Supplementary Information

Extended Experimental Procedures

Mice

Cre-inducible miR-22 expression construct was generated in which a *LoxP*-flanked tran scriptional STOP element was inserted between a CAGGS promoter and *mir-22*. This c onstruct was targeted into the mouse Collagen A1 locus using Flp recombinase-mediate d genomic integration (Firestein et al., 2008). Mouse embryonic stem cells carrying a sin gle copy of the *mir-22* STOP construct were identified by resistance to the antibiotic mar ker hygromycin and Southern blotting. Selected clones were injected into blastocysts to generate pups. To generate *MMTV-Cre;miR-22*$^{F/+}$ mice, *miR-22*$^{F/+}$ mice were bred to MM TV-Cre transgenic strain (Jackson Laboratory). To generate *MMTV-PyVT;miR-22*$^{F/+}$*;MM TV-Cre* or *MMTV-neu;miR-22*$^{F/+}$*;MMTV-Cre* compound transgenic mice, *MMTV-miR-22*$^{F/+}$ mice were crossed with *MMTV-PyVT* or *MMTV-neu* (unactivated) transgenic mice, res pectively (Jackson Laboratory). Animal experiments were approved by the Animal Care Committee of Harvard Medical School/Beth Israel Deaconess Medical Center. For genot yping, tail DNA was subjected to polymerase chain reaction (PCR) analysis with the foll owing primers. *miR-22*$^{F/+}$: 5'-ATCATGTCTGGATCCCCATC-3', 5'-AAGATCTGAGCAGG GCATGT-3'; *MMTV-Cre*: 5'-GCGGTCTGGCAGTAAAACTATC-3', 5'-GTGAAACAGCATT GCTGCACTT-3'; *MMTV-PyVT*: 5'-GGAAGCAAGTACTTCACAAGGG-3', 5'-GGAAAGT CACTAGGAGCAGGG-3'; *MMTV-neu*: 5'- CGGAACCCACATCAGGCC-3', 5'- TTTCCTG CAGCAGCCTACGC-3'.

Plasmids

PIG-miR-22 retroviral vector encoding human *pri-mir-22* was obtained as described (Poliseno et al., 2010). A 0.5 kb-pair fragment of wild-type 3'UTR of *TET2a* and *TET2b* both encompassing the miR-22 binding sites were obtained by PCR from human genomic DNA and cloned into pGLU vector. The mutant constructs of *TET2a* and *TET2b* 3'UTR were generated by site-directed mutagenesis of 2^{th} and 4^{th} base within seed-matched sequence. To generate the plasmid of miR-22 sponge, a retroviral polym erase II sponge by inserting tandem repeated miR-22 binding sites were cloned into the 3'UTR of a reporter gene encoding destabilized GFP driven by the CMV promoter (Eber t et al., 2007). CXCR4 control sponge plasmid (Addgene) was also used.

Cell culture

The human mammary epithelial cells (HMECs; primary and immortalized) were previously described (Gewinner et al., 2009). MCF-10A, MCF-7, HEK293 and 293T cell lines were obtained from ATCC and grown under cell culture conditions provided by the manufacturer.

Transfection

For transfection of siRNA/miRNAs mimic or microRNA inhibitors, MCF-10A, MCF-7 and HMEC cells (1×10^5) were seeded 16 hrs before transfection. Cells were then transfected with 20 nM siRNAs/microRNA mimic or microRNA inhibitors using Dharmafect transfection reagent according to the manufacturer's instruction. More than 90% of cells were positive to the fluorescent siGLO RISC-free control siRNA with this protocol.

