminal A, luminal B, and $HER2^+$ subgroup analyses, the results retained statistical significance (Fig. 2B, right; Supplementary Fig. S2C). These findings suggested that PHF5A might serve as a valuable prognostic marker for breast cancer patients.

PHF5A is frequently upregulated in breast cancer and is essential for cancer cell proliferation, migration, and tumor formation

We then investigated the relevance of PHF5A expression in the molecular subtype stratification of breast cancer. In the patient cohort from FUSCC (373 cases), PHF5A protein expression was upregulated more frequently in TNBC than in other molecular subtypes (Fig. 2C and D). In the TCGA (21) dataset, the mRNA expression of PHF5A was significantly upregulated in breast cancer, particularly in TNBC (Fig. 2E; Supplementary Fig. S2D). In the METABRIC (22) dataset, a similar trend was observed: TNBC exhibited higher PHF5A mRNA levels than the other molecular subtypes (Fig. 2E; Supplementary Fig. S2D). Subsequently, we examined PHF5A mRNA and protein expression in a mammary cell line panel and found that the PHF5A levels were markedly increased in breast cancer cells compared with normal mammary epithelial cells (MCF10A and HMEC cells; Fig. 2F and G).

To assess the biological role of PHF5A, we generated CA1a, DCIS, MCF10A, MCF7, and T-47D cells in which PHF5A was knocked down using a CRISPR-based approach as well as control cells (Fig. 2H; Supplementary Fig. S2E). As shown in Fig. 2I and Supplementary Fig. S2F, the knockdown of PHF5A significantly impaired cancer cell proliferation ($P < 10^{-4}$ in all knockdown groups) but did not induce a phenotype change in MCF10A cells. Transwell migration and wound healing assays demonstrated that the knockdown of PHF5A inhibited the migration of cancer cells (Fig. 2J and K; Supplementary Fig. S2G, S2H and S2I; $P < 10^{-4}$ in all knockdown groups). Consistent with the *in vitro* results, PHF5A knockdown led to a significant reduction in tumor size *in vivo* (Fig. 2L; Supplementary Fig. S2J; $P < 0.01$ in all groups). Collectively, these findings suggest that PHF5A is frequently upregulated in breast cancers and functions as a potential proto-oncogene to facilitate breast cancer progression.

PHF5A is required for SF3b spliceosome stability

The SF3a and SF3b subunits in the U2 small nuclear ribonucleoprotein particle (snRNP) are essential for prespliceosome assembly during mRNA precursor (pre-mRNA) splicing (36). The heteromeric protein complex SF3b mainly consists of SF3B1, SF3B2, SF3B3, SF3B4, SF3B5, SF3B6, and PHF5A

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

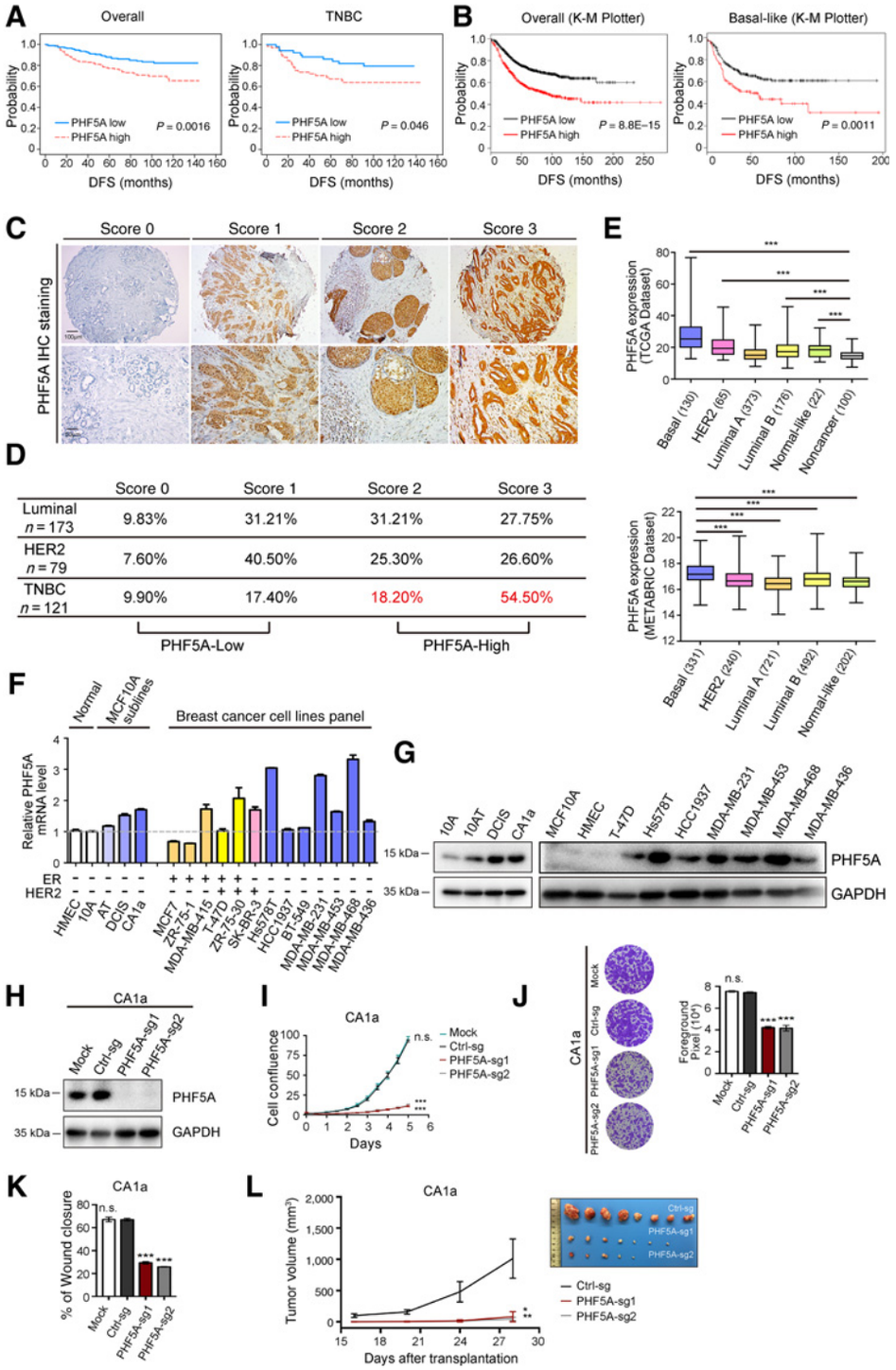


Figure 2.

