(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 December 2014 (04.12.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/193964 A2

(51) International Patent Classification:
C12Q 1/68 (2006.01)

(21) International Application Number:
PCT/US2014/039807

(22) International Filing Date:
28 May 2014 (28.05.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
61/828,103 28 May 2013 (28.05.2013) US

(72) Inventors; and
(71) Applicants : **ROSNER, Marsha** [US/US]; 4950 S. Greenwood Ave., Chicago, IL 60615 (US). **SUN, Miao** [US/US]; 929 East 57th Street, Chicago, IL 60637 (US).

(74) Agent: **STIEBEL, Thomas, Jr.**; Quarles & Brady LLP, 300 North LaSalle Street, Suite 4000, Chicago, IL 60654 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2014/193964 A2

(54) Title: PROGNOSTIC AND PREDICTIVE BREAST CANCER SIGNATURE

(57) Abstract: Compositions, kits and methods for treating cancer in a subject in need thereof are disclosed involving one or more upstream activators and/or downstream effectors of TET1.

Prognostic and Predictive Breast Cancer Signature

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Application Serial No. 61/828,103 filed May 28, 2013, which is incorporated herein by reference.

STATEMENT CONCERNING GOVERNMENT INTEREST

[0002] This work was supported by NIH grants GM087630, GM071440, NIH SPORE grant P50 CA125183-05 (DRP), and NIH grant CA127277, and by a DOD Award W81XWH-10-1-0396.

SEQUENCE LISTING

[0003] SEQ ID 1 TET1 Forward (5′–3′): TTATGTAGTTTTATTTGTTTTTTTATTGTG

[0004] SEQ ID 2 TET1 Reverse (5′–3′): CAACTCCAAACCTACACCAAC

[0005] SEQ ID 3 HOXA7 Forward (5′–3′): TATAATTTTGATTTGTGATTTGTTGTT

[0006] SEQ ID 4 HOXA7 Reverse (5′–3′): AAACCTCTTACCCTTCCATTCTAAA

[0007] SEQ ID 5 HOXA9 Forward (5′–3′): TTGGGAATTTTGATTGTTAGTTGA

[0008] SEQ ID 6 HOXA9 Reverse (5′–3′): TACCAAAACACTCCAAACAAAAAC

[0009] SEQ ID 7 TET1 Site-1 Forward (5′–3′): TTTGGGAACCGACTCCTCACCT

[0010] SEQ ID 8 TET1 Site-1 Reverse (5′–3′): TCGGGCAAACTTTCCAACTCGC

[0011] SEQ ID 9 TET1 Site-2 Forward (5′–3′): ACGCTGGGCATTTCTGATCCACTA

[0012] SEQ ID 10 TET1 Site-2 Reverse (5′–3′): TATTGTGCAGCTCGTTTAGTGCCC

[0013] SEQ ID 11 TET1 Site-3 Forward (5′–3′): ACTTTGACCTCCCAAAGTGCTGGA

[0014] SEQ ID 12 TET1 Site-3 Reverse (5′–3′): ACCTGAGTGATGCTGAGACTTCCT

[0015] SEQ ID 13 HOXA7 Site-1 Forward (5′–3′): AAAGCGCGTTCACATAATAC

[0016] SEQ ID 14 HOXA7 Site-1 Reverse (5′–3′):GTTATCATATATCACTCTACCTCGT

[0017] SEQ ID 15 HOXA7 Site-2 Forward (5′–3′): CATTCCTGCTCCGGTTT

[0018] SEQ ID 16 HOXA7 Site-2 Reverse (5′–3′): GGTCCATAAAGGCCGAAG

[0019] SEQ ID 17 HOXA7 Site-3 Forward (5′–3′): CCACCCTGCCTTGTTTCAACATCA

[0020] SEQ ID 18 HOXA7 Site-3 Reverse (5′–3′): ACCAAGTTGTCAGTGAGCCTTCCA

[0021] SEQ ID 19 HOXA9 Site-1 Forward (5′–3′): TTCATCCTCACCAGCAGTTCCAGT

[0022] SEQ ID 20 HOXA9 Site-1 Reverse (5′–3′): GGGCCATTTCGGAGTTCATTGTGT

[0023] SEQ ID 21 HOXA9 Site-2 Forward (5′–3′): CCACCCTGCCTTGTTTCAACATCA

[0024] SEQ ID 22 HOXA9 Site-2 Reverse (5′–3′): ACCAAGTTGTCAGTGAGCCTTCCA

BACKGROUND OF THE INVENTION

1. Field of the Invention

[0025] The present invention relates to the identification and treatment of cancers, including breast cancer.

2. Description of the Background of the Invention

[0026] Epigenetic changes play an important role in cancer progression as well as development. Recent studies indicate that DNA demethylation can be catalyzed by a class of methylcytosine dioxygenases termed the ten–eleven translocation (TET) family. TET1

promotes DNA demethylation by catalyzing conversion of 5-methylcytosine (5mC) primarily to 5-hydroxymethylcytosine (5hmC) as well as 5-formylcytosine or 5-carboxylcytosine. The modified cytosines are then removed through active or passive mechanisms. While TET1 is highly expressed in embryonic stem (ES) cells, loss of TET1 protein and decreased 5hmC levels have been recently shown in solid tumors relative to normal epithelial cells. However, the mechanism by which TET1 is suppressed in solid tumors has not been identified. Furthermore, the downstream targets by which TET1 regulates growth and metastasis in cancer are largely unknown. The TET family of methylcytosine dioxygenases initiates demethylation of DNA and is associated with tumorigenesis in many cancers; however, the mechanism is mostly unknown.

SUMMARY OF THE INVENTION

[0027] According to a first embodiment, a method of diagnosing a subject is provided, which includes analyzing a biological sample from the subject for expression of HMGA2, TET1, HOXA7, and HOXA9; comparing expression levels of the HMGA2, TET1, HOXA7, and HOXA9 in the biological sample to a respective expression reference level of HMGA2, TET1, HOXA7, and HOXA9 in a control sample; and diagnosing the subject with at least one of a decreased survival rate, a poor prognosis, a faster progression of the cancer, and a higher risk of relapse of the cancer if the expression level of HMGA2 in the biological sample is higher than that of the respective control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are lower than that of the respective control sample.

[0028] In another embodiment, the method further includes administering a therapeutically-effective amount of at least one of a chemotherapeutic agent or a radiotherapeutic agent to the diagnosed subject.

[0029] In yet another embodiment, the chemotherapeutic agent is a DNA demethylation agent.

[0030] In another embodiment, the DNA demethylation agent is zacitidine or decitabine, or a combination thereof.

[0031] In still another embodiment, the subject exhibits at least one of an increased chance of survival, a better prognosis, a slower progression of the disease, and a lower risk of relapse of the cancer as compare to the diagnosis before the administering of the therapeutically-effective amount of at least one of the chemotherapeutic agent or the radiotherapeutic agent to the subject.

[0032] In another embodiment, the method further includes diagnosing the subject with at least one of an increase survival rate, a better prognosis, a slower progression of the cancer, and a lower risk of relapse of the cancer if the expression level of HMGA2 in the biological sample is lower than that of the respective control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are higher than the that of the control sample.

[0033] In yet another embodiment, the cancer is breast cancer.

[0034] In one embodiment, the expression levels are determined by quantifying at least one of respective expression of a mRNA encoding HMGA2, TET1, HOXA7, or HOXA9; and a respective quantity of a nucleic acid of at least one of HMGA2, TET1, HOXA7, and HOXA9; or a respective functional fragment or variant thereof.

[0035] In another embodiment, the expression levels are determined immunochemically and based on an antibody-based detection system.

[0036] In yet another embodiment, the antibody binds specifically to a protein of at least one of the HMGA2, TET1, HOXA7, and HOXA9 or a fragment thereof.

[0037] In one embodiment, the control sample is obtained from the subject.