Immunohistochemistry

Tissues were fixed in 4% Paraformaldehyde and wash with PBS and transfer to 70% ethanol, and then embedded in paraffin in accordance with standard procedures. Sections were stained for Hematoxylin and Eosin (H&E), Ki-67 (Novocastra), GFP (Cell Signaling), MECA-32 (BD Biosciences), ERα (Abcam), TTF1 (DAKO), Cytokeratin-8 (Abcam) and Cytokeratin-14 (Novus Biologicals). For immunofluorescence analysis, the sections of tissue were fixed and stained as described (Haffner et al., 2011). After the rehydration of tissue sections, antigen retrieval consisted of steaming for 20 min in 100 mM Tris buffer (pH 9.0) followed by incubation in 3.5 N HCl for 15 min at room temperature (RT). Slides were washed and equilibrated in 0.05% Tween-20/TBS for 10 min. 5-hydroxymethylcytosine specific antibody (Active Motif) was applied at 1:10,000 dilutions. Primary antibodies were incubated for overnight at 4°C. After the extensive washing, the secondary antibodies conjugated with Alexa 488 or Alexa 594 dyes (Molec ular Probe) were incubated for 30min at RT. DNA was counterstained with DAPI. Image s were obtained with confocal microscopy (Zeiss LSM500).

Immunocytochemistry

For immunofluorescence analysis of 5hmC and 5mC, cells were fixed and stained as de scribed (Ko et al., 2010). For staining of cells with antibodies against E-cadherin, Viment in or Fibronectin, cells were fixed with 4% paraformaldehyde for 15 min and permeabiliz ed with 0.05% Triton X-100 in PBS for 15 min at room temperature. Cells were subsequ ently pre-incubated with 1% goat serum/PBS for 30 min at 4°C and incubated with prim ary antibodies for o/n at 4°C. After extensive washing with PBS, Alexa Fluor 488-conjug

ated anti-rabbit IgG (for 5-hmC staining, Molecular Probe) or Alexa Fluor 594-conjugate d anti-mouse IgG (for 5-mC staining, Molecular Probe) were added in blocking buffer for 30min at RT in the dark. DNA was counter-stained with 250 ng/ml of DAPI. Images wer e obtained with confocal microscopy (Zeiss LSM510).

Orthotopic injection

4-weeks old female Nu/Nu immunodeficient mice (Jackson Laboratory) were used for surgery. Mice were anaesthetized with isoflurane. The skin was incised and the endogenous mammary epithelial was removed from the 4^{th} fat pad from the nipple connect to the part distal of the mammary lymph node by electrocauterization before the implantation. Then, MCF-7 or LM2 cells (1×10^5) in 100μl growth medium/matrigel (BD Biosciences) were injected into the rest of mammary fat pad (n=6). For an E2 supplementation, mice have been given injections of β-estradiol (Sigma) dissolved in pure sesame oil (0.1 mg per 0.05 ml sesame oil per mouse, subcutaneously) 1 day before orthotopic injection of MCF-7 cells, and then at weekly intervals until the conclusion of the experiment (Hardman et al., 1999; Kasukabe et al., 2005; Zhou et al., 2008). The recipient mice were monitored and euthanized when the tumors reached 2 cm in diameter, and thus the metastases by MCF-7 cells were evaluated with mice carrying primary tumors of the same size (11–12 weeks post transplantation for MCF7 cells, 8-10 weeks post transplantation for LM2 cells). The primary mammary tumors and lung metastases were analyzed under a dissecting microscope equipped with GFP fluorescence imaging.

Mouse mammary stem cell analysis

Mammary glands were dissected from 7-weeks old female mice. After mechanical dissociation, the cells were prepared as described (Shackleton et al., 2006; Stingl et al., 2006). All reagents were from StemCell Technologies Inc. Cells were stained with biotinylated anti-CD45, -Ter119, -CD31 and -CD140a (eBioscience) for 1hr on ice. After extensive washing with PBS/2%FBS, cells were stained with anti–CD24-FITC, –CD29-PE, –CD49f and Streptavidin-APC-eBluor 780 (eBioscience) for 45 min on ice. For a flow cytometric analysis, DAPI-negative cells were gated and analyzed. Mammary stem cell transplantation assay was performed as previously described (Shimono et al., 2009).