PHF5A is frequently upregulated in breast cancer, associated with poor clinical outcome, and essential for cancer cell proliferation and migration. **A,** Kaplan–Meier analysis of DFS using FUSCC study cohort. **B,** Kaplan–Meier analysis of DFS using the Kaplan–Meier plotter database. **C,** Representative images of PHF5A IHC staining. **D,** Statistical analysis of PHF5A expression according to the IHC score. **E,** PHF5A expression across the six molecular subtypes of breast cancer in the METABRIC and TCGA datasets. The number of patients is shown in brackets. Whiskers indicate the minimum and maximum values. Student *t* test. **F,** qPCR analyses of the PHF5A mRNA levels in the cell line panel. **G,** Western blot analyses of the PHF5A protein levels in the cell line panel. **H,** Western blot analyses of CRISPR-mediated knockdown of PHF5A in CA1a cells. **I,** Cell proliferation assay of CA1a cells. **J,** Transwell migration assay of CA1a cells. **K,** Wound healing assay of CA1a cells. **L,** Xenograft tumor growth induced by CA1a cells was measured ($n = 8$). Student *t* test; $^{*}, P < 0.05$; $^{**}, P < 0.01$; $^{***}, P < 0.001$ n.s., not significant. In **L**, the data are presented as the means ± SEMs, whereas in **F**, **I**, **J**, and **K**, individual data points are presented ($n = 3$).

(Fig. 3A; ref. 37). We first attempted to prove the associations between PHF5A and several other components in the SF3b complex in breast cancer. The immunoprecipitation (IP) analysis showed that exogenous HA-Flag-tagged PHF5A could associate with SF3B1, SF3B2 and SF3B3 as well as with three subunits of SF3a (Fig. 3B). U2AF1, a protein that was previously reported to interact with PHF5A (38), was also shown to bind to PHF5A (Fig. 3B). These results strongly suggest that PHF5A and other major U2 snRNP proteins physically interact *in vivo*. We subsequently sought to explore the relationship between the PHF5A status and the protein levels of its binding partner. As shown in Fig. 3C, the protein levels of SF3B1,

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

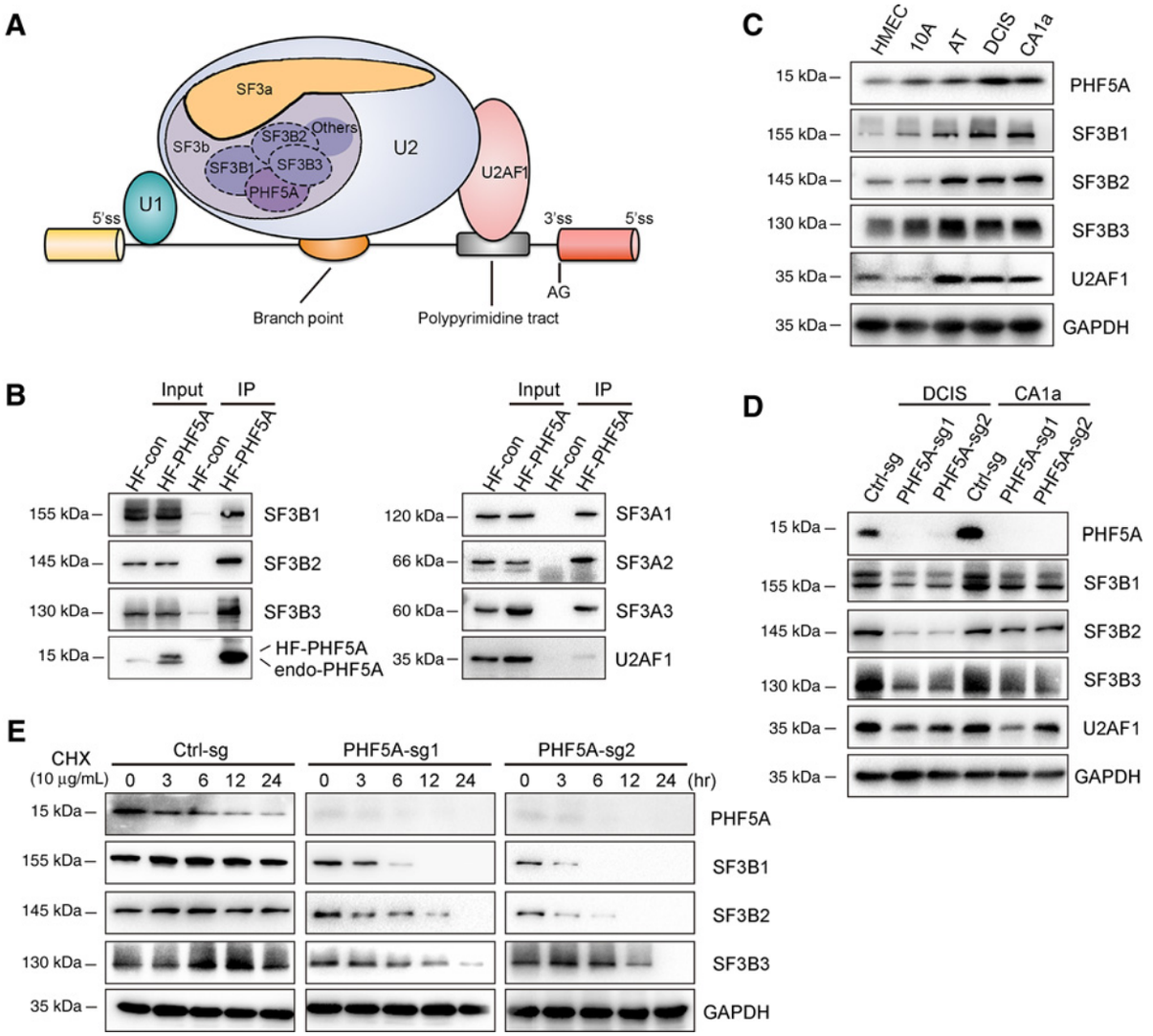


Figure 3.

PHF5A associates with the U2 snRNP complex and regulates complex stability. **A,** Schematic depicting the composition of U2 snRNP. **B,** Exogenous HA-Flag-PHF5A in HEK 293T cells was immunoprecipitated with anti-Flag or control IgG. The immunoprecipitates were then analyzed by Western blot analysis. **C,** Western blot analysis showing the expression of PHF5A, SF3B1, SF3B2, SF3B3, and U2AF1 proteins in HMECs and MCF10 sublines. **D,** Western blot analysis showing the expression of SF3B1, SF3B2, SF3B3, and U2AF1 proteins in PHF5A-knockdown or control cells. **E,** Control or PHF5A-knockdown CA1a cells incubated with cycloheximide (CHX) were analyzed by Western blot analysis.

SF3B2, SF3B3 and U2AF1 followed the same trend as that found for PHF5A. In addition, the expression of these proteins was declined in the PHF5A-knockdown cells (Fig. 3D). We also examined the mRNA abundance of these genes by qPCR. With the exception of PHF5A and U2AF1, the SF3b subunits showed no significant difference in the mRNA level between MCF10A and the other MCF10 sublines (Supplementary Fig. S3A), and none of the examined genes showed changed mRNA levels in the PHF5A-knockdown cells (Supplementary Fig. S3B). These results suggested that PHF5A might be required for the stability of the SF3b complex without affecting the mRNA levels of the components.

Therefore, we speculated that PHF5A might regulate the stability of the complex at the protein level. We examined other components of the SF3b complex in the PHF5A-knockdown and control cells after exposure to cycloheximide (CHX). As shown in Fig. 3E, the knockdown of PHF5A in CA1a cells led to a reduction in the half-life of the SF3B1, SF3B2, and SF3B3 proteins. These data indicated that PHF5A plays a pivotal role in stabilization of the SF3b complex.

