Supporting Information

Lee et al. 10.1073/pnas.1300873110

SI Materials and Methods

Dataset and Survival Analysis. Cancer subtype-specific gene expression analyses were performed on the Waddell (GSE21921), TCGA, Kao (GSE20685), and Bos (GSE12276) breast cancer datasets and the Varambally (GSE3325) prostate cancer dataset using Oncomine. Gene expression profiling was performed on a panel of breast cancer cell lines purchased from American Type Culture Collection. Primary breast tumor specimens were obtained from the John Wayne Cancer Institute, using protocols approved by the Institutional Review Board of the National University of Singapore; informed consent was obtained from each individual who provided tumor tissue. Survival analysis was performed on breast cancer datasets GSE12276, GSE1456, and GSE2034 with relevant clinical information, as described previously (1–3). Breast cancer patients were classified into groups with higher and lower expression of the indicated genes based on the mean value of gene expression.

Using survival event status and time information, the survival association of expression status (high/low expression) was computed using Cox proportional hazards model implementation (coxph) available in the R library "survival." Kaplan–Meier survival analysis was used for the analysis of clinical outcomes. A meta-analysis of overall patient survival was performed using datasets from Oxford (GSE6532), Stockholm (GSE1456), and Uppsala (GSE3439) patient cohorts.

Tumor Xenograft and Metastasis Analysis. All animal experiments were performed in accordance with protocols approved by Princeton University's Institutional Animal Care and Use Committee. Four- to 6-wk-old female athymic nude mice (National Cancer Institute) were used to evaluate lung metastatic MB231 (LM2) and sublines, and 4- to 6-wk-old female BALB/c WT mice (National Cancer Institute) were used for all xenograft experiments with mouse mammary tumor cell line 4T1. Lateral tail vein i.v. injection and orthotopic mammary fat pad injection of tumor cells were performed as described previously (4). For luciferase-labeled LM2 cells, development of metastases was monitored by bioluminescent imaging (BLI) with the IVIS Imaging System (Xenogen) and analyzed with Living Image software (Xenogen), as described previously (5). Mammary primary tumor size was measured as described previously (4). Experimental and spontaneous lung metastases were quantified based on visual examination and manual counting of Bouin's solution-fixed lungs. All raw measurements were recorded as mean ± SEM.

Comparisons of primary tumor and BLI lung metastasis growth curves were performed using repeated-measures ANOVA. Survival analysis was performed using the Kaplan–Meier estimator and log-rank test. The relative hazard ratio (HR) between different groups was calculated based on a Cox proportional hazards model and is presented as HR (95% CI). Other comparisons used the unpaired two-sided Student *t* test without the equal variance assumption or the nonparametric Mann–Whitney *U* test. All statistic calculations were performed using Excel (Microsoft), Stata 7 (StataCorp), and EPI Info (Centers for Disease Control and Prevention).

Lung colonization assays of MB231-LN overexpressed with WT and mutant *Ubiquitin-Associated and SH3 Domain-Containing B* (*UBASH3B*) were conducted in compliance with animal protocols approved by the Institutional Animal Care and Use Committee of the Agency for Science, Technology, and Research. For lung colonization assay of MB231-Luc-D3-H2LN (MB231-LN) overexpressed with WT and mutant *UBASH3B*, 2×10^5 cells of MB231-LN control and WT and mutant *UBASH3B* cells were injected via tail veins. Development of lung metastasis was monitored once a week by BLI with the IVIS Imaging System.

Transwell Matrigel Invasion Assay. For this assay, a 24-well Falcon FluorBlok Transwell insert (BD Biosciences) with a pore size of 8 μm was precoated with 500 μg/mL (or 1 mg/mL for prostate cancer cell lines) of growth factor-reduced Matrigel (BD Biosciences) for 3–4 h at 37 °C. Then 5×10^4 MB231, BT549, DU145, and PC3 with indicated treatments were seeded in each insert with DMEM containing 0.5% FBS. DMEM supplemented with 0.5% FBS and 100 ng/mL EGF was added outside the chamber as a chemoattractant. Invaded cells were fixed after 48 h of incubation using 3.7% formaldehyde and stained with 25 μg/mL propidium iodide (Sigma-Aldrich). Ten fields per insert were scanned, and invaded cells were counted using Cellomics ArrayScan.

