

Supplementary Data

Histone demethylase PHF8 promotes epithelial to mesenchymal transition and breast tumorigenesis

Peng Shao¹, Qi Liu¹, Peterson Kariuki Maina¹, Jiayue Cui², Thomas Bair³, Tiandao Li⁴, Shaikamjad Umesalma⁵, Weizhou Zhang⁵ and Hank Heng Qi^{1,*}

¹ Department of Anatomy and Cell Biology; Carver College of Medicine, The University of Iowa, IA 52242, USA

² Department of Histology and Embryology, College of Basic Medical Sciences, Jilin University, Changchun 130021, China

³ Iowa Institute of Human Genetics, Carver College of Medicine, The University of Iowa, IA 52242, USA

⁴ McDonnell Genome Institute, Washington University, St. Louis, Mo 63108, USA

⁵ Department of Pathology, Carver College of Medicine, The University of Iowa, IA 52242, USA

* To whom correspondence should be addressed. Tel: +01-319-335-3084; Fax: +01-319-335-7198; Email: hank-qi@uiowa.edu

Supplementary Materials and Methods

Supplementary Figure Legends

Supplementary References

Supplementary Figures S1-S9

Supplementary Tables S1

Supplementary Materials and Methods

Plasmids, retroviruses, lentiviruses, and stable cell lines

pOZ-N-Flag-HA-wtPHF8 and mutPHF8 (F279S) vectors were described previously (1). The sequences of the scrambled (control) shRNA and the shRNAs targeting the *PHF8* ORF and 3'UTR (1,2) were cloned into the *AgeI* and *EcoRI* sites of the Tet-pLKO-puro vector (doxycycline-inducible system) which was a gift from Dmitri Wiederschain (Addgene plasmid # 21915) (3). These shRNAs were also cloned into the *AgeI* and *EcoRI* sites of the pLKO.1-TRC vector (constitutive expression system), which was a gift from David Root (Addgene plasmid # 10878) (4). The *MYC* coding sequence (571-1890 bp) of NM_002467 was cloned into the *XhoI* and *NotI* sites of the pOZ-C vector to generate the pOZ-MYC construct. Cell lines stably expressing the pOZ constructs (MCF10A-Mock, -wtPHF8, -mutPHF8, double stable MCF10A cells expressing control or PHF8 shRNA-2 combined with empty vector (shCtrl_Mock or shPHF8-2_Mock), PHF8 shRNA resistant wtPHF8 (shPHF8-2_wtPHF8) or mutPHF8 (shPHF8-2_mutPHF8); MCF10A-MYC; HaCaT-Mock and -MYC; MDA-MB-231-Mock and -PHF8; MT2-Mock and -PHF8) were established as previously described (5). The conditionally regulated MYC (MYC-ER) construct was generated by fusing the *MYC* coding sequence in frame with the hormone-binding domain of the human estrogen receptor (ER), and was then subcloned into the pOZ-C vector. Briefly, the cells were infected with the corresponding retrovirus generated in Phoenix A cells, and were selected using dynalbead-conjugated anti-IL-2RA (Interleukin-2 receptor alpha subunit) antibodies. MYC-ER expression was stimulated with 100 nM of 4-hydroxy tamoxifen (4-OHT) (Sigma-Aldrich). Puromycin selection was used to establish Tet-On inducible cell lines stably expressing lentiviral vectors (MCF10A-control shRNA, -PHF8 ORF shRNA). Additional lines stably expressing non-inducible constructs were established in pLKO-TRC vector in similar fashion (MT-2-control shRNA, and PHF8 ORF shRNA).

RNA-seq and bioinformatics analysis

Total RNA was isolated from MCF10A/control and MCF10A/PHF8 cells using the RNeasy Plus Mini Kit (Qiagen) following the manufacturer's protocol. The efficiency of PHF8 overexpression was determined by immunoblotting and qRT-PCR. RNA-seq libraries were made using the TruSeq Stranded Total RNA LT Sample Prep Kit (with Ribo-Zero™ Human/Mouse/Rat, RS-122-2201) and multiplexed with Illumina barcodes. Two biological duplicates of MCF10A/control and MCF10A/PHF8 samples treated with TGF- β 1 for 0, 24, 48 and 72 hours were submitted for sequencing on a HiSeq2500 instrument following protocols supplied by Illumina (San Diego, CA). 50-bp paired-end reads were generated. More than 20 Mb reads were produced for each

library. The paired-end raw reads were aligned against the hg19 assembly of the human genome using the TopHat (v2.0.13) program (6). Counts were generated using featureCounts from the Subread package (Release 1.4.6) (7), and the initial differential expression was assessed with DESeq2 (8). The heat map was generated by GENE-E software hosted at the Broad Institute (<http://www.broadinstitute.org/cancer/software/GENE-E/>).

Gene set enrichment analysis

Gene set enrichment (GSEA) was performed using the GSEA software (9,10). Dataset files were developed based on normalization of expression intensities by DESeq2. Specifically, the log₂ fold change values were compared for two phenotypes in our analysis: MCF10A-Mock versus MCF10A-PHF8. GSEA was executed using the default settings, the permutation type was set to Gene_set with 1000 permutations, and the metric for ranking genes was set to Diff_of_Classes because normalized expression data by DESeq2 was log₂ transformed. Three datasets of EMT signature genes were used as references. (11-13) Each EMT signature collection was analyzed individually in multiple GSEA runs.

Transient transfection

Cells were 80-90% confluent in 6-well plates when transfection experiments were conducted. miR-22 mimics (GenePharma) or inhibitors (Integrated DNA Technologies) were used at a final concentration of 90 nM or 3 nM, respectively. In the siRNA knockdown experiments, siRNA duplexes targeting the human *PHF8* (OriGene, SR308029A and B) and *MYC* (OriGene, SR303025A and C) were used at a final concentration of 40 nM. After transfection for 48 hours (miRNA mimics and inhibitors) or 72 hours (siRNAs), cells were harvested for analysis of gene expression. All transfection of RNA duplexes was performed using the Lipofectamine RNAiMAX transfection reagent (Life technologies).

Colony formation assays

Both monolayer colony formation and anchorage-independent growth in soft agar were tested. All assays were conducted in triplicate, and three independent experiments were performed. For the monolayer colony formation assay, 500 cells that either stably expressed PHF8 or had been subjected to doxycycline-inducible PHF8 knockdown were plated in 6-well plates in triplicate, and allowed to grow in the appropriate culture medium for two weeks. Fresh medium was supplied every 3 days. Colonies were stained with 0.1% crystal violet dye after being fixed

in 4% paraformaldehyde. Colonies larger than 5 μm in diameter were counted as positive for growth.

To assay anchorage-independent growth in soft agar, MDA-MB-231 and BT474 cells subjected to doxycycline-inducible PHF8 knockdown by shRNAs were added to 1.5 ml of growth medium with 0.4% agar and layered onto 2 ml of 0.7% agar beds in six-well plates, respectively. Cells were fed with 1 ml of medium in 0.4% agar every 4 days for 3 weeks, after which the colonies were stained with 0.01% crystal violet overnight and photographed. The number of colonies larger than 100 μm in diameter was counted.