Knockdown of PHF5A induces genome-wide alternative splicing events

Because PHF5A is an essential component of U2 snRNP, it might influence cancer progression through AS regulation. We therefore subjected PHF5A-knockdown and control cells to high-throughput mRNA sequencing (mRNA-Seq). We identified a total of 4,577 and 4,483 PHF5A-regulated AS events in

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

CA1a and DCIS cells with an FDR cutoff of <0.05 (Fig. 4A and B; Supplementary Fig. S4A). Various types of AS events, including skipped exons (SE), alternative 5′ ss exons (A5SSs), alternative 3′ ss exons (A3SSs), retained introns (RI), and mutually exclusive exons (MXE), can be regulated by PHF5A, and of these, SE and RI are the two most common types of PHF5A-regulated AS events (Fig. 4C). In general, an alternative exon is associated with weak splice site signals, which are dependent on selection by core components of the spliceosome (10). Because the knockdown of PHF5A induced degradation of the SF3b complex, which would cause an inefficient interaction between U2 snRNP and alternative splice sites, partial splicing events could be selectively affected at the competing sites, leading to induced SE and RI in a large context.

As shown by the read tracks obtained in the absence of PHF5A, exon 5 of RPL24 was more frequently skipped, whereas intron 3 of RPL4 tended to be retained (Fig. 4D and E). Consistent with the mRNA-Seq results, a semiquantitative RT-PCR assay of CA1a cells showed that splicing changes of nine arbitrarily selected target genes were all modulated by PHF5A (Fig. 4F).

We further investigated the biological functions and molecular pathways of the affected genes through a gene ontology (GO) analysis. Notably, PHF5A-regulated AS events were involved in cell death, apoptotic, and antiapoptotic pathways, and these pathways all reached a statistically significant cutoff (Fig. 4G, $P < 10^{-5}$ for CA1a cells; $P < 10^{-3}$ for DCIS cells).

The PHD-like domain in PHF5A is essential for its association with histone H3

It is widely accepted that the recognition of histone modification by reader proteins aids the execution of transcription and possibly AS events (39). The SF3a complex associates with lysine 4 trimethylated histone H3 (H3K4me3) via the chromodomain of CHD1 (40), which suggests the existence of underlying linkages between the spliceosome and histone modification. Among the multiple components of the SF3b complex, PHF5A is the only molecule that might contain a histone "reader" region, a plant homeodomain (PHD)-finger-like domain (Supplementary Fig. S5A). PHD zinc fingers are structurally conserved protein modules and epigenome readers (41). The PHD domain of PHF5A and that of other reported genes were aligned (Fig. 5A), and PHF5A showed some sequence homology to the known histone binder DNMT3L.

To explore the function of PHF5A in histone binding, we performed *in vitro* GST-PHF5A pulldown assays. The binding of PHF5A to histone H3 from purified bulk histones was confirmed by direct pulldown assays (Fig. 5B, left). A Western blot analysis of the PHF5A-bound histones further revealed significant methylation of H3 on lysine 4 but not lysine 36 (Fig. 5B, right). The binding of PHF5A to nucleosome H3K4me3 but not H3K4me0 provides further evidence (Fig. 5C). The association of the SF3b complex with H3K4me3 was also confirmed by IP using CA1a cell nuclear extracts (Fig. 5D). We then explored whether PHF5A or SF3B3 colocalized with H3K4me3 through a chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-Seq) analysis. We identified 17,545 PHF5A peaks distributed in 5,367 genes in CA1a cells. The peaks were highly enriched in promoters but not introns or exons (herein defined as gene bodies; Fig. 5E). Interestingly, the average distribution of all PHF5A-bound peaks, as well as that of SF3B3-bound peaks, was predominantly detected around transcription start sites (TSS) but absent from gene bodies toward the 3′ end, recapitulating the genomic distribution of H3K4me3 rather than that of H3K36me3 (Fig. 5F; Supplementary Fig. S5B). This indicates the possibility that PHF5A carries other SF3b components to chromatin and functions as an adaptor in the U2 snRNP complex while interacting with histones. Two-thirds of all the PHF5A and SF3B3 cooccupied peaks overlapped with H3K4me3-bound genes ($P < 1 \times 10^{-15}$; Fig. 5G), indicating that promoter-associated PHF5A showed strong cooccupancy with SF3B3. For the *RPL24* gene, which is described as an example, the genome browser view of the ChIP-Seq signals showed that the PHF5A, SF3B3, and H3K4me3 peaks were distributed mainly at the gene promoter, whereas H3K36me3 was mainly found in the gene body (Fig. 5H). On the basis of alignment (Fig. 5A), we speculated that the PHD-like domain was likely to mediate the interaction with the H3K4me3 in the presence of aspartate 47 (D47) residue. *In vitro* bulk histone binding assays showed that GST-$PHF5A_{PHD}$ and GST-$PHF5A_{N50A}$ retained the ability to bind H3K4me3, whereas GST-$PHF5A_{\Delta PHD}$ and GST-$PHF5A_{D47A}$ did not show binding ability (Fig. 5I). These findings suggested that PHF5A colocalizes with H3K4me3 in the genome, and its chromatin association requires the PHD finger.

PHF5A ablation enhances apoptosis in part through FASTK AS in breast cancer

To investigate the biological role of the interaction between PHF5A and histones, we performed genetic rescue experiments by introducing sgRNA-resistant wild-type (WT) PHF5A and PHF_{D47A}-mutant virus into endogenous PHF5A-knockdown CA1a cells (Fig. 6A). Phenotypically, we observed that $PHF5A_{WT}$ completely rescued the PHF5A-depleted cell proliferation capability and tumorigenesis, whereas the reintroduction of $PHF5A_{D47A}$ failed to do so (Fig. 6B and C). These findings suggested that the association between PHF5A and histones is essential for proper PHF5A function in breast cancer progression.

As shown by the flow cytometry analysis, PHF5A ablation led to notable increases in early- and late-stage apoptosis (Fig. 6D). WT PHF5A, but not $PHF5A_{D47A}$, completely rescued the PHF5A-depleted apoptotic effect in CA1a cells (Fig. 6D). A genome-wide mapping of the ChIP-Seq and mRNA-Seq data from PHF5A-knockdown and control CA1a cells revealed 584 genes whose chromatin was occupied by PHF5A and whose AS events were induced by PHF5A knockdown (Supplementary Fig. S6A). These genes constituted approximately two-thirds of the genes alternatively spliced by PHF5A. Among the validated AS target genes, FAS-activated serine/threonine kinase (FASTK) encodes an antiapoptotic protein (42). As demonstrated through RNA-Seq analyses, PHF5A knockdown appeared to switch full-length FASTK (FASTK-L) to an intron 5-retained variant (herein termed FASTK short, FASTK-S) in CA1a cells (Fig. 6E, the PSI shifted from 0.17 to 0.48). In the rescue assays, WT PHF5A could restore the FASTK splicing patterns in endogenous PHF5A-knockout CA1a cells, whereas the $PHF5A_{D47A}$ mutant failed to achieve this effect (Fig. 6E). This finding was further verified by qPCR analysis (Fig. 6F). The genome-browser view of the FASTK ChIP-Seq signals showed that the PHF5A, SF3B3 and H3K4me3 peaks presented similar patterns distributed at the FASTK promoter,