[0038] In another embodiment, the control sample is obtained from a tissue not diagnosed with cancer.

[0039] In yet another embodiment, the HMGA2 expression in the biological sample is greater than about 20% of the expression reference level of HMGA2 in the control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are about 20% lower than the respective expression of that of the control sample.

[0040] In another embodiment, a method of treating cancer (for example, breast cancer) in a subject in need thereof is provided that includes obtaining a biological sample from the subject; measuring expression levels of HMGA2, TET1, HOXA7, and HOXA9 in the biological sample; comparing the expression levels of HMGA2, TET1, HOXA7, and HOXA9 in the biological sample to a respective reference level of HMGA2, TET1, HOXA7, and HOXA9 in a control sample; and administering a therapeutically-effective amount of at least one of a chemotherapeutic agent or a radiotherapeutic agent to the subject when the expression level of HMGA2 in the biological sample is higher than that of the respective control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are lower than that of the control sample.

[0041] In one embodiment, the expression level of HMGA2 in the biological sample is at least 20% higher than that of the respective control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are at least 20% lower than that of the respective expression of the control sample.

[0042] In yet another embodiment, the therapeutically-effective amount of at least one of the chemotherapeutic agent or the radiotherapeutic agent decreases the expression of HMGA2 in the subject.

[0043] In yet another embodiment, the therapeutically-effective amount of at least one of the chemotherapeutic agent or the radiotherapeutic agent increases expression of at least one of TET1, HOXA7, and HOXA9 in the subject.

[0044] In one embodiment, a method of determining a prognosis for survival of a subject diagnosed with cancer is provided that includes analyzing a biological sample from the subject for expression of HMGA2, TET1, HOXA7, and HOXA9; comparing expression levels of the HMGA2, TET1, HOXA7, and HOXA9 in the biological sample to a respective reference level of HMGA2, TET1, HOXA7, and HOXA9 in a control sample; and diagnosing the subject with a poor prognosis if the expression level of HMGA2 in the biological sample is higher than that of the respective control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are lower than the respective expression of the control sample; or a better prognosis if the expression level of HMGA2 in the biological sample is lower than that of the respective control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are higher than the respective expression of the control sample.\

[0045] In yet another embodiment, the method further includes administering to the subject diagnosed with the poor prognosis a therapeutically-effective amount of at least one of a chemotherapeutic agent or a radiotherapeutic agent; and/or a kit comprising at least one agent to detect the expression levels of at least one of HMGA2, TET1, HOXA7, and HOXA9 in the biological sample.

BRIEF DESCRIPTION OF THE FIGURES

[0046] FIG. 1 shows the induction of *TET1* and homeobox gene *(HOX)* gene expression upon depletion of high mobility group AT-hook 2 (HMGA2) in 1833 cells, a bone-tropic derivative of human breast cancer cell line MDA-MB-231, or in *MMTV- Wnt1* transgenic mouse breast tumors. (*A,B,D-H*) 1833 cells were stably transduced with *HMGA2* shRNA (shHMGA2) or control scrambled shRNA (SCR sh): (*A*) Gene expression array analysis showing up-regulation of *TET1* and 20 out of 39 *HOX* genes in HMGA2-depleted cells. The expression levels of *HOXA* genes are also shown in (*B*). *, fold change < 2; **, fold change > 2 based on the signal intensity of gene expression arrays. (***C***) Genomic transcription units of human HOXA genes on chromosome 7 viewed using the UCSC genome browser (39). *HOXA* genes are transcribed from right to left with the order: 5'UTR (thin blue bar), Coding Sequence (thick blue bar) and 3'UTR (thin blue bar). Bar length is proportional to length of DNA sequence. (***D-H***) QRT-PCR and immunoblotting analyses validated induction of *TET1* and *HOXA* gene expression in *HMGA2*-depleted cells: (***D***) *HMGA2*, (***E***) *TET1*, or (***F***) *HOXA4/5/6/7/9/11* mRNA analyzed by qRT-PCR (GAPDH as normalization control); (***G***) HMGA2, TET1 and HOXA9/7 protein analyzed by immunoblotting (GAPDH as control); (***H***) genome-wide 5-hydroxymethylcytosine (5hmC) levels analyzed by dot blot assay. (***I,J***) Loss of *Hmga2* in *MMTV-Wnt1* transgenic mouse breast tumors induced *Tet1* and *Hoxa9/7* expression. *Wnt1* transgenic mice were crossed with *Hmga2* specific knockout mice. Mouse primary breast tumors were obtained from *Hmga2* wildtype (*Hmga2*+/+), heterozygous (*Hmga2*+/-) or null (H*mga2*-/-) mice: (***I***) Murine *Hmga2*, *Tet1* and *Hoxa9/7* mRNA analyzed by qRT-PCR (mouse Gapdh as normalization control); (***J***) Murine TET1 and HOXA9 protein, and 5hmC levels analyzed by immunostaining. (***D-F,H,I***) Data are mean ± s.e., $n = 3$. *, $P < 0.05$; **, $P < 0.01$; ***, $P<0.001$;

[0047] FIG. 2 shows that TET1 involvement in an auto-regulation in human breast cancer cells. (*A*) TET1 binds to its own promoter. 1833 cells expressing TET1 or vector control were analyzed by ChIP assay with anti-TET1 or anti-H3K4Me3 antibody followed by qPCR analysis: TET1 and H3K4Me3 binding to the CpG island proximal to the transcription start site (TSS) of *TET1* (see site-1 and site-2 in Table S5). Site-3 is a negative control. (*B,C*) *HMGA2* depletion causes demethylation of CpG islands at the TET1 promoter region.

1833 cells stably expressing *HMGA2* shRNA (shHMGA2) or control scrambled shRNA (SCR sh) were analyzed for CpG island methylation status by multiple approaches: (***B***) *TET1* promoter region was analyzed within ±1Kb from the TSS. Methylation-specific digestions followed by qPCR distinguished between methylated CpGs versus unmethylated or other modified (e.g. 5hmC) CpGs. The percentage of methylation versus unmethylation (includes unmethylated or other modified C) is indicated; (***C***) Bisulfite sequencing of specific CpGs (see Table S4 for primers) at the *TET1* promoter proximal to the TSS. Results show unmethylated CpGs (open circles) versus methylated or modified CpGs (black circles) in 10 or more independent clones encompassing the region of interest. (***D***) 1833 cells were subjected to 5-azacytidine treatment followed by qRT-PCR analysis for *TET1* mRNA expression (*GAPDH* as normalization control). (***A,B,D***) Data are mean ± s.e., $n = 3$. *, $P < 0.05$; **, $P < 0.01$;

[0048] FIG. 3 illustrates TET1 inducement of *HOXA* gene expression. (***A,B***) Depletion of *TET1* by siRNA partially countered induction of *HOXA* genes. 1833 cells stably expressing *HMGA2* shRNA were transfected with control or *TET1* siRNA: (***A***) Analysis of *TET1* and *HOXA* gene mRNA by qRT- PCR; (***B***) Upper panel: analysis of TET1 and HOXA9/7 protein by immunoblotting; Lower panel: analysis of 5-hydroxymethylcytosine (5hmC) levels by dot blot assay. (***C,D***) Expression of TET1 dramatically induced *HOXA9* expression. 1833 cells expressing constitutive TET1 (Flag-TET1) were analyzed by (***C***) qRT-PCR for *HOXA9* mRNA and by (***D***) Upper panel: immunoblotting for TET1 (Flag-M1) and HOXA9 protein; Lower panel: dot blot assay for 5hmC levels. (***E,F***) Induced expression of TET1 in breast xenograft tumors significantly induced HOXA9 expression. 1833 cells stably expressing an inducible *TET1* expression vector were orthotopically injected into the second mammary fat pad of nude mice. Tumor tissues were collected and analyzed after 6 weeks with (+DOX) or without (-DOX) addition of Doxycycline in the mouse drinking water: (***E***) *TET1* and *HOXA9* mRNA analyzed by qRT-PCR; (***F***) TET1 and HOXA9 protein and 5hmC levels analyzed by immunostaining. (***G***) Significant positive correlation between *TET1* and *HOXA9/7* expression in breast cancer subjects (see Table S3 for patient clinical information). Correlations were determined by Pearson's correlation coefficient. *P* value is determined by Student's t test. (***A-E***) GAPDH as normalization control. Data are mean ± s.e., $n = 3$. **, $P < 0.01$; ***, $P < 0.01$;