Transwell migration assay

MDA-MB-231 and MCF10A cells subjected to doxycycline-inducible PHF8 knockdown, or expressing pOZ-PHF8-HA or control plasmid, were serum-starved for 4 hours. A total of 3×10^4 cells/well was added to the top chambers of 24-well transwell inserts (Corning Costar; 8- μm pore size). The lower compartment was filled with standard cell culture medium. Cells were allowed to migrate for 18-20 hours, after which 4% paraformaldehyde was added to each chamber for 10 minutes. Membranes were stained with 0.1% (w/v) crystal violet for 20 minutes and subsequently washed thoroughly with PBS to visualize nuclei. Cells that had not migrated to the lower compartment were removed with a cotton swab. The number of migrating cells was counted using the ImageJ software. Experiments were repeated a minimum of three times. Data are shown as the mean \pm S.D. of three independent experiments.

Acinar formation assay

Acinar formation of MCF10A cells in 3D Matrigel cultures was performed as described by Debnath et al. (14). Briefly, a total of 5×10^3 cells/well were plated in growth factor-reduced Matrigel (BD Biosciences, 354230) coated chamber slide (Falcon, 354108) and grown in 400 μl of DMEM:F12 medium supplemented with 2.5% Matrigel, 2% horse serum, 500 ng/ml Hydrocortisone, 10 $\mu\text{g/ml}$ Insulin, 100 ng/ml Cholera Toxin and 5 ng/ml EGF. Media was changed every 4 days with assay media containing 2.5% Matrigel and 5 ng/ml EGF. Acini like structures were counted at day 12 and confocal image was taken using Carl Zeiss LSM 700.

Cell proliferation assays

Cell proliferation was assessed using the MTT assay, using at least three replicates. Cells were seeded in 96-well plates at 1×10^3 cells per well, and allowed to adhere overnight. Thereafter, on each day 20 μl of MTT (5 $\mu\text{g}/\mu\text{l}$ in PBS) was added to the medium. Four hours later, the medium

was removed with a needle and syringe, and blue formazan crystals trapped in the cells were dissolved in DMSO (120µl) by incubation at 37°C for 30 minutes. The absorbance at 490 nm was then measured using a Microplate Reader (VersaMax).

Immunoblotting, immunofluorescence, immunohistochemistry and antibodies

For immunoblotting experiments, cells were washed twice with cold PBS and lysed at 4°C in RIPA buffer containing 50 mM Tris (pH 7.4), 150 mM NaCl, 0.1%SDS, 1mM EDTA, 1% (vol/vol) Triton X-100, 1% sodium deoxycholate, 0.5mM EDTA, 1 mM sodium vanadate, 50 mM NaF and complete protease inhibitors (Roche). Lysates were cleared by centrifugation at 13000 rpm for 10 minutes, and aliquots of the cell extracts containing equal amounts of protein were analyzed for protein expression by western blotting. Proteins were separated by SDS-PAGE and transferred to nitrocellulose filters (BioRad). Membranes were blocked in TBST buffer (10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.05% (vol/vol) Tween-20) containing 5% (wt/vol) milk. Proteins were detected using the Amersham ECL reagent.

For immunofluorescence experiments, cells were washed twice with PBS and fixed for 5 minutes with 4% formaldehyde in PBS, then permeabilized for 5 minutes with 0.1% Triton X-100. After rinses with PBS, blocking solution (2% BSA in PBS, pH7.5) was applied for 30 minutes and primary antibodies were added in blocking buffer for 1 h at room temperature or overnight at 4°C. After two washes with PBS/0.1% Triton X-100 for 10 minutes, cells were incubated with DAPI, Alexa Fluor® 594 Phalloidin (A12381, Life Technologies) and secondary antibodies conjugated with fluorescent dyes (Life Technologies) for 1 h, washed again with PBS/0.1% Triton X-100 for 10 minutes, and mounted in Vectashield mounting medium (Vector Laboratories). Slides were observed at 40x using a Nikon Eclipse 80i microscope, and images were recorded with a Nikon digital camera.

For immunohistochemical staining, histological tissue arrays (BR10010c, BR1503d, BR244 and BR082a) purchased from USBIOMAX were routinely deparaffinized in xylene, 3 x 5 minutes, and rehydrated through an alcohol series. Protease-induced epitope retrieval was carried out and followed by heat-induced epitope retrieval. Tissues were pre-treated with trypsin working solution (0.05%) and incubated for 10–20 minutes at 37°C in a humidified chamber. Sections were allowed to cool down to room temperature. After a short rinse in distilled water, the sections were boiled in a microwave oven at 800 Watts in 100 ml of 0.01M citrate buffer (pH 6.0) for 5 x 2 minutes. After treatment, the beaker was left on the bench to cool for 30 minutes.

Endogenous peroxidase activity was blocked with 3% H₂O₂ in PBST for 10 minutes. After blocking in the blocking reagent (Immunoperoxidase Secondary Detection System, Millipore), immunohistochemical blotting was performed using PHF8 antibody (PHF8 IHC-00343, Bethyl Laboratories) and c-Myc Tag Antibody (MA1-980, ThermoFisher) at 1:100 dilution, in PBST with 1% goat serum. A sample lacking primary antibody served as the negative control. Biotinylated secondary goat anti-mouse IgG/goat anti-rabbit IgG (Millipore) antibody was used in labelling with IHC Select® Immunoperoxidase Secondary Detection System (Millipore) for PHF8 detection. The tissue arrays were examined using a Nikon Eclipse 80i microscope, and images were recorded using a Nikon digital camera. The expression levels of PHF8 and MYC were semi-quantitatively based on staining intensity and distribution using the immunoreactive score (IRS) as described by Su et al. (2014) (15). Primary antibodies were: anti-PHF8 for use in western blotting, made in house (1); anti-CDH1 (3195S), anti-CDH2 (13116S), anti-SNAI1 (3879S), and anti-PTEN (9552S) purchased from Cell Signaling Technologies; anti-ZEB1 (sc-25388), anti-c-MYC (sc-40), anti-β-actin (sc-47778), anti-HA (sc-7392), anti-VIM (sc-6260), anti-p-SMAD2 (sc-135644), and anti-GFP (sc-9996) purchased from Santa Cruz Biotechnology; and anti-PHF8 for use in immunofluorescence, purchased from Abcam (ab36068). Secondary antibodies were anti-mouse- or anti-rabbit-conjugated horseradish peroxidase (BioRad).

Real-time PCR assay

RNA was isolated from 80% confluent plates of cells using TRIzol Reagent (Invitrogen). Total RNA (500 ng) was used for reverse transcription (RT) reactions using the M-MLV reverse transcriptase (Promega). Levels of expression for mature miRNAs and mRNAs were quantified by SYBR Green-based real-time PCR, using the iTaq Universal SYBR Green Supermix (Bio-Rad). miRNA-specific primers were designed as described previously (16). Real-time PCR reactions for miRNA were performed as follows: one cycle of 95°C for 30 seconds followed by 40 cycles of 95°C for 5 sections and 60°C for 20 sections. For mRNA detection, one cycle of 95°C for 30 sections was followed by 40 cycles of 95°C for 10 sections, 55°C for 15 sections and 72°C for 25 sections. The expression levels of mRNAs and miRNAs relative to RPL13A or GAPDH and U6 snRNA, respectively, were determined using the $2^{-\Delta\Delta C_t}$ method. Error bars represent standard deviations. Comparative real-time PCR was performed minimally in triplicate.