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

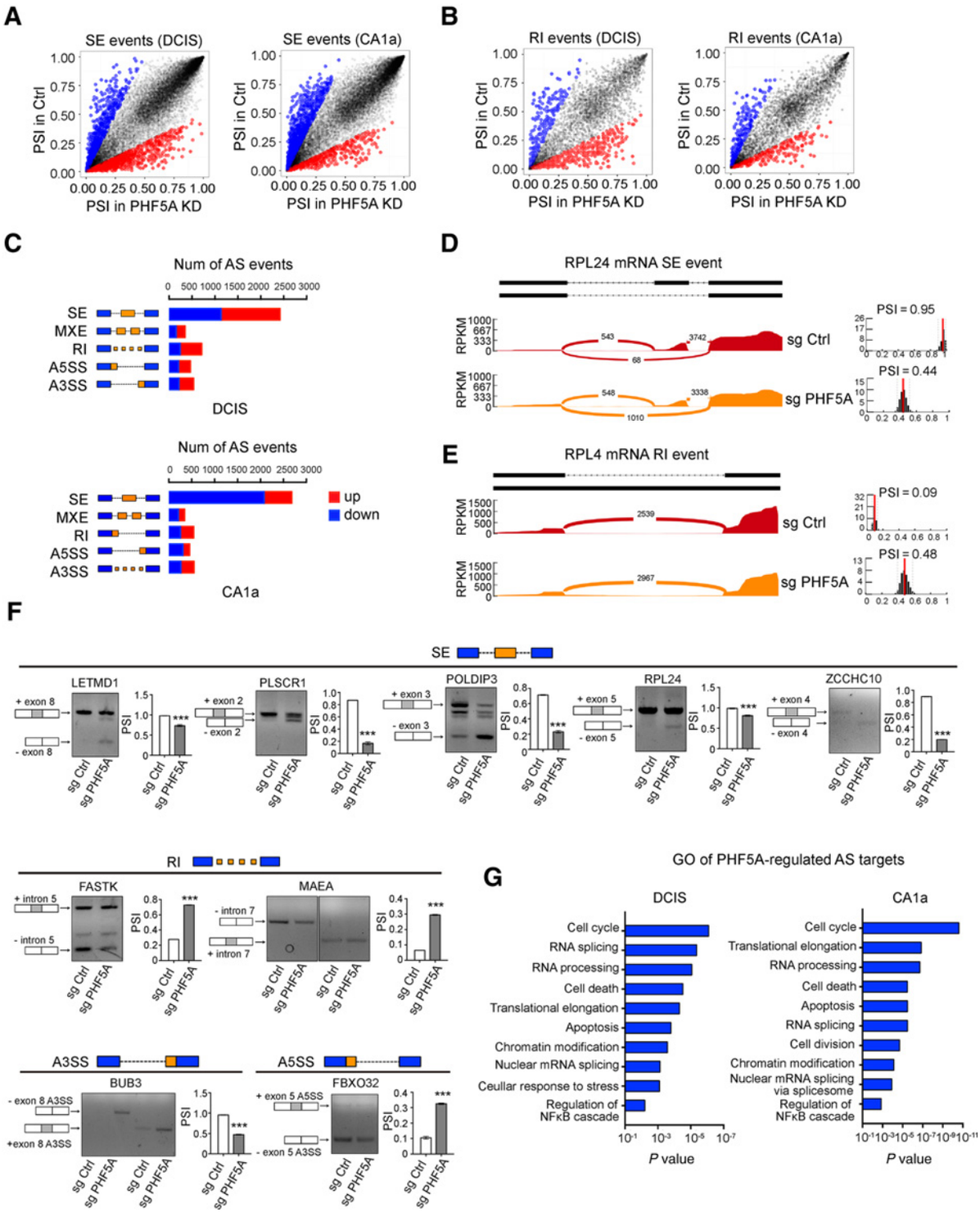


Figure 4.

PHF5A globally regulates splicing. **A,** PSI profiles of SE events identified in control and PHF5A-knockdown cells. **B,** PSI profiles of the RI events identified in control and PHF5A-knockdown cells. The colored dots represent significantly upregulated (red) or downregulated (blue) events in PHF5A knockdown cells compared with control cells (**A** and **B**). **C,** Quantification of the upregulated (red) or downregulated (blue) AS events. **D,** Example of an alternative exon. **E,** Example of a retained intron. The number of exon junction reads are indicated (**D** and **E**). **F,** Representative RT-PCR validation of PHF5A-regulated AS events. The structure of each isoform is illustrated in the diagrams. Individual data points are presented ($n = 3$; ***, $P < 0.001$; Student t test). **G,** Gene ontology analysis of PHF5A-regulated AS targets. Fisher exact P values were plotted for each enriched functional category.