[0049] FIG. 4 shows TET1 inducement of *HOXA* gene expression through binding to the promoter regions of *HOXA* genes and contributing to local demethylation in human breast cancer cells. (***A,B***) TET1 binds to the *HOXA* gene promoters. 1833 cells expressing TET1 or control were analyzed by ChIP assay with anti-TET1 or anti-H3K4Me3 antibody followed by qPCR analysis: TET1 and H3K4Me3 binding to the CpG islands proximal to the transcription start site (TSS) of (***A***) *HOXA7* (see site-1 and site-2 in Table S5). Site-3 is a negative control; or (***B***) *HOXA9* (see site-1 in Table S5). Site-2 is a negative control. (***C-E***) HMGA2 depletion causes demethylation of CpG islands at *HOXA* gene promoter regions. 1833 cells stably expressing shHMGA2 or SCR sh were analyzed for CpG island methylation status by multiple approaches (see Figs. 2*B* and 2*C* for the specificity of each method): (***C***) *HOXA* promoter regions were analyzed within -5Kb - +3Kb from the TSS. The percentage of methylation versus unmethylation is indicated; (***D,E***) Bisulfite sequencing of specific CpGs (see Table S4 for primers) at (***D***) *HOXA7* and (***E***) *HOXA9* promoters proximal to the TSS. Results show unmethylated CpGs (open circles) versus methylated or modified CpGs (black circles) in 10 independent clones encompassing the region of interest. (***F,G***) 1833 cells were subjected to 5-azacytidine treatment followed by qRT-PCR analysis for expression of (***F***) *HOXA7* or (***G***) *HOXA9* mRNA (GAPDH as normalization control). (***A-C,F,G***) Data are mean ± s.e., $n = 3$. **, $P < 0.01$; ***, $P < 0.001$;

[0050] FIG. 5 shows both TET1 and its target, HOXA9, in suppression of breast tumor growth, invasion and metastasis. (***A-D***) HMGA2/TET1/HOXA pathway regulates breast cancer cell invasion: (***A***) Inhibition of cell invasion in 1833 cells with depleted *HMGA2* expression; (***B***) Transfection of *TET1* siRNA into *HMGA2*-depleted 1833 cells increases invasion; (***C***) Transfection of HOXA7 or *HOXA9* siRNA into *HMGA2*-depleted 1833 cells increases invasion; (***D***) Decitabine (5-aza- dC) treatment of 1833 cells decreases cell invasion, and transfection of *HOXA9* siRNA into treated cells partially reversed cell invasion. (***A-D***) Data are mean ± s.e., $n = 3$. (***E-K***) 1833 cells stably expressing an inducible control, *TET1* or *HOXA9* expression vector were orthotopically injected into the mammary fat pad of nude mice. Mice were administered drinking water with (+DOX) or without (-DOX) addition of Doxycycline: (***E-G***) Both TET1 and HOXA9 suppress xenograft breast tumor growth: (***E***) Representative bioluminescence images of mice bearing 1833 cells treated as indicated; (***F***) Photograph of representative xenograft breast tumors of 1833 cells treated as indicated; (***G***)

Xenograft breast tumors of 1833 cells treated as indicated and analyzed for tumor weight. (***F,G***) Tumors were dissected at 6 weeks after implantation. (***H,I***) Both TET1 and HOXA9 suppress the proliferation in xenograft breast tumors: immunostaining showing Ki67 positive cells in tumor sample of 1833 cells with induced (+DOX) versus non- induced (-DOX) expression of (***H***) TET1; or (***I***) HOXA9. (***J,K***) Both TET1 and HOXA9 inhibit intravasation of 1833 cells. Cells isolated from the blood after 6 weeks were analyzed for *GAPDH/Gapdh* transcripts derived from human (tumor) or mouse (control) by qRT-PCR: intravasation of 1833 cells with induced (+DOX) versus non-induced (-DOX) expression of (***J***) *TET1*; or (***K***) *HOXA9*. Data are mean ± s.e., $n = 8$ per group. (***L-N***) Both TET1 and HOXA9 suppress bone metastasis of 1833 cells. 1833 cells stably expressing an inducible *TET1* or *HOXA9* expression vector were injected into the left ventricle of mice. Mice were administered drinking water with (+DOX) or without (-DOX) addition of Doxycycline, and imaged for luciferase activity after 3 weeks: (***L***) Representative bioluminescence images of mice with bone metastasis; (**M**) Quantification of bone colonization by 1833 cells with induced (+DOX) versus non-induced (-DOX) expression of *TET1* or *HOXA9*. Data are mean ± s.e.. $n = 7$-9 per group; (***N***) Kaplan-Meier survival analysis of mice over 8 weeks after injection of the tumor cells;

[0051] FIG. 6 shows the HMGA2/TET1/HOXA pathway regulates breast cancer tumorigenesis. (***A***) Comparison of the genes regulated by HMGA2, TET1 or HOXA9 in 1833 cells (human breast cancer cells, hBrCa). (***B***) Scheme illustrating HMGA2/TET1/HOXA signaling pathway in breast tumorigenesis. (***C***) Kaplan-Meier analysis of gene expression data from 101 breast tumor subjects (see Table S3 for patient clinical information). Subjects were stratified for survival using *HMGA2, TET1, HOXA9, HOXA7* or the complete pathway as indicated. Right panel: Red line, high *HMGA2* and low *TET1/HOXAs* (n=34); Blue line, low *HMGA2* and high *TET1/HOXAs* (n=35); *P*, chi-square *p* value;

[0052] FIG. S1 shows a volcano plot for gene expression in 1833 cells stably transduced with either shHMGA2 or control SCR sh. TET1 and HOX gene expression (particularly HOXA4-A9) was significantly increased in HMGA2-depleted 1833 cells. Yellow circles represent 1,012 differentially expressed genes (fold change ≥ 1.5, $P < 0.05$, and FDR < 0.01). Differentially up-regulated HOXA genes are labeled in red symbols; differentially up-regulated

HOXB genes are labeled in blue symbols; differentially up-regulated HOXC genes are labeled in green symbols; differentially up-regulated HOXD genes are labeled in black symbols; and TET1 is labeled in magenta;

[0053] FIG. S2 qRT-PCR and immunoblotting analyses showing induction of TET1 and HOXA gene expression and clinical relevance. MDA-MB-436 cells were stably transduced with HMGA2 shRNA (shHMGA2) or control SCR sh. (A and B) HMGA2 (A) and TET1 and HOXA4/5/6/7/9/11 (B) mRNA analyzed by qRT-PCR (GAPDH as normalization control). (C) HMGA2, TET1, and HOXA9/7 protein analyzed by immunoblotting (GAPDH as control). (D) HMGA2, TET1, and HOXA9 mRNA in MDA-MB-436, MDA-MB-231, or 1833 cells analyzed by qRT-PCR (GAPDH as normalization control). (E–G) Significant negative correlation between HMGA2 and TET1 expression in breast cancer subjects (E; n = 75) including estrogen receptor (ER)-negative (F; n = 20) and ER-positive (G; n = 55) subsets. Correlations were determined by Pearson's correlation coefficient. P value was determined by Student t test. (A–D) Data are means ± SEM; n = 3. *P < 0.05; **P < 0.01;