Three-Dimensional Matrigel Assay. Eight-well chamber slides (BD Biosciences) were precoated with 7.6 mg/mL growth factor-reduced Matrigel for 30 min at 37 °C. Then 5×10^3 MB231 and BT549 with indicated treatments were seeded in each well with DMEM containing 10% (vol/vol) FBS and 150 μg/mL Matrigel.

siRNA and Plasmids. siRNA and plasmid transfections were performed using Lipofectamine RNAiMax (Invitrogen) and FugeneHD (Roche Applied Science), respectively. To generate stable overexpression cell lines, target genes from their respective transient expression plasmids were subcloned into the PMN retroviral expression vector. Virally infected cells were sorted based on GFP overexpression. pcDNA4-*UBASH3B* WT was cloned using cDNA as the template, and pcDNA4-*UBASH3B* H391A was generated using the GeneTailor Site-Directed Mutagenesis Kit (Invitrogen). *UBASH3B* shRNA constructs were generated by inserting siRNA hairpin oligonucleotides into pSIREN-RetroQ (Clontech Laboratories). Target-specific siRNA and nontargeting control siRNA were purchased from 1st BASE with the following target sequences: *UBASH3B* siRNA1: 5′-CCGGCUUAUUUGAGUGGAC-3′; *UBASH3B* siRNA2: CCUCAUAAGAAGCAGCUAC-3′; *UBASH3B* siRNA3: 5′-GCACUGCAACUGAGAAAUU-3′; *CBL* siRNA: 5′-CCAAUCACAAGCUUAGUUAUCAGG-3′. All siRNAs were designed by 1st BASE except for *ETS1* siRNA1, which was referenced from previous work (6). Oligo miRNA mimics and antagomirs were purchased from Dharmacon.

For transfection, 30 nM siRNAs, 15 nM miRNA mimics, or 100 nM iRNA antagomirs were used to transfect cells. At 48 h post-transfection, cell pellets were collected and subjected to the specific assays. Knockdown of mouse *UBASH3B* in 4T1 cells was achieved using lentiviral Mission shRNA constructs (Sigma-Aldrich). The two knockdown constructs were TRCN0000099665 (targeting 3′ UTR; GCTCAGAATCATTTAGCATAT) and TRCN0000099669 [targeting coding region (CDS); GCGTTCAGACTGCACATAATA]. The two control constructs were pLKO1puro empty vector (SHC001) and pLKO1puro non-mammalian shRNA control (SHC002). Virus production, 4T1 cell infection, and puromycin selection were carried out according to the manufacturer's instructions.

Statistics. Where necessary, data were statistically analyzed to generate SE values and to determine the level of significance using the Student *t* test (one-tailed or two-tailed, as appropriate). A P value < 0.05 was considered to indicate significance. Data are reported as mean ± SEM.

Study Approval. For human tissues, informed consent was obtained under a protocol reviewed and approved by the Western Institutional Review Board. All animal experiments were approved by and performed in accordance with guidelines of the Institutional Animal Care and Use Committees at Princeton University and the National University of Singapore.

Cell Culture and Treatment All cell lines were obtained from American Type Culture Collection except for MB231-LN, which was purchased from Bioware. MB231, BT549, MCF7, T47D, BT474, MB361, MB415, MB436, Hs578T, MB157, PC3, DU145, HCT116, and HEK293T cell lines were grown in DMEM supplemented with 10% FBS. VCap cells were grown in the addition of 1% sodium pyruvate and 1% sodium bicarbonate. SKBR3 cells were maintained in McCoy's 5A medium. HCC1806, HCC1937, 22RV1, and LnCap cells were maintained in RPMI medium supplemented with 10% FBS. The MCF10A normal breast epithelial cell line was grown in DMEM/F12 supplemented with 5% (vol/vol) horse serum, 20 ng/mL EGF, 0.5 mg/mL hydrocortisone, 100 ng/mL cholera toxin, 10 μg/mL insulin, and penicillin/streptomycin (Invitrogen). All media were supplemented with 5,000 U/mL penicillin/streptomycin (Invitrogen). All cells were maintained at 37 °C with 5% CO_2.