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

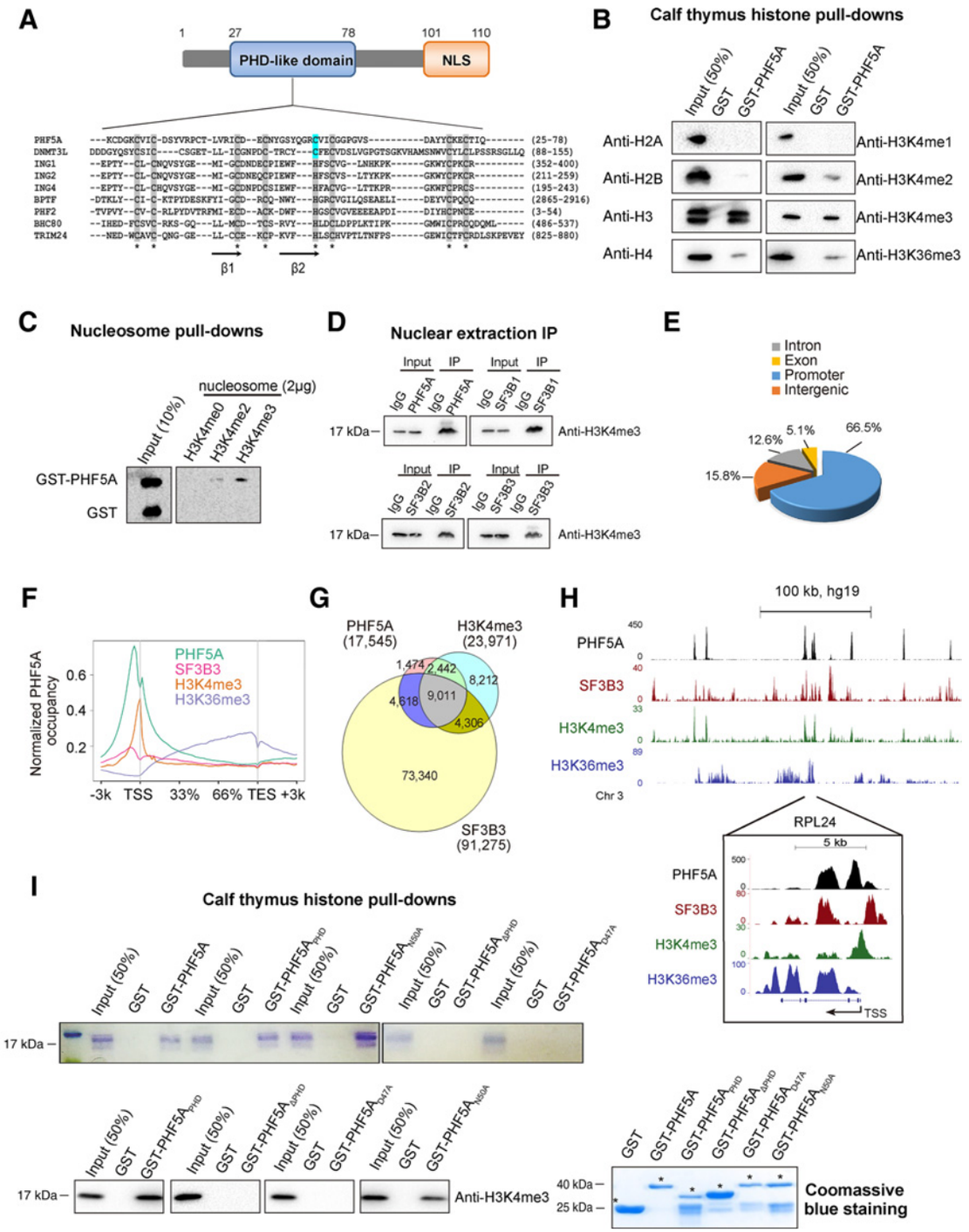


Figure 5.

The PHD-like domain in PHF5A is required for its association with histone H3. **A,** Structure-based alignments of representative PHD finger sequences. NLS, nuclear localization signal. The conserved zinc-coordinating residues are shown in gray, and the two core β strands are shown. The specific substitutions (histidine to cysteine) in PHF5A and DNMT3L are shown in blue. **B,** Western blot analysis of GST-PHF5A pulldowns from calf thymus histone proteins. **C,** Western blot analysis of reconstituted nucleosomes incubated with GST-tagged full-length PHF5A protein. **D,** Western blot analyses of H3K4me3 immunoprecipitated from CA1a cell nuclear extracts. **E,** Genomic distribution of ChIP-Seq peaks for PHF5A. **F,** Average genome-wide occupancies of PHF5A, SF3B3, H3K4me3, and H3K36me3 along the transcription unit. The gene body length is aligned on the basis of percentage from the TSS to the transcription end site (TES). **G,** Venn diagram showing the overlap among PHF5A-, H3K4me3-, and SF3B3-occupied genes. $P < 1 \times 10^{-15}$ (Pearson χ^2 test). **H,** Genome browser view of ChIP-Seq peaks in chromosome 3q12.3 regions. **I,** Coomassie blue staining (top) and Western blot analyses (bottom left) of pulldowns from calf thymus histone proteins. Coomassie blue staining (bottom right) of GST-PHF5A–purified truncations, mutants, and controls is shown.

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

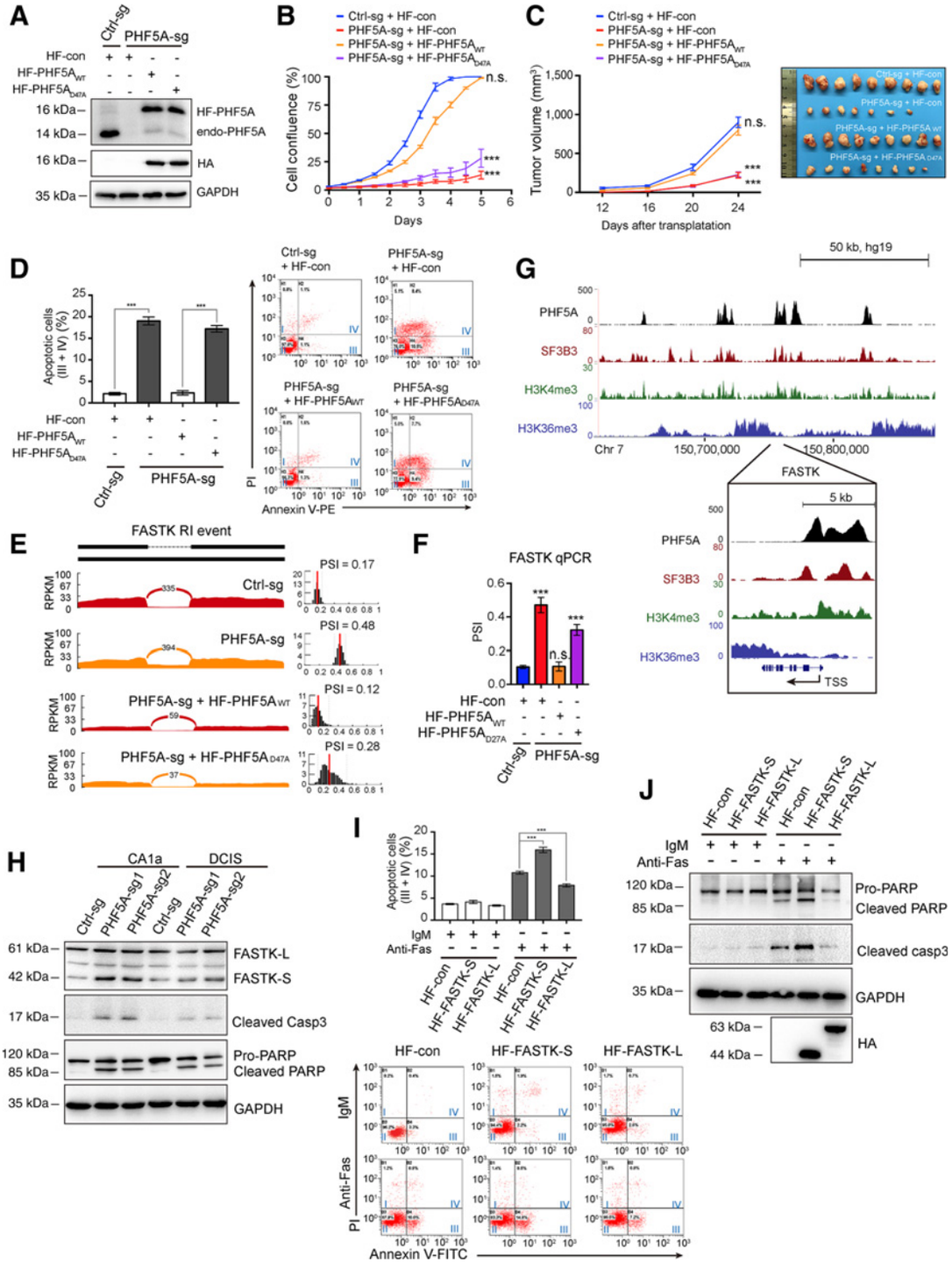


Figure 6.