Dual luciferase reporter assay

Dual-luciferase reporters were derived from the psiCHECK-2 vector (Promega). The 3' untranslated region (UTR) of human PHF8 containing a wild type (wt) miR-22 binding site (PHF8-3'UTR-wt) or a mutated site complementary to the seed region of miR-22 (PHF8-3'UTR-mut) was cloned into the downstream of the 3' UTR of the Renilla luciferase (hRluc) gene. HEK 293T cells were plated in 48-well plates (4×10^4 per well) one day before transfection, and then were cotransfected with 100 ng of reporter plasmid (psiCHECK-3'UTR) and 20 nM of the synthetic miR-22 mimics or negative control (GenePharma). Transfection was performed using lipofectamine 2000 (Invitrogen), according to the manufacturer's protocol. Luciferase activities were measured 30 h post-transfection, using the Dual-Luciferase® Reporter Assay System (Promega) and an FB12 Luminometer (Berthold). For each sample, Renilla luciferase activity was normalized to firefly luciferase expression (Rluc/Fluc). Data shown are the Mean \pm SD from at least three independent experiments performed in duplicate.

Bioluminescent imaging

For bioluminescent imaging, FVB/N mice were anesthetized and injected with 100 μ l of D-luciferin (15 mg/ml). Animals were imaged in an IVIS 200 chamber within 6 min after D-luciferin injection, and the photon flux were recorded using Living Image software (Xenogen) at the indicated time points. To measure tumor growth, photon flux was calculated for each mouse by using a circular region of interest over the tumor and normalized to the value obtained from the signal of control tumors at day 0.

Oligonucleotides

Primers or Probes		Sequences	Annotation
PHF8			
PHF8_3'UTR	Forward	GCCTTCTCCACTGAGGAGCAGGTA	qPCR targeting the 3'UTR of PHF8
	Reverse	CTCCTCATCCTGCCTTCCAGCTCT	
PHF8_ORF	Forward	GCAAACCGCAGCACCACACCT	qPCR targeting the ORF of PHF8
	Reverse	CGAGTCTCTGCTTTGCTGTG	
shPHF8-1	Sense	CCGGTGGCCTAGAAATGCCAACTTCACTCGAG TGAAGTTGGCATTCTAGGCCTTTTG	pLKO-TET-ON-PHF8 shRNA; sequence targeting the 3' UTR of PHF8 is underlined.
shPHF8-2	Sense	CCGGGCTTCATGATCGAGTGTGACACTCGAGT GTCACACTCGATCATGAAGCTTTTG	pLKO-TET-ON-PHF8 shRNA; sequence targeting the ORF of PHF8 is underlined.
shPHF8-3	Sense	CCGGTGGAGACACCGAAGATTGTTTCGCTCGAG CGAACAACTCTCGGTGTCTCCTTTTG	pLKO-TET-ON-PHF8 shRNA; sequence targeting the ORF of PHF8 is underlined.

siPHF8-A		rArGrCrArArGrArArGrUrArGrArCrArGrGrCrUrAGG	siRNA for PHF8
siPHF8-B		rGrGrArGrGrArCrUrArUrArCrArCrArGrArUrGrArGrGAC	siRNA for PHF8
MYC and MYC target genes			
siMYC-A		rUrCrCrUrGrArGrACrArGrArUrCrAGrCArCrArArCCG	siRNA for MYC
siMYC-C		rCrCrGrArGrGrArGrArUrGrUrCrArGrArGrGrCrGrAAC	siRNA for MYC
MYC	Forward	TTCGGGTAGTGGAAAACCAG	
	Reverse	CCTCCTCGTCGCAGTAGAAA	
NCL	Forward	TGGCCCAGTCCAAGGTAAC	
	Reverse	TGGCCCAGTCCAAGGTAAC	
CDK1	Forward	GGCCAGAAGTGAATCTTTACA	
	Reverse	GGATCATAGATTAACATTTTC	
CDK4	Forward	GAAACTCTGAAGCCGACCAG	
	Reverse	AGGCAGAGATTGCTTGTGT	
CDC25A	Forward	TAAGACCTGTATCTCGTGGCTG	
	Reverse	CCCTGGTTCAGTCTATCTCT	
CCNA2	Forward	CTGCATTTGGCTGTGAACTAC	
	Reverse	ACAAACTCTGCTACTTCTGGG	
	Reverse	TGAGCCATTCGCAGTTTCAC	
EMT-related genes and control			
TWIST1	Forward	CCGGAGACCTAGATGTCATTG	
	Reverse	CACGCCCTGTTTCTTTGAAT	
ZEB1	Forward	CTGATTCTACACCGCCCAA	
	Reverse	AGCGCTTTCCACATTTGTCA	
ZEB2	Forward	CGCAAACAAGCCAATCCCA	
	Reverse	CACACTAGCTGGACTCGTCT	
CDH1	Forward	CGAGAGCTACACGTTACGG	
	Reverse	GGGTGTCGAGGGAAAAATAGG	
CDH2	Forward	CGACGAATGGATGAAAGACC	
	Reverse	CATAGTCCTGCTCACCACCA	
SNAI1	Forward	CCCTCAAGATGCACATCCGAA	
	Reverse	TGGCACTGGTACTTCTTGACATCTGA	
GAPDH	Forward	ATGCCTCCTGCACCACCAAC	
	Reverse	GGGGCCATCCACAGTCTTCT	
RPLP0 (36B4)	Forward	CCTGGAGGAGAAGAGGAAAGAGA	
	Reverse	TTGAGGACCTCTGTGTATTTGTCAA	
CHGA	Forward	GGAGGATCGACCGACAGAC	
	Reverse	GAGCAGAAGAGCCAGGACAG	
ChIP-qPCR			
SNAI1_1	Forward	GGCCTAGCGAGTGGTTCTTC	Targeting the TSS of SNAI1
	Reverse	CCAACGCACCTGGATTAGAG	
SNAI1_2	Forward	CTCTGAGCGGTGAGGGTTAG	Targeting the gene body of SNAI1
	Reverse	CCCCAAAATAGAGCCCTGTGA	