[0054] FIG. S3 shows loss of Hmga2 in MMTV–Wnt1 transgenic mice suppresses primary breast tumor growth. Wnt1 transgenic mice were crossed with Hmga2-specific knockout mice (SI Materials and Methods). Mouse primary breast tumors were obtained from Hmga2 wild-type (Hmga2+/+) or null (Hmga2−/−) mice. Immunostaining for H&E (Left) and anti-Ki67 (Right) was conducted on those tumor samples;

[0055] FIG. S4 show the effect of HMGA2 expression on TET1 in 1833 and MDA-MB-436 cells. (A–C) We analyzed 1833 cells transfected with HMGA2 lacking the let-7 binding region by qRT-PCR for HMGA2 (A) and TET1 (B) mRNA and by immunoblotting for HMGA2 and TET1 protein (C). (D–F) MDA-MB-436 cells transfected with HMGA2 lacking the let-7 binding region were analyzed by qRT-PCR for HMGA2 (D) and TET1 mRNA (E) and by immunoblotting for HMGA2 and TET1 protein (F). GAPDH was the normalization control for mRNA and protein expression. Data are means ± SEM; n = 3. *P < 0.05; **P < 0.01;

[0056] FIG. S5 shows induction of HOXA9 by a demethylation reagent suppresses cell invasion and proliferation. (A–D) We analyzed 1833 cells subjected to mock (Control) or decitabine treatment for HOXA9 mRNA by qRT-PCR (A), cell invasion from 3-d treatment

(B), HOXA9 protein from 3-d treatment by immunoblotting (C), or cell proliferation (D). (E–H) Inhibition of HOXA9 expression by HOXA9 siRNA after decitabine treatment. We transiently transfected HOXA9 siRNA into 1833 cells treated with decitabine and analyzed for HOXA9 mRNA by qRT-PCR (E), HOXA9 protein by immunoblotting (F), cell invasion (G), or cell proliferation (H). (I–K) Inhibition of HOXA7 or HOXA9 expression by siRNAs. HMGA2-depleted 1833 cells transfected with siRNA for HOXA7 or HOXA9 were analyzed for HOXA7 (I) or HOXA9 mRNA (J) by qRT-PCR or for HOXA7 and HOXA9 protein by immunoblotting (K). (A, E, I, and J) GAPDH as normalization control. (C, F, and K) a-tubulin or GAPDH as control. (A, B, D, E, and G–J) Data are means ± SEM; n = 3. *P < 0.05; **P < 0.01; ***P < 0.001; and

[0057] FIG. S6 shows induced expression of Tet1 or HOXA9 suppresses breast cancer cell growth and enhances survival, and the HMGA2/HOXA pathway stratifies subjects for metastasis-free survival. (A) In vitro analysis of induced TET1 and HOXA9 protein by immunoblotting upon induction of Tet1 expression. (B) In vitro analysis of induced HOXA9 protein by immunoblotting upon induction of HOXA9 expression. (C) Relative growth of cells with induced (+DOX) vs. noninduced (−DOX) Tet1 or HOXA9 expression. Data are means ± SEM; n = 3. *P < 0.05; **P < 0.01. (D and E) Xenograft breast tumors with (+DOX) or without (−DOX) induction of HOXA9 were analyzed for HOXA9 mRNA by qRT-PCR (D) or HOXA9 protein by immunostaining (E). ***P < 0.001. (F–H) We injected 1833 cells stably expressing an inducible control, Tet1, or HOXA9 expression vector into the left ventricle of mice. Mice were administered drinking water with (+DOX) or without (−DOX) addition of doxycycline. Kaplan–Meier survival analysis of mice over 8 wk after injection of the tumor cells (n = 7–9 per group). (F) Induced (+DOX) vs. noninduced (−DOX) control group. (G) Noninduced Tet1 (−DOX) vs. noninduced HOXA9 (−DOX) group. (H) Induced Tet1 (+DOX) vs. induced HOXA9 (+DOX) group. P values were determined by Student t test. (I) Kaplan–Meier analysis of gene expression data from 735 breast tumor subjects. Subjects were stratified for metastasis-free survival by using HMGA2, HOXA9, HOXA7, HOXA4, or the HMGA2/HOXA pathway as indicated. (Right) Red line, high HMGA2 and low HOXAs (n = 69); blue line, low HMGA2 and high HOXAs (n = 72); P, $\chi 2$ P value.

DESCRIPTION

[0058] Treatment of a cancer in a subject in need thereof is provided herein, as are compositions, kits, and methods for treating cancer; methods for identifying genes and pathways having a role in the treatment and prognosis of cancer; therapies to treat and identify cancers based on these genes and pathways; and a method for monitoring the effectiveness of a course of treatment for a subject diagnosed cancer. While the present disclosure may be embodied in different forms, several specific embodiments are discussed herein with the understanding that the present disclosure is to be considered only an exemplification and is not intended to limit the invention to the illustrated embodiments.

[0059] While not wishing to be bound by theory, we believe we have identified not only an important upstream regulator (HMGA2) of TET1, but also a new downstream regulatory pathway for TET1 involving *HOXA* genes; we also believe that TET1 and HOXA9 play an important role not only in breast tumor invasion and growth, but also in metastasis via commonly regulated genes. Because HMGA2 is a genomic architectural factor, we also contemplate that HMGA2 might be able to regulate *TET1* gene expression by direct binding to the *TET1* promoter or alteration of its chromatin structure.

[0060] While this work was under completion, it was reported that TET1 inhibits growth and metastasis in prostate and breast cancer (Winter N, Nimzyk R, Bösche C, Meyer A, Bullerdiek J (2011) Chromatin immunoprecipitation to analyze DNA binding sites of HMGA2. PLoS ONE 6(4):e18837). In that report, TET1 was shown to inhibit invasion in culture in part via tissue inhibitors of metalloproteinases (TIMPs). By contrast, we did not observe significant induction of *TIMP* expression by TET1. Instead, we identified a group of genes commonly altered by *HMGA2* depletion or induction of either *TET1* or *HOXA9*, including a subset of induced genes that promote development and a subset of inhibited genes that promote cell proliferation, consistent with a role for TET1/HOXA9 in suppression of breast tumor growth and metastasis.

[0061] The TET1/HOXA9 signaling pathway we identify here also highlights the importance of cell context in determining the pathological function of TET1. In contrast to our results for breast cancer, the MLL-TET1 fusion protein and the HOXA9 protein both promote

leukemogenesis (See, Hsu CH, et al. (2012) TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases. Cell Rep 2(3):568–579; Ono R, et al. (2002) LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). Cancer Res 62(14):4075–4080). Recently, HOX family members were reported to play key roles in regulating tumorigenesis including the epithelial/mesenchymal transition, invasion and apoptosis (Shah N, Sukumar S (2010) The Hox genes and their roles in oncogenesis. Nat Rev Cancer 10(5):361–371). Highly methylated *HOXA* gene loci have been reported in human breast cancer (Faber J, et al. (2009) HOXA9 is required for survival in human MLL-rearranged acute leukemias. Blood 113(11):2375–2385), although mutations in these genes are not common. Whether these genes function in similar ways or promote different phenotypes is an interesting question that requires further investigation.

[0062] Further, in one embodiment, we have identified a gene signature comprising three mechanistically linked genes (*HMGA2, TET1, HOXA9*) that is prognostic for breast cancer survival and other cancers as well. This signature has the potential to identify subjects harbouring breast or other tumors with suppressed TET1/HOXA9 signaling who might benefit from DNA demethylation agents currently used in the clinic.