Human patient samples

A series of low- and high-grade breast cancers was assembled from 108 women breast cancer patients. Age at diagnosis of the breast cancer and determination of whether the cancer was a first or subsequent cancer for the patient was determined from medical re cord review. Each cancer was scored for the following pathologic features: histologic typ e; Nottingham combined histologic grade, with each of the three components of grade (t hat is, tubule formation, nuclear grade and mitotic rate) recorded separately; presence o f geographic necrosis or fibrotic focus; extent of lymphocytic infiltrate; and tumor margin characteristics (invasive or pushing). Information regarding ER, PR and HER-2 status, a ssessed as part of the routine clinical evaluation, was abstracted from institutional patho logy reports. This study was approved by the institutional review boards of Dana Farber/ Harvard Cancer Center.

Real-time quantitative PCR

Total RNA was extracted from cells using the RNeasy kit (Qiagen) including DNase

digestion. cDNA was obtained with a First-strand cDNA synthesis kit (USB) or a Taqman microRNA reverse transcript kit (Applied Biosystems). Taqman probes were obtained from Applied Biosystems and selected to avoid detection of genomic DNA. Amplifications were run in a 7900 Real-Time PCR System (Applied Biosystems). Each value was adjusted by using β-actin, PBGD, RNU6B, SNORD44 or SNORD55 as appropriate references.

Glucosylation of genomic 5hmC followed by methylation sensitive qPCR (glucMS-qPCR)

Genomic DNA was treated with T4 Phage β-glucosyltransferase (T4-BGT, New England Biolabs) according to the manufacturer's instruction. Glucosylated genomic DNA (100 ng) was digested with 10 U of HpaII, MspI or no enzyme (mock digestion) at 37°C overnight, followed by inactivation for 20 min at 80°C. The HpaII- or MspI-resistant fraction was quantified by qPCR using primers designed around at least one HpaII/MspI site, and normalizing to the mock digestion control. Resistance to MspI directly translates into percentage of 5hmC, whereas 5mC levels were obtained by subtracting the 5hmC contribution from the total HpaII resistance (Davis and Vaisvila, 2011). Primers used for miR-200c/141 are 5'-AGCCCCTTCGTCTCCCCAGCACCCACTCTC-3', 5'-AAGCCCCTCGACCGTCGGCCCGCCGCCTCT-3'.

Quantitative analysis of 5hmC and 5mC levels using a dot-blot analysis

Genomic DNA was denatured with 0.4 M NaOH, 10 mM EDTA at 95ºC for 10 min, and then neutralized by adding an equal volume of cold 2 M ammonium acetate (pH 7.0).

The DNA was then spotted on nitrocellulose membranes and subjected to Western blotting using antibodies against 5hmC (Active Motif) or 5mC (Eurogentec) as described (Ficz et al., 2011).

Luciferase reporter assay

293T cells were seeded at a density of $1x10^5$ cells per 6-well plate 16 hrs before transfection. 300 ng of luciferase plasmids (pGLU) encoding wild type or mutant 3'UTR of *TET2* were co-transfected with 2 ng of pRL-TK plasmid. The Effectene reagent (Qiagen) was used as transfectant. 36h after transfection, luciferase activity was measured and normalized. Alternatively, cells were co-transfected with 200 ng of luciferase plasmids, 2 ng of pRL-TK and 20 nM of miRNA/the inhibitor of miRNA using Dharmafect 4.

Western blot analysis

Cells were collected and lysed (50mM Tris pH7.5, 1mM EDTA, 1mM $MgCl_2$, 150mM NaCl, 0.5% Triton X100, 1 mM β-glycerophosphate, 1 mM Na_3VO_4, 1 mM NaF, protease inhibitors). Tumor tissues were homogenized in homogenized solution (50mM Tris pH7.5, 1mM EDTA, 1mM $MgCl_2$, 150mM NaCl, 1% NP-40, 1 mM β-glycerophosphate, 1 mM Na_3VO_4, 1 mM NaF, protease inhibitors) and then sonicated. The resulting lysates were subjected to Western blotting with anti-TET2 (Abcam, 1:2000), anti-TET3 (Abcam, 1:500), anti-E-cadherin (Cell signaling, 1:2000), anti-N-cadherin (BD Bioscience, 1:1000), anti-Fibronectin (Sigma, 1:1000), anti-Vimentin (Cell signaling, 1:2000), anti-Zeb1 (Abcam, 1:1000), anti-Zeb2 (Abcam, 1:1000), anti-β-actin (Sigma, 1:20000), anti-HSP90 (BD Bioscience, 1:2000). After washing with 0.05% Tween-20/TBS, membranes