ZEB1_1	Forward	CGGCTTTACGACATCACCTT	Targeting the TSS of ZEB1
	Reverse	CCACCACACCTGAGGAAAAC	
ZEB1_2	Forward	TAGGGGAATGCGAGTGTGT	Targeting the gene body of ZEB1
	Reverse	CTCAAGGAGAAAACACATCCCT	
GAPDH_1	Forward	GCTCTCTGCTCCTCCTGTTC	Targeting the TSS of GAPDH
	Reverse	CCTTCAGGCCGTCCCTAGC	
GAPDH_2	Forward	GGCCTCCAAGGAGTAAGACC	Targeting the 3'UTR of GAPDH
	Reverse	AGGGGTCTACATGGCAACTG	
CHGA_1	Forward	GGAGGATCGACCGACAGAC	Targeting the TSS of CHGA
	Reverse	GAGCAGAAGAGCCAGGACAG	
CHGA_2	Forward	TTGAGGTTGTGCTCTTGTGG	Targeting the gene body of CHGA
	Reverse	TGATCCCTGGTCCTGATTGT	
Luciferase assay			
PHF8_miR22 target_WT	Sense	TCGAGTTGTCCATTAACACCCCTTCTTGATCTT TCAAAGGCAGCTAATTGCTAGCAGC	The wild-type of miR-22 target site is underlined.
	Antisense	GGCCGCTGCTAGCAATTA <u>GCTGCC</u> CTTTGAAAG ATCAAGAAGGGGTGTTAATGGACAAC	
PHF8_miR22 target_MUT	Sense	TCGAGTTGTCCATTAACACCCCTTCTTGATCTT TCAAAGCGAGGTTATTGCTAGCAGC	The mutant of miR-22 target site is underlined
	Antisense	GGCCGCTGCTAGCAATA <u>ACCTCGC</u> TTTGAAAG ATCAAGAAGGGGTGTTAATGGACAAC	
miRNA			
let-7c-5p_RT	RT_primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGC ACTGGATACGACAACCAT	
let-7c-5p_F	Forward	GCCAGCTGAGGTAGTAGGTTGTA	
miR-31-5p_RT	RT_primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGC ACTGGATACGACAGCTAT	
miR-31-5p_F	Forward	ACTCAAGGCAAGATGCTGGCA	
miR-22-3p_RT	RT_primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGC ACTGGATACGACACAGTTC	
miR-22-3p_F	Forward	GCGACAAGCTGCCAGTTGAA	
miR-182-5p_RT	RT_primer	GTCGTATCCAGTGCAGGGTCCGAGGTATTCGC ACTGGATACGACTGTGTG	
miR-182-5p_F	Forward	TGGCTTTGGCAATGGTAGAACT	
miRNA universal qPCR primer	Reverse	GTGCAGGGTCCGAGGT	
U6	Forward	CGCTTCGGCAGCACATATAC	Control for the miRNA assay
	Reverse	AAAATATGGAACGCTTCACGA	

Supplementary Figure Legends

Supplementary Figure S1. PHF8 regulates acinar formation in MCF10A cells. (A) Western blot analysis of PHF8 expression in MCF10A-Mock and MCF10A-wtPHF8 cells (left panel), and cells expressing control shRNAs or PHF8 shRNAs (right panel). (B) The size of acini was measured on day 12. Immunostaining was performed for F-actin (red) and the apoptosis marker cleaved caspase-3 (green). Nuclei were co-stained with DAPI (blue). Scale bars, 50 μ m. The dot plot shows the size distribution for each condition (right panel), with the median indicated by a horizontal line (at least 100 acini per experiment; $**P < 0.001$, Mann-Whitney test).

Supplementary Figure S2. PHF8 regulates the cell morphology and expression of EMT-related genes induced by TGF- β 1. (A) Western blot analysis of proteins encoded by selected EMT-related genes in A549 cells stably overexpressing vector only (Mock), wild type PHF8 (WT), or mutant PHF8 (Mut). The relative signal intensities for indicated proteins are normalized to β -actin and shown below each lane. (B) Western blot analysis of PHF8 expression in A549 cells with inducible expression of control shRNA or two different PHF8 shRNAs. (C) Morphological change induced by TGF- β 1 at indicated time points in A549 cells expressing control or PHF8 shRNAs. (D) qPCR analysis of *RPLP0* in MCF10A-Mock and MCF10A-wtPHF8 cells treated with TGF- β 1 treatment for 0 and 1.5 hours. (E) A heat map of the expression of Taube's EMT genes (up and down regulated), retrieved from normalized RNA-seq data as in Figure 1F. (F) Normalized enrichment scores for the genes between MCF10A-Mock and MCF10A-wtPHF8 cells treated with TGF- β 1 for 24, 48 and 72 hours, using the EMT gene signatures of Taube et al. (2010) (13), Huang et al. (2012) (11) and Groger et al. (2012) (12), respectively. (G) qPCR analysis of *GAPDH* and *CHGA* expression in MCF10A cells expressing control shRNAs or PHF8 shRNA-2 without or with TGF- β 1 treatment for 1.5 hours. (H) Western blot analysis of EMT-related genes in A549 cells expressing control or two different PHF8 shRNAs with TGF- β 1 treatment at different time points. (I and J) Comparison of cell morphology and proteins encoded by EMT-related genes in double stable MCF10A cells expressing control or PHF8 shRNA-2 combined with empty vector (shCtrl_Mock or shPHF8-2_Mock), PHF8 shRNA-2 resistant wtPHF8 (shPHF8-2_wtPHF8) or mutPHF8 (shPHF8-2_mutPHF8).

Supplementary Figure S3. ChIP-qPCR assay on TSS (amplicon 1) and gene body (amplicon 2) for *GAPDH*, *ZEB1* and *CHGA*. (A) The amplicons of *GAPDH*, *ZEB1* and *CHGA* for ChIP-qPCR were illustrated. (B) ChIP-qPCR assays with PHF8 antibody were performed for *GAPDH*, *ZEB1* and *CHGA*.

Supplementary Figure S4. Western blot analysis of PHF8 expression in MCF10A and human breast cancer cell lines. The relative signal intensity for PHF8, normalized to β -actin, is shown below the lane.

Supplementary Figure S5. PHF8 overexpression blocks TGF- β induced cell cycle arrest. Quantitation of the cycle distributions in MCF10A-Mock and MCF10A-wtPHF8 cells following treatment with TGF- β for the indicated time, as assessed by fluorescence activated cell sorting (FACS). P values represent three independent experiments. *: $p < 0.05$, **: $p < 0.001$; unpaired two-tailed Student's t-test.

Supplementary Figure S6. Schematic illustration of predicted sites of miRNA targeting within the *PHF8* 3' UTR using TargetScan Human (Release 6.2). The seed sequences and complementary PHF8 sequences are shown.

Supplementary Figure S7. PHF8 positively regulates anchorage-independent growth and tumor growth of breast cancer cells *in vitro* and *in vivo*. (A) Western blot analysis of PHF8 expression in MDA-MB-231, BT474 cells expressing control shRNA, siRNA or PHF8 shRNAs and siRNAs, respectively. (B) Soft agar colony assays in BT474 cells expressing doxycycline inducible control or PHF8 shRNAs. The quantification is shown at the right panel. (C) Western blot analysis of PHF8 expression in MDA-MB-231 overexpressing empty vector (Mock) or wild type PHF8. (D) Knockdown of PHF8 by shRNAs in MDA-MB-231 cells downregulated the expression of cell cycle and migration related genes, as detected by qPCR. SD was obtained from three independent experiments. *: $p < 0.05$, **: $p < 0.01$; unpaired two-tailed Student's t-test. (E) Representative image of tumor growth of MDA-MB-231 cells stably expressing control and PHF8 shRNA implanted in nude mice. Tumors collected at week 6 after injections were weighed and assessed for volume, as quantitated at the right panel. **: $p < 0.01$; unpaired two-tailed Student's t-test. (F) Western blot analysis of PHF8 expression using anti-PHF8 antibody in MT2 stable cell lines expressing control shRNAs or PHF8 shRNAs and SKBR3 cells. (G) IHC analysis of PHF8 expression in MT2 xenograft tumors. (H) MT2 stable cell lines expressing empty vector (Mock) or HA-tagged PHF8 was confirmed by western blotting using anti-HA antibody. (I) Representative image of tumor growth of mouse breast cancer MT2 cells in FVB/N mice into which MT2 cells stably overexpressing empty vector (Mock) or PHF8 were implanted.

Tumors collected at day 60 were weighed and assessed for volume, as quantitated at the right panel. Error bars denote SD. *: $p < 0.05$, **: $p < 0.01$; Mann Whitney test (two-tailed) was used.