For Western blot analysis after EGF treatment, cells were treated with 100 ng/mL EGF in DMEM supplemented with 0.5% FBS for 24 h and then harvested for downstream assays.

Microarray Gene Expression Analyses. Total RNA, including small RNAs, was isolated using the RNeasy Mini Kit (Qiagen). Microarray hybridization was performed using the Illumina HumanHT-12 V4.0 expression beadchip, and data analysis was performed using GeneSpring software (Agilent). Reverse-transcription and quantitative PCR (qPCR) assays were performed using the High-Capacity cDNA Archive Kit and Kapa SyBr Fast qPCR Kit (Kapa Biosystems), respectively. TaqMan MicroRNA assays were used to quantify the level of mature miRNAs as described previously (7). In brief, 10 ng of total RNA was reverse-transcribed, and the product was subjected to a TaqMan stemloop miRNA assay (Applied Biosystems). *RNU6B* was used to normalize the data. For quantification of mRNA levels, 18S level was used as an internal control. All reactions were analyzed in an Applied Biosystems 7500 Fast Real-Time PCR system in 96-well plate format, using the following primer sequences:

Gene	Forward primer	Reverse primer
UBASH3B	AGCCCGCGCACAAAAAGCCT	CGGGGCAGGGGGTCATCCAG
UBASH3A	CGGAGTCGTGGGATCAAAGA	GCGTCCCCTGCAATTCTG
ACOT9	TGGTGGATAAGATTGATATGTGTAAGAAG	TCCCGACCCAGCTAACATG
CDK6	GAGTGTTGGCTGCATATTTGCA	GATCAACATCTGAACTTCCACGAA
HYAL3	GGCCCCTATGTGATCAATGTG	ATGGCACCGCTGGTGACT
PSAT1	AATGGAGGTGCCGCGGCCAT	GCCCGGATGCCTCCCACAGAC
B3GNT5	AAGCCGACCTCCGATTTGGACA	CCTTCAGGAAGCGTGGTGGGC
AGPAT4	GTGTGGCTGGAGCCTGTCCG	TCCTGCTCCCACTTGCGCGA

The microarray data reported herein have been deposited at the National Center for Biotechnology Information's Gene Expression Omnibus (accession no. GSE36693).

Western Blot Analysis. Western blot analysis was performed as described previously (8). Anti-UBASH3B (ab34781) was purchased from Abcam. Anti-epidermal growth factor receptor (EGFR) (cs2232), anti-CBL (cs2747), and anti–pan-phosphor-Tyr (cs9411) were purchased from Cell Signaling Technology. Anti-ZEB1 (sc-25388) and anti-ZEB2 (sc-48789) were obtained from Santa Cruz Biotechnology. Anti-Myc and anti-actin were purchased from Roche Applied Science and Sigma-Aldrich, respectively.

Coimmunoprecipitation. Coimmunoprecipitation assays were performed as described previously (8). 293T and BT549 whole-cell lysates were extracted and subjected to immunoprecipitation using protein A/G agarose (Roche) according to the manufacturer's instructions. Anti-Myc, anti-Flag M2 affinity gel (A2220; Sigma-Aldrich), or a nonspecific IgG (sc-2025; Santa Cruz Biotechnology) was used in the coimmunoprecipitation assay.