PHF5A regulates FASTK splicing to antagonize apoptotic signaling in breast cancer cells. **A,** Western blot analysis of endogenous PHF5A-knockdown CA1a cells with stable expression of exogenous PHF5A$_{WT}$ or PHF5A$_{D47A}$. **B,** Cell proliferation assay of the cells described in **A**. **C,** Xenograft tumor growth induced by the cells described in **A** ($n = 9$). The data are presented as the means ± SEMs (***, $P < 0.001$; Student t test). The rest of the groups were all compared with Ctrl-sg + HF-con. **D,** Annexin V staining of the cells described in **A**. **E,** RI events of FASTK are shown. The number of exon junction reads are indicated. **F,** qPCR analysis of FASTK PSI in CA1a cells. **G,** Genome browser view of ChIP-Seq peaks in chromosome 7q36.1 regions (*FASTK* gene). **H,** PHF5A-knockdown cells were used to examine apoptotic markers by Western blot analysis. **I,** Annexin V staining of the CA1a cells expressing FASTK-S or FASTK-L. For **B**, **D**, **F**, and **I**, individual data points are presented ($n = 3$, ***, $P < 0.001$; Student t test). **J,** The cells described in **I** were used to examine apoptotic markers by Western blot analysis.

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

whereas the H3K36me3 peaks were distributed in the gene body (Fig. 6G). The ChIP-qPCR results validated the finding that $PHF5A_{D47A}$ failed to bound to the promoter of some target genes compared with $PHF5A_{WT}$ (Supplementary Fig. S6B), and a Western blot analysis indicated that the $PHF5A_{D47A}$ mutant could keep the stability of the SF3b complex (Supplementary Fig. S6C). These data suggested that the ability of PHF5A to regulate RNA splicing is also dependent on its anchoring to histones via its PHD-like domain.

Intron 5 retention introduced a frame shift and generated a premature stop codon, resulting in the possibility of producing a truncated FASTK protein (FASTK-S protein) consisting of 384 amino acids (Supplementary Fig. S6D). In addition, the FASTK-S transcript encodes different amino acids from FASTK-L within the range of 349 to 384 (Supplementary Fig. S6D). These two FASTK transcripts were further confirmed by Sanger sequencing of the RT-PCR products (Supplementary Fig. S6E). We then investigated whether AS of *FASTK* gene exerts an apoptotic effect on cells. As shown in Fig. 6H, the knockdown of PHF5A resulted in cleavage of caspase-3 (Casp3) and PARP. Moreover, we observed conversion of the FASTK-L (61 kDa) and FASTK-S (42 kDa) proteins. The 42-kDa bands were enriched in the protein extracts from PHF5A-knockdown cells, particularly CA1a cells (Fig. 6H). FASTK is a component of a signaling cascade that is initiated by ligation of the Fas receptor (43). Subsequently, CA1a cells carrying exogenous HA-Flag-tagged FASTK-S or FASTK-L protein were treated with anti-Fas CH11 antibody and subjected to apoptotic assessments. Intriguingly, a flow cytometry analysis revealed that cells transduced with exogenous FASTK-S showed the most significant apoptotic effect, whereas the FASTK-L group presented a decreased apoptotic effect (Fig. 6I). Consistently, a Western blot analysis revealed the same trends for the levels of caspase-3 and PARP cleavage (Fig. 6J). These results suggested that PHF5A depletion sensitizes cancer cells to apoptotic signaling partially through AS-mediated FASTK isoform conversion.

To explore whether the PHF5A-modulated FASTK-AS axis is relevant in a clinical setting, we analyzed the correlation between PHF5A expression and FASTK-AS events using mRNA-Seq data of 40 paired TNBC and adjacent normal breast tissues from FUSCC. The PHF5A ratios of paired nontumor to tumor tissue were negatively correlated with the FASTK PSI differences between nontumor and tumor tissues (Fig. 7A, R^2 = -0.662; P = 0.000003). We then evaluated the levels of PHF5A and cleaved caspase-3 in 373 breast cancer tissue specimens from FUSCC using IHC (Fig. 7B). A strong negative correlation was found between the PHF5A and cleaved caspase-3 levels (for total cases, $P < 10^{26}$; for TNBC cases, $P < 10^{7}$, Fig. 7B). These results supported the hypothesis that PHF5A might be an apoptosis suppressor in breast cancer. As illustrated in the model shown in Fig. 7C, in high PHF5A-expressing cancer cells, PHF5A carries the U2 snRNP complex, recognizes histones via its PHD-like domain and localizes to gene promoters, and subsequently, RNA splicing is properly regulated. In contrast, in low PHF5A-expressing noncancer cells, the level of U2 snRNP decreases, and the insufficient complex localizes to gene promoter regions, leading to AS events. As a result, PHF5A epigenetically controls the balance between pro- and antiapoptotic pathways and therefore promotes breast cancer progression partially through regulation of FASTK AS.

Discussion

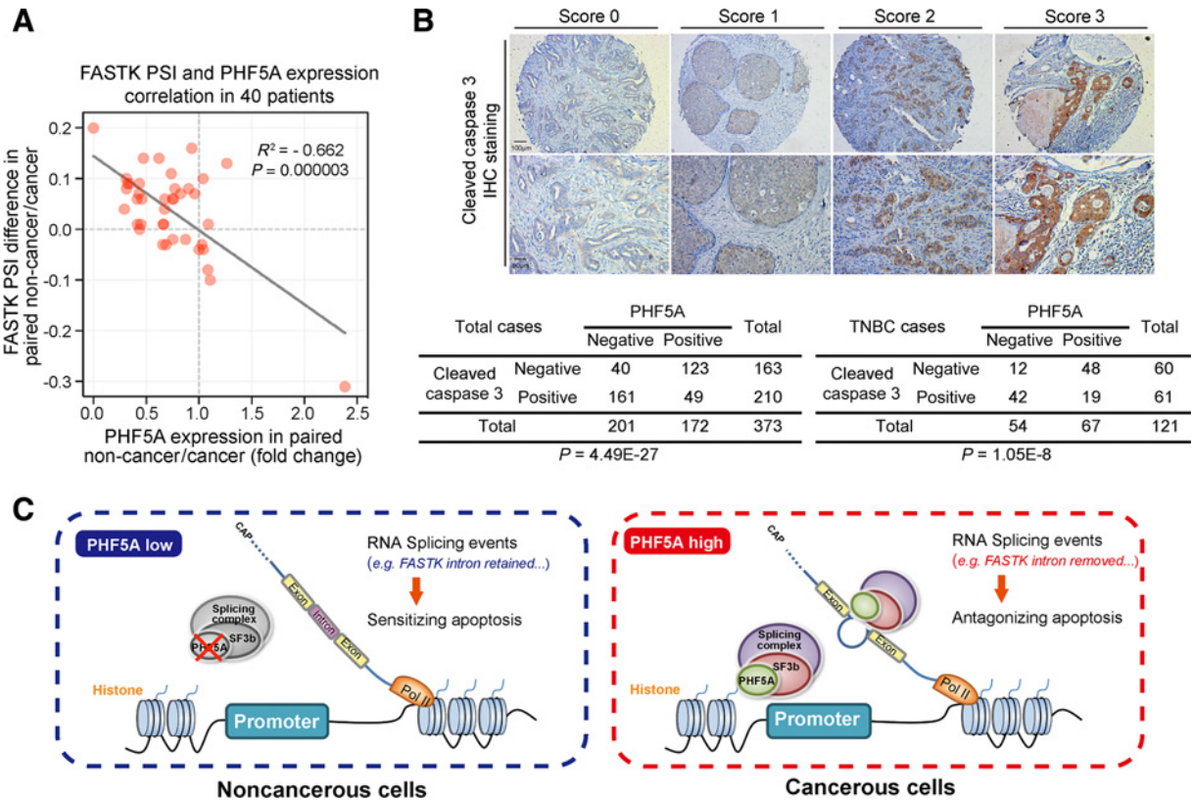
In this study, we performed an unbiased genetic screen to comprehensively investigate RBPs that act as promoters during breast cancer progression. To avoid the functional redundancy that masks phenotypic outcomes and reduce the false-negative rate in loss-of-function screens, we constructed a small CRISPR-Cas9 library to ensure a higher success rate. PHF5A functions as a breast cancer promoter and a strong prognostic indicator of poor postoperative survival in breast cancer patients.