Supplementary Figure S8. PHF8 is upregulated in breast cancers. (A) Levels of the *PHF8* mRNA in multiple cancer types based on RNA-seq data from The Cancer Genome Atlas (TCGA). (B) RNA-seq data from The Cancer Genome Atlas (TCGA) for *PHF8* expression in breast cancers versus normal tissue was analyzed using StartBase 2.0. The case numbers are indicated in the parentheses. (C) Levels of the *PHF8*, *JHDM1D* and *PHF2* mRNAs based on ONCOMINE analysis using the indicated thresholds. (D) *PHF8* mRNA expression in breast cancer samples from ONCOMINE. (E) Correlation of *PHF8* expression with *ERBB2*, *ER* and *PR* status, based on TCGA data for 593 cases of breast cancer. The number of cases is shown in parentheses.

Supplementary Figure S9. PHF8 protein levels are positively correlated with metastasis in the MMTV-ERBB2 mouse model of breast cancer. IHC analysis of PHF8 expression in small and large tumors obtained from six and eight month old MMTV-ERBB2/FVB mice with anti-PHF8 antibody (1). Nuclei were counterstained with Surgipath Hematoxylin. Tumor stages are indicated. Images are at 20X magnification unless otherwise indicated.

Supplementary References

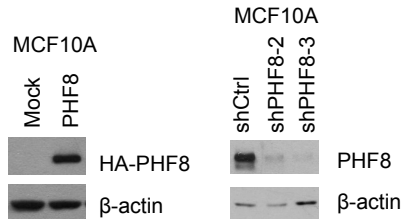
1. Qi, H.H., Sarkissian, M., Hu, G.Q., Wang, Z., Bhattacharjee, A., Gordon, D.B., Gonzales, M., Lan, F., Ongusaha, P.P., Huarte, M. *et al.* (2010) Histone H4K20/H3K9 demethylase PHF8 regulates zebrafish brain and craniofacial development. *Nature*, **466**, 503-507.
2. Fortschegger, K., de Graaf, P., Outchkourov, N.S., van Schaik, F.M., Timmers, H.T. and Shiekhata, R. (2010) PHF8 targets histone methylation and RNA polymerase II to activate transcription. *Mol Cell Biol*, **30**, 3286-3298.
3. Wiederschain, D., Wee, S., Chen, L., Loo, A., Yang, G., Huang, A., Chen, Y., Caponigro, G., Yao, Y.M., Lengauer, C. *et al.* (2009) Single-vector inducible lentiviral RNAi system for oncology target validation. *Cell Cycle*, **8**, 498-504.
4. Moffat, J., Grueneberg, D.A., Yang, X., Kim, S.Y., Kloepper, A.M., Hinkle, G., Piquini, B., Eisenhaure, T.M., Luo, B., Grenier, J.K. *et al.* (2006) A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. *Cell*, **124**, 1283-1298.

5. Qi, H.H., Ongusaha, P.P., Myllyharju, J., Cheng, D., Pakkanen, O., Shi, Y., Lee, S.W. and Peng, J. (2008) Prolyl 4-hydroxylation regulates Argonaute 2 stability. *Nature*, **455**, 421-424.
6. Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R. and Salzberg, S.L. (2013) TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol*, **14**, R36.
7. Liao, Y., Smyth, G.K. and Shi, W. (2013) The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res*, **41**, e108.
8. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.
9. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S. *et al.* (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*, **102**, 15545-15550.
10. Mootha, V.K., Lindgren, C.M., Eriksson, K.F., Subramanian, A., Sihag, S., Lehar, J., Puigserver, P., Carlsson, E., Ridderstrale, M., Laurila, E. *et al.* (2003) PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet*, **34**, 267-273.
11. Huang, S., Holzel, M., Knijnenburg, T., Schlicker, A., Roepman, P., McDermott, U., Garnett, M., Grenrum, W., Sun, C., Prahallad, A. *et al.* (2012) MED12 controls the response to multiple cancer drugs through regulation of TGF-beta receptor signaling. *Cell*, **151**, 937-950.
12. Groger, C.J., Grubinger, M., Waldhor, T., Vierlinger, K. and Mikulits, W. (2012) Meta-analysis of gene expression signatures defining the epithelial to mesenchymal transition during cancer progression. *PLoS One*, **7**, e51136.
13. Taube, J.H., Herschkowitz, J.I., Komurov, K., Zhou, A.Y., Gupta, S., Yang, J., Hartwell, K., Onder, T.T., Gupta, P.B., Evans, K.W. *et al.* (2010) Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc Natl Acad Sci U S A*, **107**, 15449-15454.
14. Debnath, J., Muthuswamy, S.K. and Brugge, J.S. (2003) Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods*, **30**, 256-268.

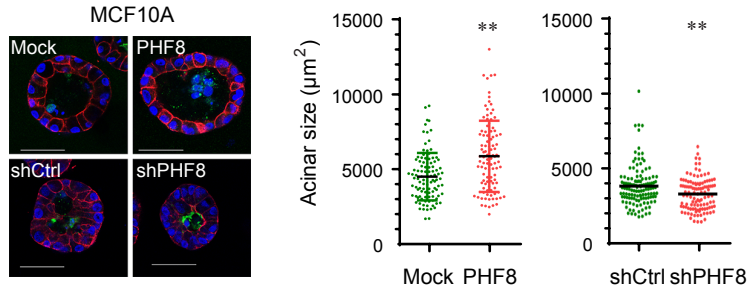
15. Su, S., Liu, Q., Chen, J., Chen, J., Chen, F., He, C., Huang, D., Wu, W., Lin, L., Huang, W. *et al.* (2014) A positive feedback loop between mesenchymal-like cancer cells and macrophages is essential to breast cancer metastasis. *Cancer cell*, **25**, 605-620.
16. Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R. *et al.* (2005) Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res*, **33**, e179.