1. Bos PD, et al. (2009) Genes that mediate breast cancer metastasis to the brain. *Nature* 459(7249):1005–1009.
2. Pawitan Y, et al. (2005) Gene expression profiling spares early breast cancer patients from adjuvant therapy: Derived and validated in two population-based cohorts. *Breast Cancer Res* 7(6):R953–R964.
3. Wang Y, et al. (2005) Gene expression profiles to predict distant metastasis of lymph node-negative primary breast cancer. *Lancet* 365(9460):671–679.
4. Korpal M, et al. (2011) Direct targeting of Sec23a by miR200s influences cancer cell secretome and promotes metastatic colonization. *Nat Med* 17(9):1101–1108.
5. Lu X, et al. (2011) VCAM-1 promotes osteolytic expansion of indolent bone micrometastasis of breast cancer by engaging α4β1-positive osteoclast progenitors. *Cancer Cell* 20(6):701–714.
6. Zhang Y, et al. (2011) miR-125b is methylated and functions as a tumor suppressor by regulating the ETS1 proto-oncogene in human invasive breast cancer. *Cancer Res* 71(10):3552–3562.
7. Feng M, et al. (2011) Myc/miR-378/TOB2/cyclin D1 functional module regulates oncogenic transformation. *Oncogene* 30(19):2242–2251.
8. Lee ST, et al. (2011) Context-specific regulation of NF-κB target gene expression by EZH2 in breast cancers. *Mol Cell* 43(5):798–810.