were incubated with HRP-conjugated secondary antibodies and developed using the ECL kit (Amersham) after subsequent washing.

Migration and invasion assay

MCF-10A cells were transfected with the indicated molecules. 12 hrs later, cells were starved in assay medium (Debnath et al., 2003) for additional 12 hrs and then trypsinized, and seeded on the top chambers of 24-well transwell culture inserts (BD Biosciences). Assay medium supplemented with 1 ng/ml EGF was used as chemoattractant. After 24 hrs, cells were fixed in 10% formalin solution for 10 min at room temperature. For invasion assays, BD Biocoat Matrigel invasion chambers (8 μm pores) were used according to the manufacturer's instructions. The non-motile or non-invasive cells on the upper side of the filter were removed, while the motile or invasive cells on the lower side of the filter were stained with crystal violet and counted. 5 fields for each of well were counted under a Nikon Eclipse TE300 microscope. Images were acquired using IPLab software.

Supplementary Figure Legends

Figure S1. Related to Figures 1 and 2.

miR-22 promotes mammary epithelial cell migration, stemness and tumor invasion

(A) MCF-10A cells transfected with control miRNA or miR-22 were subjected to the cell migration assay in the presence or absence of epidermal growth factor (EGF). Representative fields of the migrated cells are shown (left). Scale bars, 100 μm. The migrated cells were also quantified.

(B) Morphology of control miRNA- and miR-22–transfected HMEC cells is shown by a phase contrast. Scale bars, 100 μm.

(C) MCF-7 cells infected with the miR-22 expressing or empty vector were subjected to immunofluorescence analysis with E-cadherin. Scale bars, 20 μm.

(D) MCF-10A cells infected with the miR-22 expressing or empty vector were cultured for 6 days in suspension. Representative images of mammaosphere (>50 μm) are shown. Scale bars, 50 μm.

(E) Cell lysates from the mammospheres of MCF-10A cells infected with the miR-22 expressing or empty vector were subjected to Western blot analysis for the indicated proteins.

(F) GFP (top)- and Ki-67 (bottom)-stained sections of primary mammary tumors formed by MCF-7 cells infected with the miR-22 expressing or empty vector. Scale bars, 100 μm.

(G) *In situ* hybridization analysis of primary mammary tumors formed by MCF-7 cells infected with the miR-22 expressing or empty vector, at 12 weeks after orthotopic transplantation. RNU6B was used as an internal small RNA control. Scale bars, 100 μm.

(H) Mouse mammary epithelial cells obtained from *miR-22$^{F/+}$;MMTV-Cre* mice or *MMTV-Cre* littermate controls subjected to the cell migration assay. Representative fields of the migrated cells are shown (left). Scale bars, 100 μm. The migrated cells were also quantified (right).

(I) Cell lysates from mammary epithelial cells obtained from *miR-22$^{F/+}$;MMTV-Cre* mice or *MMTV-Cre* littermate controls were subjected to Western blot analysis for the indicated proteins.

(J) Mouse MSCs obtained from *miR-22$^{F/+}$;MMTV-Cre* mice or *MMTV-Cre* littermate controls were cultured for 6 days in suspension. Representative images of mammosphere (>50 μm) are shown (left). Scale bars, 50 μm. The number of mammospheres per 1000-plated cells in each culture was also quantified (right). The data are represented as mean ±SD from three independent experiments.

Figure S2. Related to Figure 3.