Supplementary Figure S1

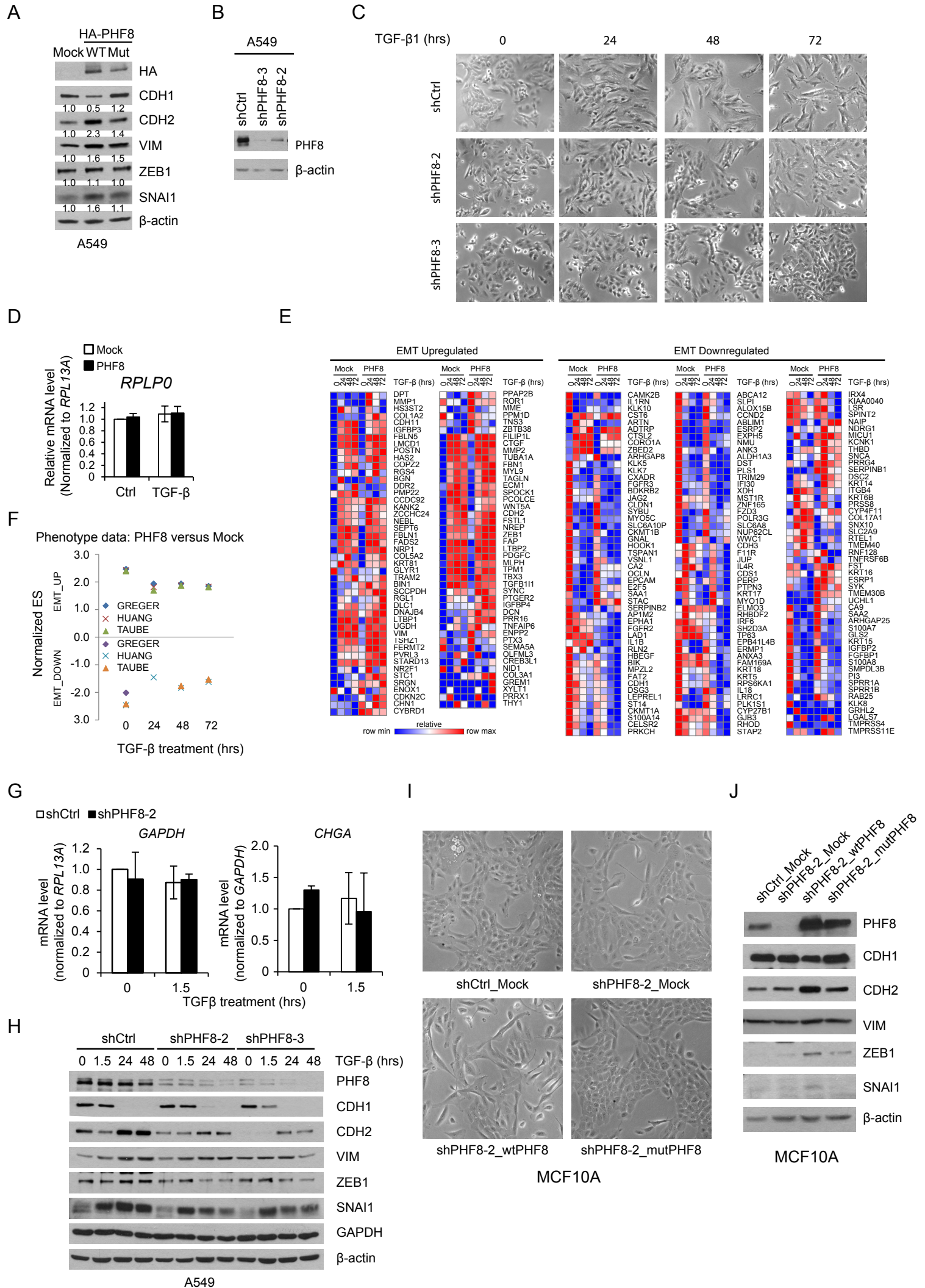
A



B

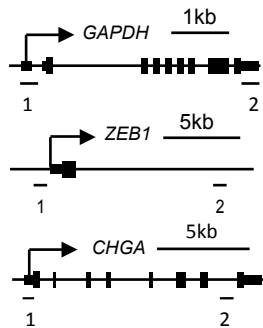


Supplementary Figure S2

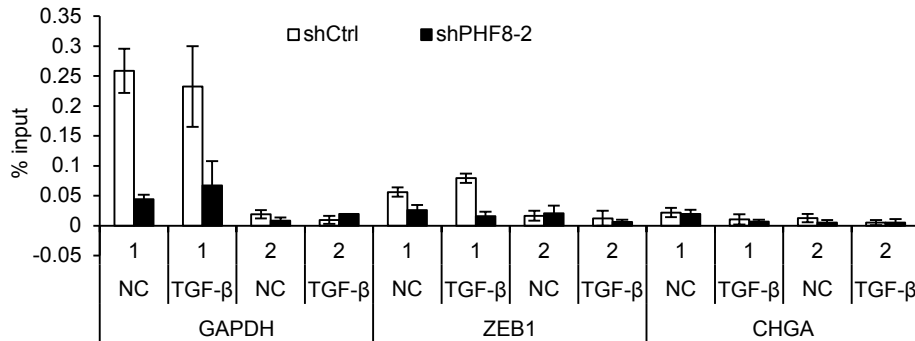


Supplementary Figure S3

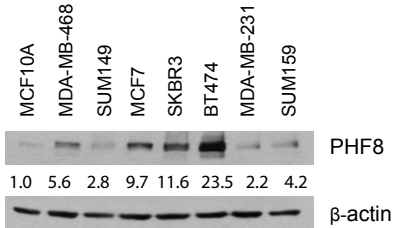
A



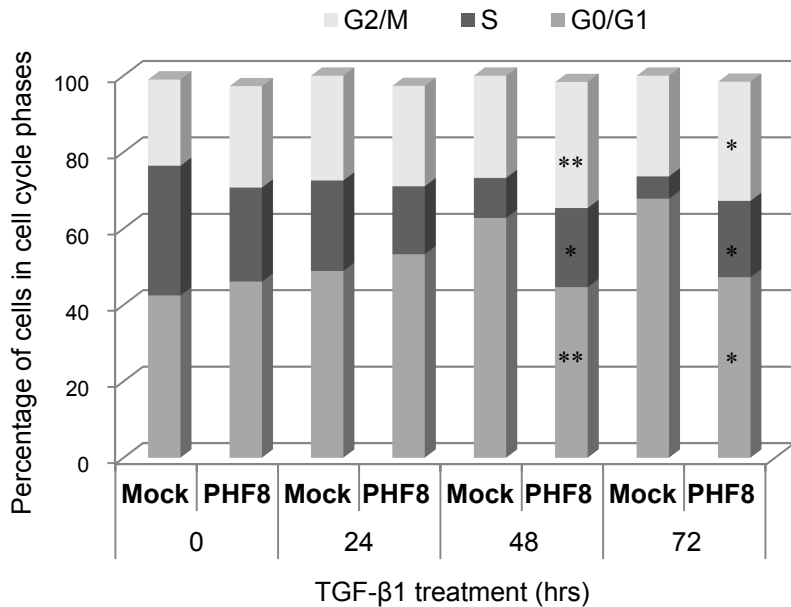
B



Supplementary Figure S4

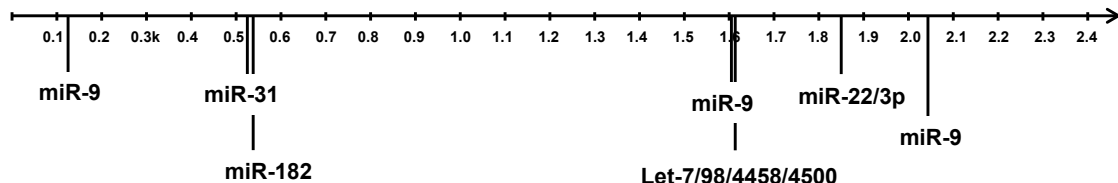


Supplementary Figure S5



Supplementary Figure S6

Human PHF8 mRNA 3' UTR (2478nt, NM_015107)