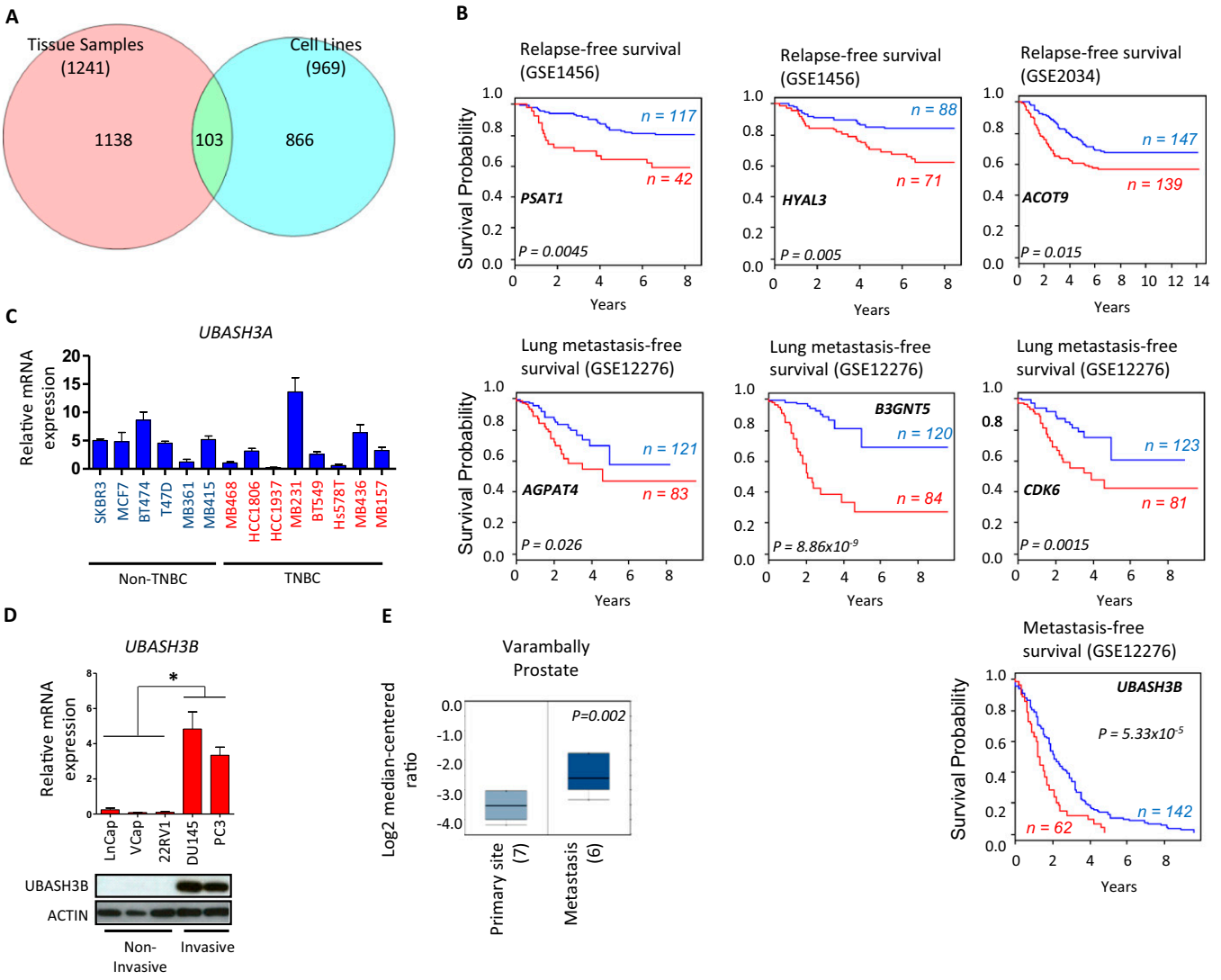

Fig. S1. Functional genomics identifies *UBASH3B* overexpression in triple-negative breast cancer (TNBC), associated with poor prognosis. (*A*) Venn diagram showing the overlap of genes overexpressed in TNBC tissues and cell lines. (*B*) Kaplan–Meier analyses of disease-specific survival of breast cancer patients in the indicated published datasets. Patients with tumors with gene expression exceeding the mean value are labeled in red, and those with tumors with gene expression below the mean are in blue. *P* values were calculated using a Cox proportional hazards regression model and the Wald test. (*C*) qPCR analysis of *UBASH3A* in a panel of breast cancer cell lines. (*D*) qPCR and Western blot analysis of *UBASH3B* expression in a panel of prostate cancer cell lines. $*P < 0.05$, unpaired two-tailed *t* test. (*E*) Boxplots showing mRNA expression of *UBASH3B* in an Oncomine prostate cancer dataset. Error bars represent mean ± SEM.

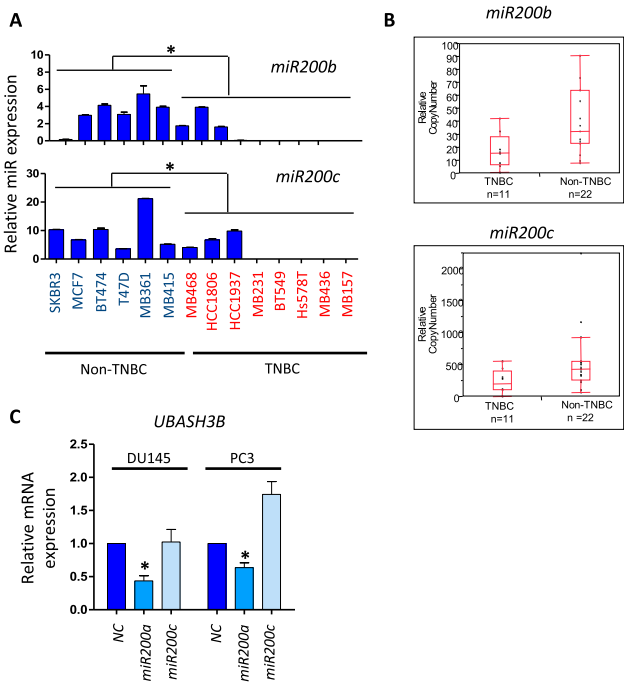

Fig. S2. MicroRNA200s (*miR200*s) are associated with non-TNBC status, but only *miR200a* regulates *UBASH3B* expression. (*A*) qPCR analysis of *miR200b/c* in a panel of breast cancer cell lines. **P* < 0.05, unpaired two-tailed *t* test. (*B*) Boxplot showing *miR200b/c* expression assessed by qPCR in breast cancer specimens. *P* values were calculated by the unpaired two-tailed *t* test. (*C*) qPCR of *UBASH3B* expression in DU145 and PC3 prostate cancer lines overexpressing *miR200a* or *miR200c* mimics. **P* < 0.05, paired two-tailed *t* test. Error bars represent mean ± SEM.