[0063] Additionally, it is believed that the high mobility group AT-hook 2 (HMGA2), a chromatin-remodeling factor, binds to AT-rich regions in DNA, altering chromatin architecture to either promote or inhibit the action of transcriptional enhancers. HMGA2 is highly expressed in ES cells but is generally low or lacking in normal somatic cells. Interestingly, HMGA2 is highly expressed in most malignant epithelial tumors, including, for example, breast, pancreas, oral squamous cell carcinoma, and non-small-cell lung cancer. HMGA2 overexpression in transgenic mice causes tumor formation, whereas Hmga2-knockout mice have a pygmy phenotype indicative of a growth defect. We have reported that HMGA2 promotes tumor invasion and metastasis in breast cancer in part through regulation of prometastatic genes, including Snail, osteopontin, and CXCR4. To systematically identify critical downstream mediators of HMGA2 that regulate invasion and metastasis, we performed gene expression array analysis by knocking down HMGA2 in breast cancer cells. Here we show that TET1 is an important effecter of HMGA2 in breast cancer. We further show that TET1 regulates homeobox A (HOXA) genes, including HOXA7 and HOXA9. Both

TET1 and HOXA9 suppress breast tumor growth and metastasis. Our study reveals a regulatory pathway that stratifies subject survival.

[0064] Further, as described more fully herein, we have identified upstream activators and downstream effectors of TET1 in a breast cancer model using human breast cancer cells and a genetically engineered mouse model. We show that depleting the architectural transcription factor HMGA2 induces TET1. TET1 binds and demethylates its own promoter and the promoter of HOXA genes, enhancing its own expression and stimulating expression of HOXA genes including HOXA7 and HOXA9. Both TET1 and HOXA9 suppress breast tumor growth and metastasis in mouse xenografts. The genes comprising the HMGA2- TET1-HOXA9 pathway are believed to be coordinately regulated in breast cancer and together encompass a prognostic signature for subject survival. These results implicate the HMGA2-TET1-HOX signaling pathway in the epigenetic regulation of human breast cancer and highlight the importance of targeting methylation in specific subpopulations as a potential therapeutic strategy.

[0065] From our studies, we provide a method for at least one of: determining the progression of a disease and/or a prognosis for survival of a subject diagnosed with cancer; determining the risk of a relapse of cancer in a subject diagnosed with a cancer, and a treatment regimen for cancer in a diagnosed subject. In one embodiment, the method includes: (a) determining the expression levels of at least one of HMGA2, TET1, HOXA7, and HOXA9 in a biological sample obtained from the subject, and (b) comparing the expression levels of HMGA2, TET1, HOXA7, and HOXA9 in the sample to respective reference levels of HMGA2, TET1, HOXA7, and HOXA9 in a control non-disease state sample. When the expression profile in the subject exhibits at least one of: (a) a higher level of HMGA2 expression, (b) a lower level of TET1 expression, (c) a lower level of HOXA7, and (d) a lower level of HOXA9 expression as compare to the respective reference levels, the expression profile correlates with at least one of: (a) decreased survival, (b) poor prognosis, (c) faster progression of the disease, and (d) higher risk of relapse of the cancer in the subject. In one embodiment, the expression profile in the subject has a higher level of HMGA2 expression, a lower level of TET1 expression, a lower level of HOXA7, and a lower level of HOXA9 expression as compared to the reference levels and correlates with a decreased survival rate, a poorer prognosis, a faster progression of the

disease, and a higher risk of relapse of the cancer in the subject as compare to a subject exhibiting a different expression profile.

[0066] In yet another embodiment, when the expression profile of the subject exhibits at least one of: (a) a lower level of HMGA2 expression, (b) a higher level of TET1 expression, (c) a higher level of HOXA7 expression, and (d) a higher level of HOXA9 expression as compare to the reference levels in a control non-disease state sample, the expression profile correlates with at least one of: (a) increased chance of survival, (b) better prognosis, (c) slower progression of the disease, and (d) lower risk of relapse of the cancer in the subject. In one embodiment, the expression profile in the subject has a lower level of HMGA2 expression, a higher level of TET1 expression, a higher level of HOXA7 expression, and a higher level of HOXA9 expression as compare to the reference levels, the expression profile correlates with an increased chance of survival, a better prognosis, a slower progression of the disease, and a lower risk of relapse of the cancer in the subject. as compare to a subject exhibiting a different expression profile.

[0067] In yet another embodiment, a method for stratifying a subject diagnosed with cancer is provided to determine a therapy regimen for the treatment of the cancer. The method comprises: (a) determining the relative and/or absolute expression levels of at least one of HMGA2, TET1, HOXA7, and HOXA9 in a biological sample obtained from the subject; and (b) comparing the levels obtained in step (a) to the relative and/or absolute expression levels of HMGA2, TET1, HOXA7, and/or HOXA9 in a control non-disease state sample.

[0068] In one embodiment, the relative and/or absolute expression levels of HMGA2, TET1, HOXA7, and/or HOXA9 in the biological sample obtained from the subject are obtained prior to beginning therapy with the subject or obtained in the early stages of therapy of the subject.

[0069] In yet another embodiment, the expression levels of HMGA2, TET1, HOXA7, and/or HOXA9 are determined by quantifying the levels of a functional fragment or variant as is known in the art.

[0070] In yet another embodiment, the expression levels of HMGA2, TET1, HOXA7, and/or HOXA9 are determined using primers readily identified by those skilled in the art including, for example, web-based tools such as Primer3 (Untergrasser A, Cutcutache I,

Koressaar T, Ye J, Faircloth BC, Remm M,Rozen SG (2012)Primer3 - new capabilities and interfaces.*Nucleic Acids Research*40(15):e115; Koressaar T, Remm M (2007) Enhancements and modifications of primer design program Primer3*Bioinformatics* 23(10):1289-91; Source code available at http://sourceforge.net/projects/primer3/). Such primers are also commercially available, including from, for example, Taqman® and Applied Biosystems® from Life Technologies, Inc.

[0071] In another embodiment, the expression levels of HMGA2, TET1, HOXA7, and/or HOXA9 are determined immunochemically, for example, using an antibody-based detection system known to those skilled in the art. In one embodiment, the antibody binds specifically to a protein encoded by the respective gene or a fragment thereof.

[0072] In still another embodiment, the expression levels of HMGA2, TET1, HOXA7, and/or HOXA9 in the subject are determined by quantifying the respective expression of mRNA encoding HMGA2, TET1, HOXA7, and/or HOXA9 as known in art, or quantifying a nucleic acid comprising a sequence determined by those skilled in the art.

[0073] In another embodiment, an increased HMGA2 expression, a lower level of TET1 expression, a lower level of HOXA7, and a lower level of HOXA9 expression in a subject as compared to a control tissue indicates that the subject will likely have decreased survival, poor prognosis, faster progression of the disease, and/or higher risk of relapse, as compared to, for example, a normal non-disease state subject. Such subjects will also benefit from chemotherapy, including a DNA demethylation agent, such as azacitidine or decitabine, or radiotherapy, to treat or alleviate the symptoms and/or severity of the disease. On the other hand, lower level of HMGA2 expression, a higher level of TET1 expression, a higher level of HOXA7 expression, and a higher level of HOXA9 expression in a subject as compared to a control tissue correlates with an increased chance of survival and better prognosis, slower progression of the disease, or lower risk of relapse. In another embodiment, a therapeutic regime to treat a cancer in a subject is determined based on the expression levels of HMGA2, TET1, HOXA7, and/or HOXA9. For example, expression levels are measured before and after a subject is treated with chemotherapy and/or radiotherapy. If a lower level of HMGA2 expression, and a higher level of TET1, HOXA7, and HOXA9 expression are detected in the subject post-treatment, continued dosing of the subject with the chemotherapy or radiotherapy is desirable and/or recommended. If little or

no difference in post-therapy expression levels are detected, alternate therapeutic regimes should be utilized until the subject positively responds to treatment indicated by a decreased level of HMGA2 expression and higher levels of TET1, HOXA7, and/or HOXA9 expression. It is contemplated that the expression level profiles required to effectively treat a particular cancer are subject and/or cancer type and/or stage related. In some instances, a complete shut down (0% expression as compared to a control) in expression of HMGA2 may be required to effectuate treatment, while in other instances, the level may be about 5%, 10%, 20%, 30% 40%, 50%, 60% 70%, 80% or 90% of the control and beneficial results of a therapeutic regime may be seen in a subject. In most circumstance, however, it is believed that a reduction of HMGA2 expression of about 20% will result in the desired therapeutic result. Likewise, a small increase in expression of TET1, HOXA7, and/or HOXA9 may have the required therapeutic effect in certain subjects, while in others it may require an increase of about 5%, 10%, 20%, 30% 40%, 50%, 60% 70%, 80%, 90%, 100% or 200% or more in expression levels of one or more of these genes to show the desired therapeutic effect in the subject. As above, in most circumstance, it is believed that an increase in expression of TET1, HOXA7, and/or HOXA9 of about 20%, respectively, will result in the desired therapeutic result. Therapeutic effects and the prognosis, progression and/or regression of the cancer can be determined by those skilled in the art.