miR-9 sites

Seed sequence 1: 131-137

5' ...GCUUUCUUGGGACUUACCAAAGG...
 |||||
 3' AGUAUGUCGAUCUAUUGGUUUCU

Seed sequence 2: 1602-1608

5' ...UCUCCAUCUAGUCACACCAAAGG...
 |||||
 3' AGUAUGUCGAUCUAU-UGGUUUCU

Seed sequence 3: 2036-2042

5' ...CAGAGUUAUGCUUUUCCAAAGAG...
 |||||
 3' AGUAUGUCGAUCUAUUGGUUUCU

Let-7a* seed sequence: 1622-1828

5' ...AGGC AAAAGCCUGGCUACCUCC...
 |||||
 3' UUGAUAGUUGGAUGAUGGAGU

miR-22 seed sequence: 1840-1847

5' ...UCUUGAUCUUUCAA-GGCAGCUA...
 |||||
 3' UGUCAAGAAGUUGACCGUCGAA

miR-31 seed sequence: 528-535

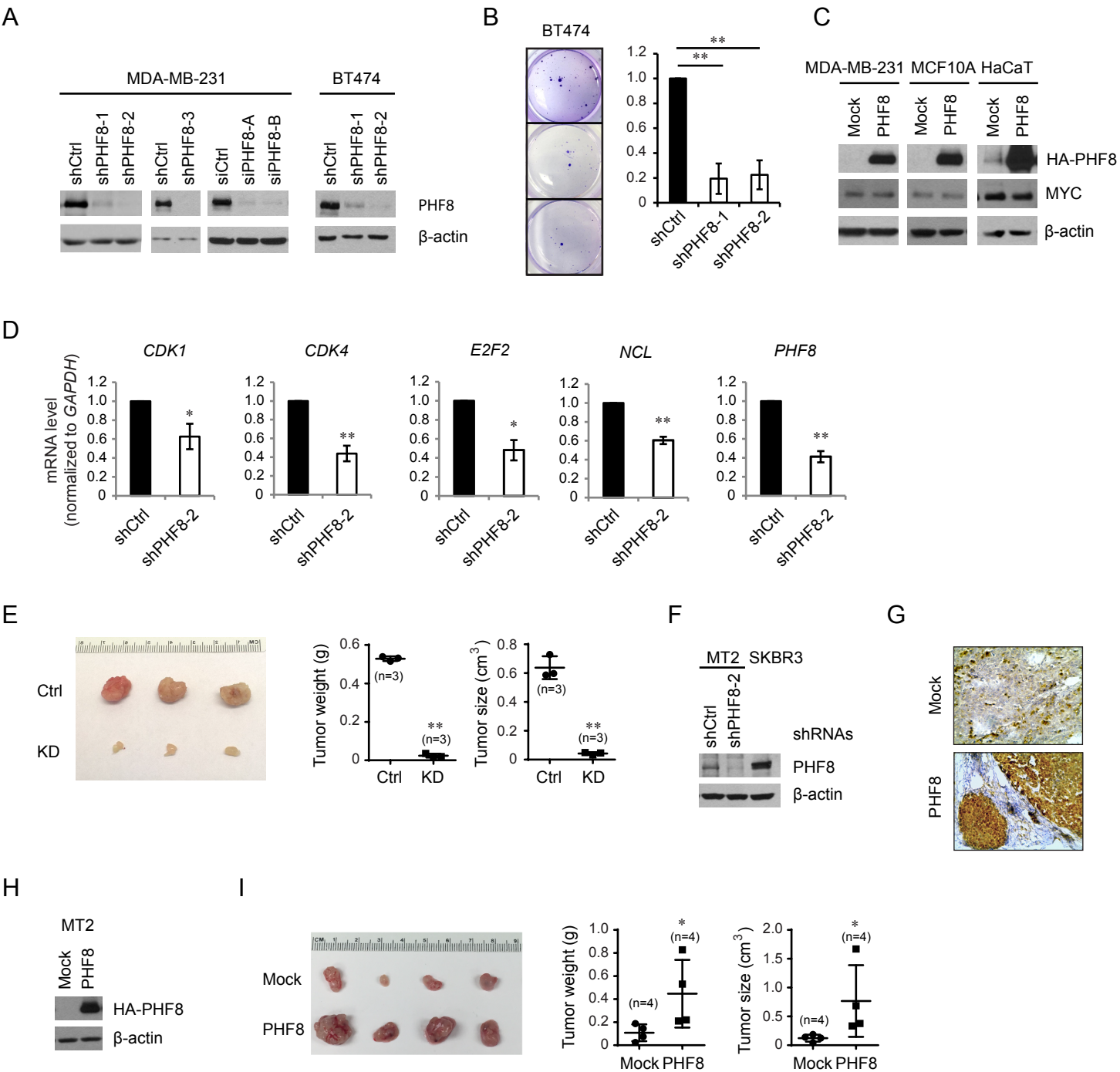
5' ...GUGGUCUCUGCCCC-UCUUGCCA...
 |||||
 3' UCGAUACGGUCGUAGAACGGA

miR-182 seed sequence: 530-537

5' ...GGUCUCUGCCCCCUCUUGCCAAA...
 |||||
 3' UCACACUCAAGAUGGUAACGGUUU

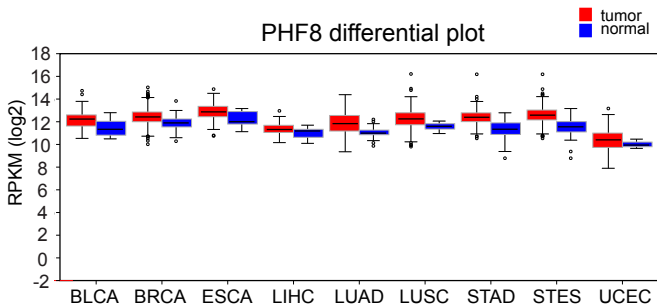
* : This seed sequence is shared by all 8 let-7 isoforms and miR-450, 4458, 98

Supplementary Figure S7



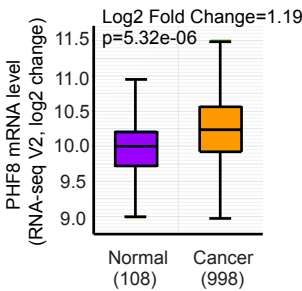
Supplementary Figure S8

A

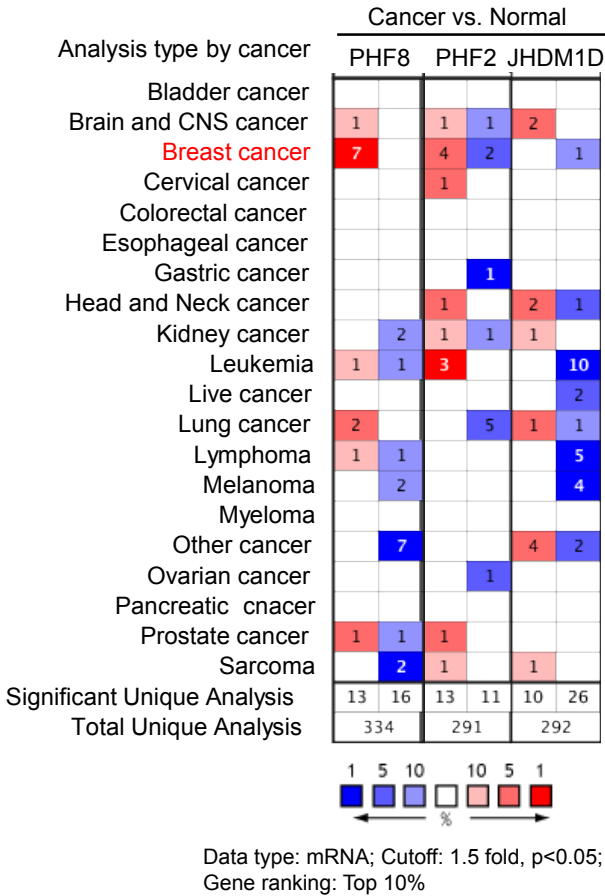


BLCA: Bladder Urothelial Carcinoma
BRCA: Breast invasive carcinoma
ESCA: Esophageal carcinoma
LIHC: Liver hepatocellular carcinoma
LUAD: Lung adenocarcinoma
LUSC: Lung squamous cell carcinoma
STAD: Stomach adenocarcinoma
STES: Stomach and Esophageal carcinoma
UCEC: Uterine Corpus Endometrial Carcinoma