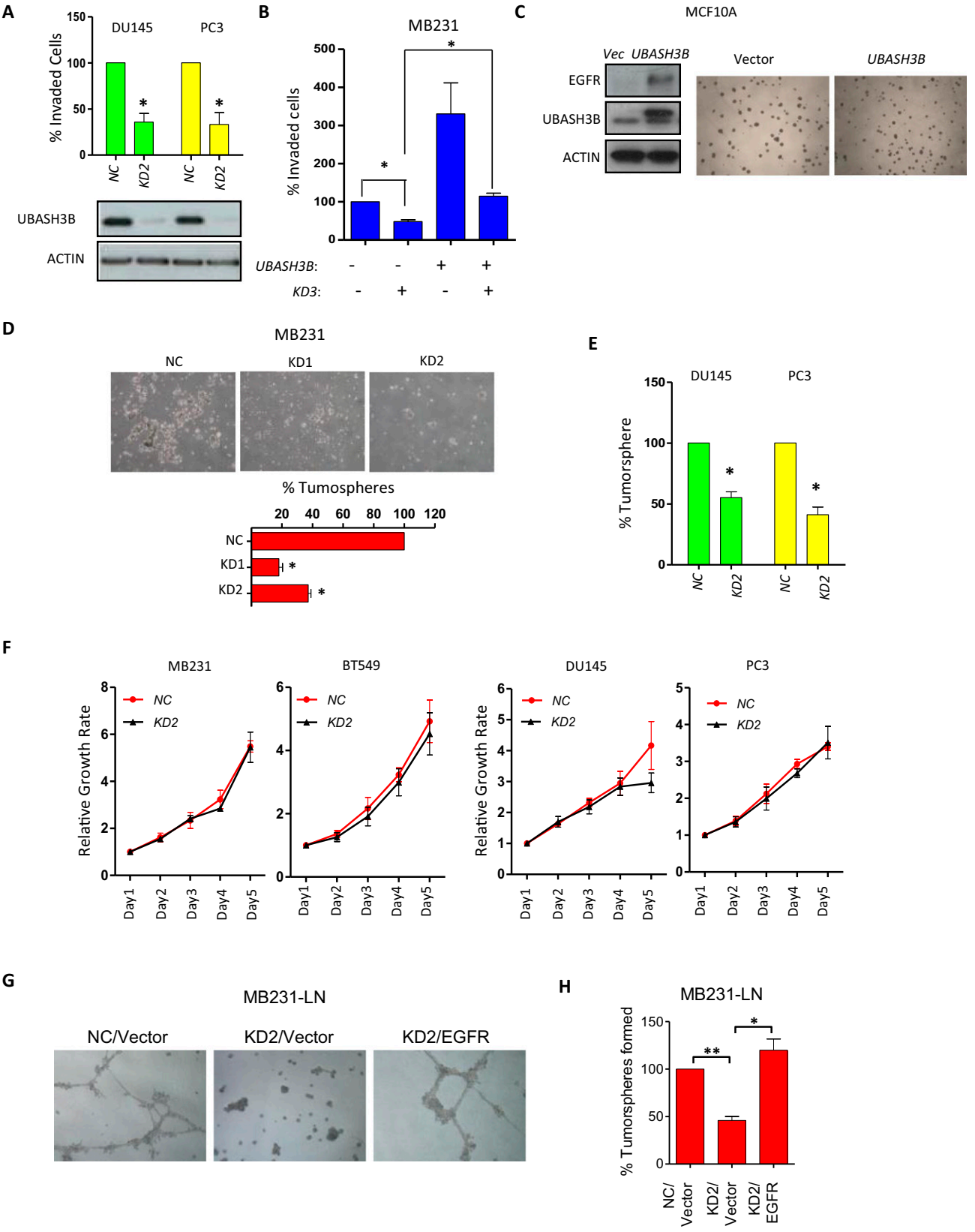

Fig. S3. *UBASH3B* depletion reduces cancer cell aggressiveness. (*A*) (*Upper*) Transwell Matrigel invasion assay on DU145 and PC3 prostate cancer lines in the presence of EGF as a chemoattractant. *$P < 0.05$, paired two-tailed *t* test. (*Lower*) *UBASH3B* knockdown efficiency as assessed by Western blot analysis. (*B*) Transwell Matrigel invasion assay conducted on MB231 cells with stable overexpression of *UBASH3B*, followed by the depletion of endogeneous *UBASH3B* using 3′ UTR siRNA. *$P < 0.05$, paired two-tailed *t* test. (*C*) Western blot analysis of EGFR and UBASH3B expression in normal immortalized breast cell line