[0074] Expression levels in a subject can be measured in cells of a biological sample obtained from the subject by methods known to those skilled in the art. For example, a tissue sample can be removed from a subject by conventional biopsy techniques. In another example, a body fluid sample, such as a lymph, blood or serum sample, or an exudate fluid sample such as a cancerous organ exudate (for example, exudate from the breast) may be used as the sample. A blood sample can be removed from the subject and white blood cells can be isolated for DNA extraction by standard techniques. The fluid or tissue sample obtained from the subject can be done prior to the initiation of radiotherapy, chemotherapy or other therapeutic treatment. A corresponding control tissue or blood sample can be obtained from unaffected or non-disease state tissues of the subject, from a normal (non-disease or non-cancerous) subject or population of normal subjects, or from cultured cells corresponding to the majority of cells in the subject's sample. The control tissue or blood sample is then processed along with the sample from the subject, so that the levels of expression in cells from the subject's sample can be compared to the corresponding expression levels from cells of the control sample.

[0075] The level of a gene product in a sample can be measured using any technique that is suitable for detecting RNA expression levels in a biological sample. Suitable techniques for determining RNA expression levels in cells from a biological sample are well known to those of skill in the art, including, for example, Northern blot analysis, RT-PCR, in situ hybridization. In one embodiment, the level of gene product is detected using Northern blot analysis. For example, total cellular RNA can be purified from cells by homogenization in the presence of nucleic acid extraction buffer, followed by centrifugation. Nucleic acids are precipitated, and DNA is removed by treatment with DNase and precipitation. The RNA molecules are then separated by gel electrophoresis on agarose gels according to standard techniques, and transferred to nitrocellulose filters. The RNA is then immobilized on the filters by heating. Detection and quantification of specific RNA is accomplished using appropriately labelled DNA or RNA probes complementary to the RNA in question. See, for example, *Molecular Cloning: A Laboratory Manual*, J. Sambrook et al., eds., 2nd edition, Cold Spring Harbor Laboratory Press, 1989, Chapter 7, the entire disclosure of which is incorporated by reference.

[0076] Once the gene expression level of a sample in a subject is measured, the survival, prognosis, progression of the disease, and risk of relapse of the subject can be determined by comparing the gene expression of the sample of the reference control sample (i.e., disease free). As used herein, when the level of expression in the sample is greater than that of the control sample, the expression is termed "up-regulated." When the level of expression in the sample is less than that of the control sample, the expression is termed "down-regulated." In one embodiment, the HMGA2 expression level in the sample is greater than the level of corresponding HMGA2 expression in the control sample, that is, the HMGA2 expression in the sample is "up-regulated." In yet another embodiment, at least one expression level of TET1, HOXA7, and HOXA9 in the sample is less than the level of the corresponding TET1, HOXA7, and HOXA9 expression level in the control sample, that is the expression level of the TET1, HOXA7, and HOXA9 in the sample is "down-regulated." When the HMGA2 expression is up-regulated and the TET1, HOXA7, and/or HOXA9 expression levels are down-regulated in the subject's test sample, the subject will likely experience a decreased survival rate, a poor prognosis, a faster progression of the disease, and/or higher risk of relapse. However, if the HMGA2 expression is down-regulated, and the TET1, HOXA7, and/or HOXA9 expression levels are up-regulated in the subject's test sample, the subject will likely experience an increased

chance of survival a better prognosis, a slower progression of the disease, and/or a lower risk of relapse.

[0077] In one embodiment, a kit is provided to determine the levels of HMGA2, TET1, HOXA7, and/or HOXA9 expression in the sample of a subject. Such a kit may include a reagent for detecting either the DNA encoding HMGA2, TET1, HOXA7, and/or HOXA9, the mRNA encoding HMGA2, TET1, HOXA7, and/or HOXA9, the HMGA2, TET1, HOXA7, and/or HOXA9 polypeptides, or any combination thereof. The reagent may include one or more molecules capable of specifically binding a nucleic acid sequence (DNA or RNA) encoding HMGA2, TET1, HOXA7, and/or HOXA9, or the HMGA2, TET1, HOXA7, and/or HOXA9 polypeptides.

[0078] The kit may include one or more nucleic acid reagents for the detection of either DNA encoding HMGA2, TET1, HOXA7, and/or HOXA9, mRNA encoding HMGA2, TET1, HOXA7, and/or HOXA9, or both. The one or more nucleic acid reagents may be used for hybridization or amplification with the DNA and/or mRNA encoding HMGA2, TET1, HOXA7, and/or HOXA9. The kit may include one or more pairs of primers for amplifying the DNA and/or mRNA encoding HMGA2, TET1, HOXA7, and/or HOXA9. The kit may further include samples of total mRNA derived from tissue of various physiological states, such as normal, and metastatically progressive tumor, for example, to be used as controls. The kit may also include buffers, nucleotide bases, and other compositions to be used in hybridization and/or amplification reactions. Each solution or composition may be contained in a vial or bottle and all vials held in close confinement in a box for commercial sale. Another embodiment of the present invention encompasses a kit for use in detecting the DNA and/or mRNA encoding HMGA2, TET1, HOXA7, and/or HOXA9 in cancer cells in a biological sample that includes oligonucleotide probes effective to bind with high affinity to DNA and/or mRNA encoding HMGA2, TET1, HOXA7, and/or HOXA9 in vitro or in situ and containers for each of these probes.

[0079] In a further embodiment, a kit is provided for use in determining the level of HMGA2, TET1, HOXA7, and/or HOXA9 expression in a biological sample that includes one or more agents, such as, for example, one or more antibodies, specific for one or more HMGA2, TET1, HOXA7, and/or HOXA9 polypeptides. In one particular embodiment, the kit will include

one or more agents and one or more nucleic acid markers wherein the agents and nucleic acid markers are modified in a fashion appropriate for carrying out immuno-polymerase chain reaction assays.

[0080] In another embodiment, a kit is provided for determining a prognosis for survival for a subject with cancer, characterized in that the kit includes compounds capable of detecting the levels of HMGA2, TET1, HOXA7, and/or HOXA9 expression in a biological sample. In a further embodiment, such compounds may be hydrolysis probes targeting any of SEQ ID NO 1-22 for determining HMGA2, TET1, HOXA7, and/or HOXA9 expression. In a further embodiment, the kit may be adapted for RT-PCR and where the kit includes primers amplifying any one or more of SEQ ID NO 1-22. In a further embodiment, such compounds may be one or more antibodies, for example a polyclonal antibody or a monoclonal antibody, wherein the antibody interacts with one of the HMGA2, TET1, HOXA7, or HOXA9 polypeptides. In a further embodiment, a kit is provided for measuring HMGA2, TET1, HOXA7, and/or HOXA9 expression by a DNA, RNA, or protein array.