Analysis of *miR-22$^{F/+}$;MMTV-Cre* transgenic mice

(A) Cytokeratin (CK) 8- and CK14- stained sections of primary mammary tumors from 8-10 months old *miR-22$^{F/+}$;MMTV-Cre* mice. Insets are high magnification of images. Scale bars, 100 μm.

(B and C) H&E-stained sections of primary mammary tumors (left) or lungs (right) isolated from 3-months old *MMTV-PyVT;miR-22$^{F/+}$;MMTV-Cre* mice and *MMTV-PyVT*

littermates (B). Arrows indicate clusters of metastatic cells in the lung. Scale bars, 200 μm. The incidence of metastases to the lung in mice at 3-months old age was also quantified (C).

(D) H&E-stained sections of primary mammary tumors from 12-months old *MMTV-neu;miR-22$^{F/+}$;MMTV-Cre* mice and *MMTV-neu* littermates. Scale bars, 200 μm.

(E) Total RNAs from mammary tumors of *MMTV-neu* or neu*;miR-22$^{F/+}$;MMTV-Cre* were subjected to a real-time qPCR analysis for Zeb1 mRNA.

(F) CK8- and CK14-stained sections of primary mammary tumors of *MMTV-neu;miR-22$^{F/+}$;MMTV-Cre* mice or *MMTV-neu.* N, Normal mammary glands; T, Breast tumors. Scale bars, 100 μm.

(G) Cells isolated from mammary tumors of *MMTV-neu;miR-22$^{F/+}$;MMTV-Cre* mice or *MMTV-neu* were stained with CD90 in the distribution of CD45^{neg}CD31^{neg}CD140a^{neg}Ter119^{neg} in DAPI^{neg} cells. Representative flow cytometric analyses are shown (left) and the quantifications of the analysis are shown (right) (n=3).

Figure S3. Related to Figure 4.

miR-22 controls the level of 5-hydroxymethylcytosine by directly targeting the TET family members

(A) Restoration of miR-200a expression by the treatment with DNA-demethylating agent 5'-aza-2'-deoxycytidine (5'-Aza) in MCF-10A cells infected with the miR-22 expressing or empty vector. The data are represented as mean ±SD from three independent experiments.

(B) Seed sequences of miR-22 on human and mouse *TET* family members were analyzed by using the prediction algorithms (for an example, *http://www.targetscan.org*).

(C and D) Total RNAs isolated from control miRNA- or miR-22–expressing 293T (C) or MCF-10A (D) cells were subjected to a real-time qPCR analysis for TET1, 2 and 3 mRNA.

(E and F) 293T cells were transiently transfected with a combination of a pGL3 firefly luciferase reporter plasmid encoding 3'UTR sequences of human *TET2a* (left) or *TET2b* (right), a Renilla luciferase reporter and miR-22 (E) or the inhibitor of miR-22 (represented as IR-22) (F). Luciferase activities were then measured 48 hrs after transfection. The data are represented as mean ±SD from three independent experiments.

(G) HEK293 cells expressing control miRNA (top) or miR-22 (bottom) were fixed, denatured and stained with antibodies against 5hmC (left) and 5mC (right). Scale bars, 40 μm.

(H) 5hmC- and DAPI-stained sections of the duct of mammary glands isolated from 20-weeks old *Pten*$^{+/-}$ mice or wild-type littermates. Scale bars, 20 μm.

Figure S4. Related to Figures 5 and 6.

TET family members regulates mammary epithelial cell migration, stemness and tumor invasion

(A) miR-22 expression in various human breast cancer cell lines was evaluated by using a real-time qPCR analysis. The data are represented as mean ±SD from three independent experiments.

(B) A GFP sponge by inserting multiple miR-22 binding sites into the 3'UTR of a destabilized *GFP* reporter gene driven by the CMV promoter was generated (top). The imperfect pairing between miR-22 and a sponge with bulged binding sites is diagramed

(bottom).

(C) Real-time qPCR analysis of TET2 (left) or TET3 (right) expression with RNAs from MCF-10A cells infected with the lentiviral vector expressing TET2 or TET3 shRNA. The data are represented as mean ±SD from three independent experiments.