Legend continued on following page

MCF10A; 3D Matrilgel growth is shown on the right. (*D*) Tumorsphere assay of MB231 depleted of *UBASH3B* using two independent siRNAs. *$P < 0.05$, paired two-tailed *t* test. Representative images of the tumorspheres are shown. (*E*) Tumorsphere assay on DU145 and PC3 depleted of *UBASH3B*. *$P < 0.05$, paired two-tailed *t* test. (*F*) Cell viability assay on MB231, BT549, DU145, and PC3 depleted with *UBASH3B*. (*G*) Three-dimensional Matrigel assay using MB231-LN cells with *UBASH3B* depletion and EGFR overexpression. (*H*) Tumorsphere assay using MB231-LN cells with *UBASH3B* depletion and *EGFR* overexpression. *$P < 0.05$, paired two-tailed *t* test. Error bars represent mean ± SEM.

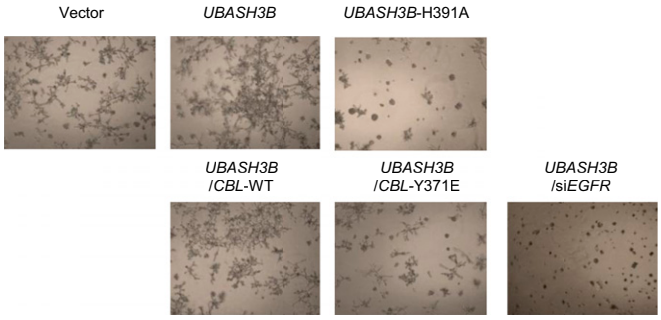

Fig. S4. CBL and EGFR function downstream of UBASH3B in breast cancer cell aggressiveness. Representative images of the 3D Matrigel assay using MB231-LN cells with *UBASH3B* WT and phosphatase dead mutant H391A ectopic overexpression, and combinations of *UBASH3B* overexpression with *CBL* WT, constitutive active ubiquitin E3 ligase mutant Y371E, and EGFR depletion lines.

Table S1. Supplemental data for Fig. 1*A*

ILMN_gene	ILMN_gene	ILMN_gene	ILMN_gene
TOX2	*AGPAT4*	*ZRANB2*	*SFRP1*
SMOX	*ITM2C*	*LOC100132139*	*FAM57A*
CENPV	*HS.571502*	*SLC43A3*	*HS.554507*
CD44	*APEG1*	*SLC43A3*	*FLJ13305*
CDC42BPA	*MAPK12*	*MPP6*	*NXN*
DCBLD1	*ADORA2B*	*LYN*	*ACTN1*
COTL1	*TM4SF1*	*STS-1*	*C11ORF41*
IL8	*SERPINE2*	*UBASH3B*	*CCDC102A*
MAP1B	*UGP2*	*WWTR1*	*MICAL1*
MOXD1	*EGFR*	*B3GNT5*	*CALD1*
CYP26B1	*GSTP1*	*RASAL2*	*PRMT2*
LOC643977	*PVRL3*	*LOC100131541*	*PRICKLE1*
ZCCHC6	*PIM1*	*IGF2BP3*	*HS.184721*
CD70	*ALDH1A3*	*PRSS12*	*LOC100132439*
PCOLCE2	*MFGE8*	*TP53BP2*	*HSD11B1L*
LOC100132564	*PVRL3*	*PELI1*	*HOMER3*
SNORD3D	*ACOT9*	*FOXC1*	*NPAT*
SNORD3A	*CDK6*	*FAM171A1*	*CPA4*
SNORD3C	*HEATR1*	*SERPINB7*	*ARHGAP22*
WNT5B	*CDC42EP1*	*SERPINB7*	*HYAL3*
FXYD5	*SLC25A37*	*EN1*	*CDC20*
FXYD5	*DLL3*	*FSCN1*	*LBR*
TMEM158	*GPR161*	*CHST3*	*PSAT1*
AHNAK2	*RGS20*	*PLOD1*	*LOC729779*
C14ORF78	*ST3GAL6*	*LOC346887*	*SNORA77*
TIAM2	*IGF2BP2*	*AK2*	

Full list of Illumina gene names (ILMN_gene) upregulated in TNBC.