B



C

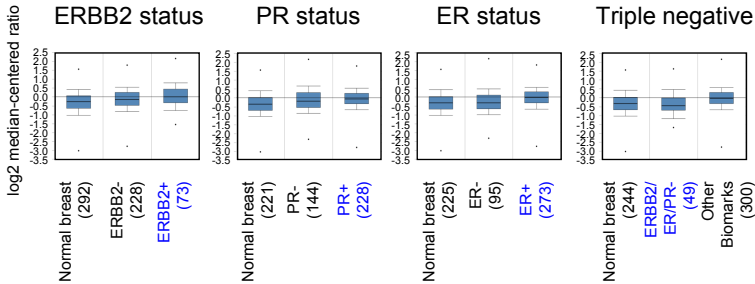


D

Breast cancer (Number of cases)	Gene ranking Top %	Fold change	p value	Reference
Intraductal Cribriform Breast Adenocarcinoma (3)	1%	2.1	1.14E-12	TCGA Breast ¹
Mixed Lobular and Ductal Breast Carcinoma (7)	2%	2.2	7.90E-6	TCGA Breast
Male Breast Carcinoma (3)	3%	1.6	7.51E-7	TCGA Breast
Mucinous Breast Carcinoma (4)	5%	1.6	0.002	TCGA Breast
Invasive Lobular Breast Carcinoma (36)	7%	1.9	3.63E-9	TCGA Breast
Invasive Ductal Breast Carcinoma (389)	11%	2.0	7.15-17	TCGA Breast
Invasive Ductal and Lobular Carcinoma (3)	19%	3.5	0.028	TCGA Breast
Invasive Breast Carcinoma Stroma (53)	2%	2.3	1.30E-19	Finak Breast ²
Invasive Lobular Breast Carcinoma (7)	10%	1.6	0.041	Radvanyi Breast ³
Invasive Breast Carcinoma (154)	21%	1.5	0.019	Gluck Breast ⁴

¹ compared with 61 normal tissues; ² compared with 6 normal tissues; ³ compared to 9 normal tissues; ⁴ compared to 4 normal tissues

E



Supplementary Figure S9

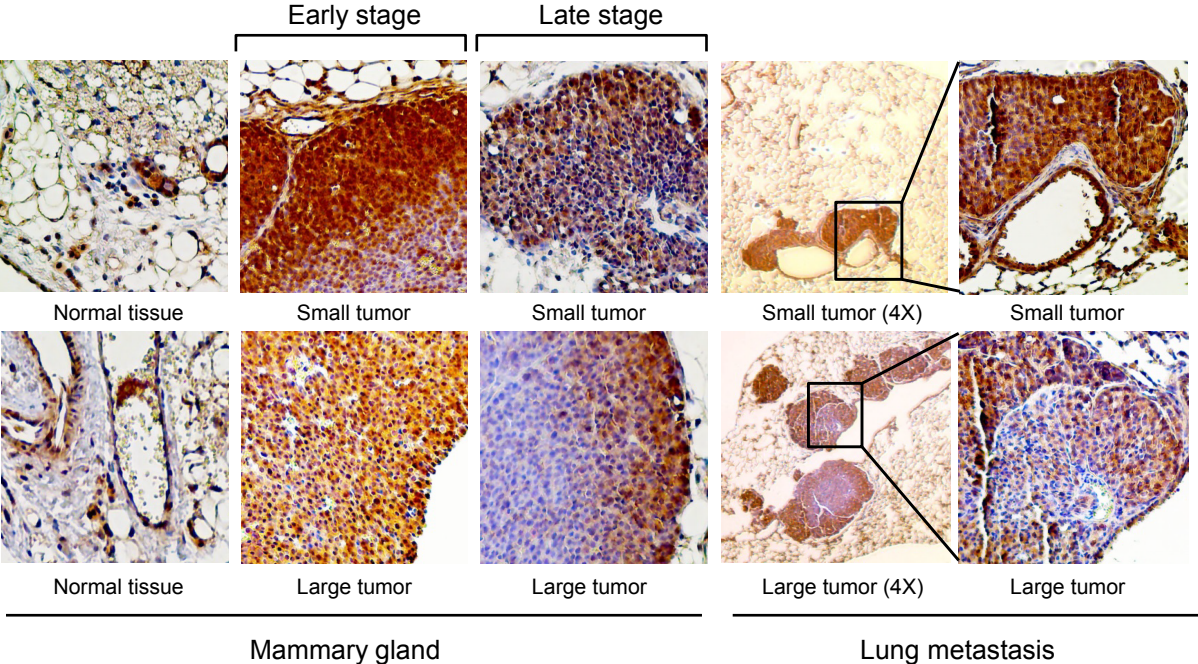


Table S1. Summary of biological duplicated RNA-seq information.

Sample ID	Index	Yield (Mbases)	# Reads	% Align (PF)	% of raw clusters per lane	% of Q30 Bases (PF)	Mean Quality Score (PF)
First round							
MCF10A/Mock-NC	CGATGT	1,477	29,545,742	69.7	11.6	97.3	37.7
MCF10A/Mock-TGFβ1-24hr	TGACCA	1,554	31,084,892	70.8	12.2	97.4	37.8
MCF10A/Mock-TGFβ1-48hr	ACAGTG	1,554	31,086,940	69.9	12.2	97.2	37.7
MCF10A/Mock-TGFβ1-72hr	TAGCTT	1,599	31,985,712	68.4	12.5	97.2	37.7
MCF10A/PHF8-NC	CCGTCC	1,206	24,126,904	62.7	9.4	97.2	37.7
MCF10A/PHF8-TGFβ1-24hr	GTCCGC	1,557	31,148,960	62.6	12.2	97.0	37.6
MCF10A/PHF8-TGFβ1-48hr	GTTTCG	1,799	35,989,074	64.3	14.1	97.3	37.8
MCF10A/PHF8-TGFβ1-72hr	ACTGAT	1,739	34,772,914	63.3	13.6	97.6	37.8
Secondary round							
MCF10A/Mock-NC	CGATGT	1,702	34,031,810	68.0	13.0	97.0	37.6
MCF10A/Mock-TGFβ1-24hr	TGACCA	1,544	30,883,576	68.6	11.8	97.6	37.8
MCF10A/Mock-TGFβ1-48hr	ACAGTG	1,578	31,568,542	70.9	12.1	97.8	37.9
MCF10A/Mock-TGFβ1-72hr	TAGCTT	1,872	37,436,972	70.4	14.3	97.6	37.8
MCF10A/PHF8-NC	CCGTCC	1,366	27,328,602	65.4	10.5	97.6	37.8
MCF10A/PHF8-TGFβ1-24hr	GTCCGC	1,613	32,268,858	67.0	12.3	97.5	37.8
MCF10A/PHF8-TGFβ1-48hr	GTTTCG	1,765	35,293,844	68.5	13.5	97.5	37.8
MCF10A/PHF8-TGFβ1-72hr	ACTGAT	1,381	27,616,924	68.0	10.6	97.9	37.9