[0081] One embodiment of the invention is directed to a kit for determining the levels of HMGA2, TET1, HOXA7, and/or HOXA9 expression in a mammalian biological sample, there the levels of HMGA2, TET1, HOXA7, and/or HOXA9 expression is an indicator of the prognosis of a cancer, such as breast cancer. The kit includes: a) one or more antibodies that specifically bind to the HMGA2, TET1, HOXA7, or HOXA9 polypeptides or antigen binding fragments thereof, b) a reagent useful for detecting the extent of interaction between the antibody(ies) and HMGA2, TET1, HOXA7, or HOXA9 polypeptides; c) one or more reagents or solutions useful for antigen retrieval; and c) positive and/or negative control samples. The antibody may be directly linked to an indicator reagent, where the indicator reagent may include one or more fluorescent, colorimetric, immunoperoxidase and isotopic reagents. Alternatively, the kit may further include a second indicator antibody linked to an indicator reagent, where the indicator reagent may include one or more fluorescent, calorimetric, immunoperoxidase and isotopic reagents.

[0082] In one embodiment, the kit contains at least one primary antibody (e.g., anti-HMGA2, TET1, HOXA7, or HOXA9 monoclonal antibodies), at least one labeled secondary antibody (e.g., anti-human HMGA2, TET1, HOXA7, or HOXA9 polyclonal antibodies labeled

with a detection enzyme such as HRP), and at least one substrate (e.g., TMB). Alternatively, the kits can contain radiolabeled secondary antibody in place of the secondary antibody labelled with an enzyme. The kits may also contain disposable supplies for carrying out detection assays (e.g., microtiter plates, pipettes).

[0083] A kit is also provided for use in treating a cancer in a subject, and/or determining prognosis or survival, progression of, or stratification of a subject diagnosed with cancer. The kit may include, for example, compounds capable of detecting the levels of expression levels of HMGA2, TET1, HOXA7, and HOXA9 in a biological sample obtained from a subject diagnosed with cancer; supplies to take a sample from a subject; and/or instructions for us.

[0084] A kit may also include an agent that induces the expression of at least one of TET1 or HOXA9 in the subject and/or an agent that regulates the HMGA2/TET1/HOXA pathway directed at one or more targets as identified by the present disclosure. It is envisioned that a particular kit may be designed for a particular type of cancer and/or a specific tissue. The kit may further include means for administering the agent to a subject in need thereof. In addition, the kit may also include one or more chemotherapeutic and radiotherapeutic agents directed at the specific type of cancer against which the kit is directed.

[0085] Kits may further be a packaged collection of related materials, including, for example, a single and/or a plurality of dosage forms each approximating an therapeutically effective amount of an active ingredient, such as, for example, an expression inhibitor and/or a pharmaceutical compound as described herein that slows, stops, or reverses the growth or proliferation of a tumor or cancer or kills tumor or cancer cells, and/or an additional drug. The included dosage forms may be taken at one time, or at a prescribed interval. Contemplated kits may include any combination of dosage forms.

[0086] Conveniently, HMGA2, TET1, HOXA7, and/or HOXA9 expression may be evaluated using a kit including at least one probe suitable for detecting one or more HMGA2, TET1, HOXA7, and/or HOXA9 markers. As used herein, a probe may include any molecule capable of detecting an HMGA2, TET1, HOXA7, and/or HOXA9 marker, including, but not limited to, monoclonal and polyclonal antibodies and oligonucleotides. For example, the kit may include an antibody specific for an epitope of any of the HMGA2, TET1, HOXA7, and/or HOXA9 proteins encoded by any of the HMGA2, TET1, HOXA7, and/or HOXA9 genes, an

oligonucleotide probe complementary to at least a portion of any of the HMGA2, TET1, HOXA7, and/or HOXA9 genes or to at least a portion an RNA (e.g., mRNA) encoded by any of the HMGA2, TET1, HOXA7, and/or HOXA9 genes, or primer pairs suitable for evaluating HMGA2, TET1, HOXA7, and/or HOXA9 gene expression by a polymerase chain reaction (PCR)-based method, such as real time PCR or reverse transcription PCR. Other methodologies for measuring expression of an HMGA2, TET1, HOXA7, and/or HOXA9 marker may include ribonuclease protection assay, S1 nuclease assay, and Northern blot analysis. Optionally, the kits may include instructions for detecting HMGA2, TET1, HOXA7, and/or HOXA9 detection or for performing the methods of the invention.

[0087] The kit may include a microarray that may be used to determine expression of at least one HMGA2, TET1, HOXA7, and/or HOXA9 marker by a tumor sample and instructions for analyzing the information for use in the methods of the invention. The microarray includes at least one oligonucleotide complementary to a sequence of at least one of SEQ ID NO 1-22. Preferably, the microarray includes a set of oligonucleotides complementary to a set of at least one each of the HMGA2, TET1, HOXA7, and/or HOXA9 sequences selected from SEQ ID NO 1-22. The term "microarray" refers to an ordered arrangement of hybridizable array elements, e.g. oligonucleotide probes, on a substrate, e.g. glass slide or silica. Illustratively, the microarray includes control probes to allow for detection of expression levels that can be used in TSP classifiers to determine HMGA2, TET1, HOXA7, and/or HOXA9 status.

[0088] Although not wishing to be bound by theory, it is believed that the agent depletes transcription factor HMGA2 in the subject when provided in an amount that induces TET1 expression. It is further believed that the TET1 binds and demethylates its own promoter and a promoter of HOXA genes. The TET1 expression and/or HOXA9 expression in the subject is therefore enhanced or stimulated. Illustratively, an amount of the agent administered to the subject is in an amount that suppresses cancer tumor growth or metastasis in the subject as determined by those skilled in the art and described herein. It is also contemplated that such agent(s) can be administered as pharmaceutical compositions in therapeutically effective amounts to subject, and if desired and/or beneficial, in combination with one or more other chemotherapeutic and radiotherapeutic agents, or as part of a kit as described herein. It is further contemplated that a method of treating a cancer in a subject can be provided by administering a

therapeutically effective amount of an agent that at least one of: (a) induces the expression (up-regulates) of at least one of TET1, HOXA7, and HOXA9 in the subject, or (b) down-regulates HMGA2 expression in the subject. In one embodiment, the agent induces the expression TET1 in the subject. In yet another embodiment, the agent induces the expression of HOXA7 in the subject. In still another embodiment, the agent induces the expression of HOXA9 in the subject. In another embodiment the agent down-regulates HMGA2 expression in the subject. In one embodiment, the subject experiences at least one of an increased chance of survival, a better prognosis, a slower progression of the disease, and/or a lower risk of relapse after the agent that resulted in the desired up-regulation of TET1, HOXA7, and/or HOXA9, and/or the desired down-regulation of HMGA2 is administered to the subject. In one embodiment, the prognosis of the subject is determined about 12, 24, 36, 48, 72, 96, 120, or 144 hours after the agent is initially administered to the subject, or about 1, 2, 3, or 4 weeks after the agent is initially administered to the subject.

[0089] Examples of agents useful in the present disclosure that down-regulate HMGA2 inlcude: 1) panobinostat (LBH-589, Novartis, CAS 404950-80-7 (U.S. Patent No. 7,067,551)), and 2) microRNA let-7. Panobinostat is a pan-deacetylase inhibitor and has recently been described by Di Fazio, et al., Exp Cell Res. 2012 Sep 10;318(15):1832-43. doi: 10.1016/j.yexcr.2012.04.018. Epub 2012 Jun 8. Downregulation of HMGA2 by the pan-deacetylase inhibitor panobinostat is dependent on hsa-let-7b expression in liver cancer cell lines. MicroRNA let-7 and has been described by Liu Qi, et al. (See Liu Qi, et al., Histopathology. 2014 Feb 26. doi: 10.1111/his.12401. [Epub ahead of print] HMGA2 is down-regulated by microRNA let-7 and associated with epithelial-mesenchymal transition in oesophageal squamous cell carcinomas of Kazakhs.).