Components of the spliceosome machinery were recently shown to be dysregulated in cancer (9), and PHF5A is implicated in pre-mRNA processing in brain tumors (38). The sequence context sets constraints for the recognition of AS sites by the core splicing machinery, and the level or functionality of the machinery also contributes to splicing control (10). Accordingly, several previous studies have shown that the depletion of certain spliceosome-associated RBPs could exert selective effects on certain AS events (44). Here, we showed that PHF5A binds with other U2 snRNP proteins and regulates SF3b complex stability. Therefore, high PHF5A expression in breast cancer might influence AS site selection by conferring U2 snRNP stability. This finding might provide a potential mechanism involving spliceosome dysregulation during tumor progression.

The data obtained in this study suggest that FASTK-AS regulation by PHF5A not only exists in cancer cell lines but also represents a common phenomenon in breast cancer patient specimens. Interestingly, a switch from FASTK-L to FASTK-S is obtained with the knockdown of PHF5A in CA1a but not DCIS cells. It is widely believed that genetic changes and molecular diversity are involved in the progression of breast carcinoma (45, 46). Therefore, AS event changes could be related to the multistep process of tumor progression, as either a contribution factor or an accompanying phenomenon, and this hypothesis warrants further study. In addition, we found that a transcript of FASTK, FASTK-S, encodes a truncated protein comprising 384 amino acids. Li and colleagues (42) reported that FASTN (a truncation mutant comprising amino acids 1–372), which is similar to FASTK-L, prevents Fas-induced activation of caspase-3, whereas FASTC (a truncation mutant comprising amino acids 372–550) does not significantly affect this pathway. Hence, FASTK-S transcripts encoding different amino acids within the range of 349–384 might acquire apoptotic activity. A similar example in which RI produces a novel protein is cyclin D1b, which presents more potent transforming activity than other isoforms in cancers (47). These cases emphasize the notion that gene transcripts whose introns are retained are not always subject to RNA nonsense-mediated decay but convert transcripts into multiple functional proteins (48). Although we cannot exclude the possibility that the AS of other genes might be involved in antiapoptotic signaling, the relatively lower PSI differences in other AS events obtained with the knockout of PHF5A suggests that their roles might not be critical.

PHD finger proteins are one of the largest families of epigenetic effectors and are capable of recognizing posttranslational histone modifications and unmodified histone tails (49). In this study, we identified that histone H3 is recognized by PHF5A and demonstrated that the ability of PHF5A to regulate splicing is dependent on its ability to bind chromatin. Our data strongly suggest the involvement of histone modification in PHF5A–SF3b complex–mediated AS, leading to the elucidation of a novel pathway that

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514



Total cases		PHF5A Negative	PHF5A Positive	Total
Cleaved caspase 3	Negative	40	123	163
	Positive	161	49	210
Total		201	172	373

P = 4.49E-27

TNBC cases		PHF5A Negative	PHF5A Positive	Total
Cleaved caspase 3	Negative	12	48	60
	Positive	42	19	61
Total		54	67	121

P = 1.05E-8

Figure 7.

The PHF5A implicates the existence of apoptotic signaling in clinical specimens. **A,** The scatterplot shows the correlation between the FASTK PSI and the PHF5A expression level in 40 paired TNBC and adjacent normal breast tissues. Gray line, the fitting curve (calculated by Pearson correlation). **B,** Representative images of cleaved caspase-3 IHC staining. Statistical analyses of the correlation between PHF5A and cleaved caspase-3 based on the IHC staining are shown (calculated by Pearson correlation). **C,** Diagram illustrating the epigenetic regulation of RNA splicing by PHF5A.

contributes to complicated communication between chromatin and RNA splicing in tumor progression.

Collectively, this study emphasizes the importance of the splicing factor PHF5A, which couples histone to the suppression of apoptosis signaling in tumor progression and serves as a new prognostic biomarker in breast cancer. In addition, the targeting of PHF5A might inhibit cancer growth via modulation of the core spliceosome. A recent study proposed that PHF5A, a core component of the SF3b complex, is a common cellular target of splicing-modulating chemical probes (37). Our data indicate that this epigenetic apoptotic suppressor plays a key role in breast cancer progression and should be critically considered for optimization of the current therapeutic strategy. Thus, our findings are of high translational relevance and are expected to have a substantial impact on the development of a promising epigenetic approach for investigating targeted therapies for breast cancers with high PHF5A expression.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors' Contributions

Conception and design: X. Hu, Z.-M. Shao
Development of methodology: Y.-Z. Zheng, X. Hu
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): Y.-Z. Zheng, H.-J. Shen, X.-G. Li, D. Ma, Y. Gong, Y.-R. Liu, F. Qiao, H.-Y. Xie, W.-L. Sun, H.-Y. Zhao, L. Yao
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Y.-Z. Zheng, M.-Z. Xue, H.-J. Shen, P. Wang
Writing, review, and/or revision of the manuscript: Y.-Z. Zheng, W.-J. Zuo, X. Hu
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): Y.-Z. Zheng, H.-J. Shen, B. Lian, D.-Q. Li, X. Hu
Study supervision: X. Hu

Acknowledgments

We thank Fei Lan for providing expert advice, reagents, and technical support. This study was supported by grants from the Ministry of Science and Technology of China (National Key R&D Program of China, MOST2016YFC0900300 to X. Hu), the National Natural Science Foundation of China (81572583 to Z.M. Shao and 81672601 to X. Hu) and the Shanghai Committee of Science and Technology Funds (15410724000 to X. Hu).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received November 16, 2017; revised March 15, 2018; accepted April 19, 2018; published first April 26, 2018.