(D) MCF-10A cells expressing TET2 (top) or TET3 (bottom) shRNA were subjected to the cell migration assay. Representative fields of the migrated cells are shown. Scale bars, 100 μm.

(E) Mammospheres derived from MCF-10A cells expressing TET2 (left) or TET3 (right) shRNA were measured. The number of mammospheres per 1000-plated cells in each culture was then quantified. The data are represented as mean ±SD from three independent experiments.

(F) H&E-, Ki67- and ERα-stained sections of lungs isolated from mice that received orthotopic injection of TET2 shRNA-transduced MCF-7 cells, at 12 weeks after transplantation. Scale bars, 50 μm.

(G) Methylation-specific PCR (MSP) analysis of *mir-200c* CpG islands with genomic DNAs purified from MCF-10A cells infected with a combination of the miR-22 and TET2b expressing vector.

(H) LM2 cells infected with a combination of the control or miR-22 sponge and the TET2 or TET3 shRNA expressing vector were subjected to the cell invasion assay. Representative fields of the migrated cells are shown. Scale bars, 100 μm.

Figure S5. Related to Figures 7.

miR-22 is highly expressed and correlates with the TET3-miR-200 axis in non-TNBC patients

(A) miR-22 expression profiling was analyzed from a previously published Illumina Human RefSeq-8 and miRNAv1 array dataset (superSeries GSE22220). Breast tumors were classified by molecular subtypes, estrogen receptor (ER), progesterone receptor (PR), ERBB2 (HER-2) and epithelial clusters.

(B) Anti-correlation between miR-22 and TET3 expression was analyzed using a real-time qPCR analysis with RNAs from non-TNBC patient samples.

(C) Co-expression analysis of TET3 and miR-200a/miR-200c was analyzed by using a real-time qPCR analysis with RNAs from non-TNBC patient samples.

Table S1. Related to Figure 7.

The list of genes upregulated in breast cancer patients highly expressing miR-22

Supplementary References

Davis, T., and Vaisvila, R. (2011). High sensitivity 5-hydroxymethylcytosine detection in Balb/C brain tissue. J. Vis. Exp. pii: 2661. doi: 10.3791/2661.

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.

Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. (2011). Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature *473*, 398-402.

Firestein, R., Blander, G., Michan, S., Oberdoerffer, P., Ogino, S., Campbell, J., Bhimavarapu, A., Luikenhuis, S., de Cabo, R., Fuchs, C.*, et al.* (2008). The SIRT1 deacetylase suppresses intestinal tumorigenesis and colon cancer growth. PloS One *3*, e2020.

Gewinner, C., Wang, Z.C., Richardson, A., Teruya-Feldstein, J., Etemadmoghadam, D., Bowtell, D., Barretina, J., Lin, W.M., Rameh, L., Salmena, L.*, et al.* (2009). Evidence that inositol polyphosphate 4-phosphatase type II is a tumor suppressor that inhibits PI3K signaling. Cancer Cell *16*, 115-125.

Hardman, W.E., Moyer, M.P., and Cameron, I.L. (1999). Fish oil supplementation enhanced CPT-11 (irinotecan) efficacy against MCF7 breast carcinoma xenografts and ameliorated intestinal side-effects. Br. J. Cancer *81*, 440-448.

Kasukabe, T., Okabe-Kado, J., Kato, N., Sassa, T., and Honma, Y. (2005). Effects of combined treatment with rapamycin and cotylenin A, a novel differentiation-inducing

agent, on human breast carcinoma MCF-7 cells and xenografts. Breast Cancer Res. *7*, R1097-R1110.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84-88.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993-997.

Zhou, J., Zhang, H., Gu, P., Bai, J., Margolick, J.B., and Zhang, Y. (2008). NF-kappaB pathway inhibitors preferentially inhibit breast cancer stem-like cells. Breast Cancer Res. Treat. *111*, 419-427.