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

References

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin 2015;65:87–108.
2. Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, et al. Alternative isoform regulation in human tissue transcriptomes. Nature 2008;456:470–6.
3. Venables JP. Aberrant and alternative splicing in cancer. Cancer Res 2004;64:7647–54.
4. David CJ, Manley JL. Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes Dev 2010;24:2343–64.
5. Oltean S, Bates DO. Hallmarks of alternative splicing in cancer. Oncogene 2014;33:5311–8.
6. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–74.
7. Lee SC, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. Nat Med 2016;22:976–86.
8. Chen M, Manley JL. Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. Nat Rev Mol Cell Biol 2009; 10:741–54.
9. Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. Oncogene 2016;35:2413–27.
10. Fu XD, Ares M Jr. Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet 2014;15:689–701.
11. Kornblihtt AR. Epigenetics at the base of alternative splicing changes that promote colorectal cancer. J Clin Invest 2017;127:3281–3.
12. Yuan H, Li N, Fu D, Ren J, Hui J, Peng J, et al. Histone methyltransferase SETD2 modulates alternative splicing to inhibit intestinal tumorigenesis. J Clin Invest 2017;127:3375–91.
13. Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27:1160–7.
14. Nagaraja GM, Othman M, Fox BP, Alsaber R, Pellegrino CM, Zeng Y, et al. Gene expression signatures and biomarkers of noninvasive and invasive breast cancer cells: comprehensive profiles by representational difference analysis, microarrays and proteomics. Oncogene 2006;25: 2328–38.
15. Hu X, Kim JA, Castillo A, Huang M, Liu J, Wang B. NBA1/MERIT40 and BRE interaction is required for the integrity of two distinct deubiquitinating enzyme BRCC36-containing complexes. J Biol Chem 2011;286: 11734–45.
16. Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science 2014;343:84–7.
17. Luo B, Cheung HW, Subramanian A, Sharifnia T, Okamoto M, Yang X, et al. Highly parallel identification of essential genes in cancer cells. Proc Natl Acad Sci U S A 2008;105:20380–5.
18. Barbie DA, Tamayo P, Boehm JS, Kim SY, Moody SE, Dunn IF, et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature 2009;462:108–12.
19. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531–7.
20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. PNAS 2005;102: 15545–50.
21. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. Nature 2012;490:61–70.
22. Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 2012;486:346–52.
23. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov 2012;2:401–4.
24. Shi X, Hong T, Walter KL, Ewalt M, Michishita E, Hung T, et al. ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. Nature 2006;442:96–9.
25. Luger K, Rechsteiner TJ, Richmond TJ. Preparation of nucleosome core particle from recombinant histones. Methods Enzymol 1999;304:3–19.
26. Luger K, Rechsteiner TJ, Flaus AJ, Waye MM, Richmond TJ. Characterization of nucleosome core particles containing histone proteins made in bacteria. J Mol Biol 1997;272:301–11.
27. Guo R, Zheng L, Park JW, Lv R, Chen H, Jiao F, et al. BS69/ZMYND11 reads and connects histone H3.3 lysine 36 trimethylation-decorated chromatin to regulated pre-mRNA processing. Mol Cell 2014;56:298–310.
28. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). Genome Biol 2008;9:R137.
29. Katz Y, Wang ET, Airoldi EM, Burge CB. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat Methods 2010;7:1009–15.
30. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 2009;37:1–13.
31. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2009;4:44–57.
32. Soule HD, Maloney TM, Wolman SR, Peterson WD Jr, Brenz R, McGrath CM, et al. Isolation and characterization of a spontaneously immortalized human breast epithelial cell line, MCF-10. Cancer Res 1990; 50:6075–86.
33. Hu M, Yao J, Carroll DK, Weremowicz S, Chen H, Carrasco D, et al. Regulation of in situ to invasive breast carcinoma transition. Cancer Cell 2008;13:394–406.
34. Gerstberger S, Hafner M, Tuschl T. A census of human RNA-binding proteins. Nat Rev Genet 2014;15:829–45.
35. Sotiriou C, Neo SY, McShane LM, Korn EL, Long PM, Jazaeri A, et al. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci 2003;100:10393–8.
36. Brosi R, Groning K, Behrens SE, Luhrmann R, Kramer A. Interaction of mammalian splicing factor SF3a with U2 snRNP and relation of its 60-kD subunit to yeast PRP9. Science 1993;262:102–5.
37. Teng T, Tsai JH, Puyang X, Seiler M, Peng S, Prajapati S, et al. Splicing modulators act at the branch point adenosine binding pocket defined by the PHF5A-SF3b complex. Nat Commun 2017;8:15522.
38. Hubert CG, Bradley RK, Ding Y, Toledo CM, Herman J, Skutt-Kakaria K, et al. Genome-wide RNAi screens in human brain tumor isolates reveal a novel viability requirement for PHF5A. Genes Dev 2013;27:1032–45.
39. Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. Regulation of alternative splicing by histone modifications. Science 2010; 327:996–1000.
40. Sims RJ III, Millhouse S, Chen CF, Lewis BA, Erdjument-Bromage H, Tempst P, et al. Recognition of trimethylated histone H3 lysine 4 facilitates the recruitment of transcription postinitiation factors and pre-mRNA splicing. Mol Cell 2007;28:665–76.
41. Sanchez R, Zhou MM. The PHD finger: a versatile epigenome reader. Trends Biochem Sci 2011;36:364–72.
42. Li W, Simarro M, Kedersha N, Anderson P. FAST is a survival protein that senses mitochondrial stress and modulates TIA-1-regulated changes in protein expression. Mol Cell Biol 2004;24:10718–32.
43. Li W, Kedersha N, Chen S, Gilks N, Lee G, Anderson P. FAST is a BCL-X(L)-associated mitochondrial protein. Biochem Biophys Res Commun 2004; 318:95–102.
44. Saltzman AL, Pan Q, Blencowe BJ. Regulation of alternative splicing by the core spliceosomal machinery. Genes Dev 2011;25:373–84.
45. Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, et al. Gene expression profiles of human breast cancer progression. Proc Natl Acad Sci U S A 2003;100:5974–9.
46. Yao J, Weremowicz S, Feng B, Gentleman RC, Marks JR, Gelman R, et al. Combined cDNA array comparative genomic hybridization and serial analysis of gene expression analysis of breast tumor progression. Cancer Res 2006;66:4065–78.
47. Lu F, Gladden AB, Diehl JA. An alternatively spliced cyclin D1 isoform, cyclin D1b, is a nuclear oncogene. Cancer Res 2003;63:7056–61.
48. Wong JJ, Au AY, Ritchie W, Rasko JE. Intron retention in mRNA: No longer nonsense: Known and putative roles of intron retention in normal and disease biology. Bioessays 2016;38:41–9.
49. Musselman CA, Kutateladze TG. Handpicking epigenetic marks with PHD fingers. Nucleic Acids Res 2011;39:9061–71.

Published OnlineFirst April 26, 2018; DOI: 10.1158/0008-5472.CAN-17-3514

Cancer Research
The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

AACR American Association for Cancer Research

PHF5A Epigenetically Inhibits Apoptosis to Promote Breast Cancer Progression

Yi-Zi Zheng, Meng-Zhu Xue, Hong-Jie Shen, et al.

Cancer Res 2018;78:3190-3206. Published OnlineFirst April 26, 2018.