[0090] It is also believed that HOXA9 expression is regulated by several genes, including UTX (Ubiquitously transcribed tetratricopeptide repeat, X chromosome), WHSC1 (Wolf-Hirschhorn syndrome candidate 1), MLL (Myeloid/lymphoid or mixed-lineage leukemia) and MEN1 (Multiple endocrine neoplasia I). It is contemplated that an agent that regulates one or more of these genes that results in the up-regulation of HOXA9 are also useful in the present disclosure. Similarly, it is believed that one or more genes regulate the expression of TET1 and/or HOXA7 and can also be identified by those skilled in the art. Agents that target such

genes that result in upregulation of TET1 and/or HOXA7 are also useful in the present disclosure. Combinations of gene regulatory agents can also be used in the present disclosure.

[0091] The term "subject" refers to any organism classified as a mammal, including mice, rats, guinea pigs, rabbits, dogs, cats, cows, horses, monkeys, and humans, and is used interchangeably with the term "patient" when refereeing to a human subject.

[0092] As used herein, the term "cancer" refers to a class of diseases of mammals characterized by uncontrolled cellular growth. The term "cancer" is used interchangeably with the terms "tumor," "solid tumor," "malignancy," "hyperproliferation" and "neoplasm." Cancer includes all types of hyperproliferative growth, hyperplasic growth, neoplastic growth, cancerous growth or oncogenic processes, metastatic tissues or malignantly transformed cells, tissues, or organs, irrespective of histopathologic type or stage of invasiveness. Illustrative examples include, lung, prostate, head and neck, breast and colorectal cancer, melanomas and gliomas (such as a high grade glioma, including glioblastoma multiforme (GBM), the most common and deadliest of malignant primary brain tumors in adult humans).

[0093] As used herein, the phrase "solid tumor" includes, for example, lung cancer, head and neck cancer, brain cancer, oral cancer, colorectal cancer, breast cancer, prostate cancer, pancreatic cancer, and liver cancer. Other types of solid tumors are named for the particular cells that form them, for example, sarcomas formed from connective tissue cells (for example, bone cartilage, fat), carcinomas formed from epithelial tissue cells (for example, breast, colon, pancreas) and lymphomas formed from lymphatic tissue cells (for example, lymph nodes, spleen, thymus). Treatment of all types of solid tumors regardless of naming convention is within the scope of this invention.

[0094] As used herein, the phrases "chemotherapeutic agent," "cytotoxic agent," "anticancer agent," "antineoplastic agent" and "antitumor agent" are used interchangeably and refer to an agent that has the effect of inhibiting the growth or proliferation, or inducing the killing, of a tumor or cancer cell. The chemotherapeutic agent may inhibit or reverse the development or progression of a tumor or cancer, such as for example, a solid tumor.

[0095] As used herein, the term "chemotherapy" refers to administration of at least one chemotherapeutic agent to a subject having a tumor or cancer.

[0096] An illustrative antineoplastic agent or chemotherapeutic agent includes, a DNA demethylation agent, such as, for example, azacitidine (Vidaza®, Celgene Corporation, CAS 310-6702) or decitabine (Dacogen®, Eisai, Inc., CAS 2353-33-5). Another example is a standard taxane. Taxanes are produced by the plants of the genus *Taxus* and are classified as diterpenes and widely uses as chemotherapy agents including, for example, paclitaxel, (Taxol®, Bristol-Meyers Squibb, CAS 33069-62-4) and docetaxel (Taxotere®, Sanofi-Aventis, CAS 114977-28-5). Other chemotherapeutic agents include semi-synthetic derivatives of a natural taxoid such as cabazitaxel (Jevtana®, Sanofi-Aventis, CAS 183133-96-2). Other chemotherapeutic agents also include an androgen receptor inhibitor or mediator. Illustrative androgen receptor inhibitors include, a steroidal antiandrogen (for example, cyperterone, CAS 2098-66-0); a non-steroidal antiandrogen (for example, flutamide, Eulexin®, Schering-Plough, CAS 13311-84-7); nilutamide (Nilandron®, CAS 63612-50-0); enzalutamide (Xtandi®, Medivation®, CAS 915087-33-1); bicalutamide (Casodex, AstraZeneca, CAS 90357-06-5); a peptide antiandrogen; a small molecule antiandrogen (for example, RU58642 (Roussel-Uclaf SA, CAS 143782-63-2); LG120907 and LG105 (Ligand Pharmaceuticals); RD162 (Medivation, CAS 915087-27-3); BMS-641988 (Bristol-Meyers Squibb, CAS 573738-99-5); and CH5137291(Chugai Pharmaceutical Co. Ltd., CAS 104344603904)); a natural antiandrogen (for example, ataric acid (CAS 4707-47-5) and N-butylbensensulfonamide (CAS 3622-84-2); a selective androgen receptor modulator (for example, enobosarm (Ostarine®, Merck & Company, CAS 841205-47-8); BMS-564,929 (Bristol-Meyer Squibb, CAS 627530-84-1); LGD-4033 (CAS 115910-22-4); AC-262,356 (Acadia Pharmaceuticals); LGD-3303 (Ganolix Lifescience Co., Ltd., 9-chloro-2-ethyl-1-methyl-3-(2,2,2-trifluoroethyl)-3H-pyrrolo[3,2-f]quinolin-7(6H)-one; S-40503, Kaken Pharmaceuticals, 2-[4-(dimethylamino)-6-nitro-1,2,3,4-tetrahydroquinolin-2-yl]-2-methylpropan-1-ol); andarine (GTx-007, S-4, GTX, Inc., CAS 401900-40-1); and S-23 (GTX, Inc., (2S)-N-(4-cyano-3-trifluoromethylphenyl)-3-(3-fluoro-4-chlorophenoxy)-2-hydroxy-2-methyl-propanamide)); or those described in U.S. Patent Appln. No. 2009/0304663. Other neoplastic agents or chemotherapeutic agents that may be used include, for example: alkylating agents such as nitrogen mustards such as mechlorethamine (HN_2), cyclophosphamide, ifosfamide, melphalan (L-sarcolysin) and chlorambucil; ethylenimines and methylmelamines such as hexamethylmelamine, thiotepa; alkyl sulphonates such as busulfan; nitrosoureas such as carmustine (BCNU), lomustine (CCNU), semustine (methyl-CCNU) and streptozocin

(streptozotocin); and triazenes such as decarbazine (DTIC; dimethyltriazenoimidazole-carboxamide); antimetabolites including folic acid analogues such as methotrexate (amethopterin); pyrimidine analogues such as fluorouracil (5-fluorouracil; 5-FU), floxuridine (fluorodeoxyuridine; FUdR) and cytarabine (cytosine arabinoside); and purine analogues and related inhibitors such as mercaptopurine (6-mercaptopurine; 6-MP), thioguanine (6-thioguanine; TG) and pentostatin (2'-deoxycoformycin); natural products including vinca alkaloids such as vinblastine (VLB) and vincristine; epipodophyllotoxins such as etoposide and teniposide; antibiotics such as dactinomycin (actinomycin D), daunorubicin (daunomycin; rubidomycin), doxorubicin, bleomycin, plicamycin (mithramycin) and mitomycin (mitomycin C); enzymes such as L-asparaginase; biological response modifiers such as interferon alphenomes; other agents such as platinum coordination complexes such as cisplatin (cis-DDP) and carboplatin; anthracenedione such as mitoxantrone and anthracycline; substituted urea such as hydroxyurea; methyl hydrazine derivative such as procarbazine (N-methylhydrazine, MTH); adrenocortical suppressant such as mitotane (o,p'-DDD) and aminoglutethimide; taxol analogues/derivatives; hormone agonists/antagonists such as flutamide and tamoxifen; and GnRH and analogues thereof. Examples of other chemotherapeutic can be found in Cancer Principles and Practice of Oncology by V. T. Devita and S. Hellman (editors), 6.sup.th edition (Feb. 15, 2001), Lippincott Williams & Wilkins Publishers.

[0097] As used herein, the term "radiotherapy" refers to administration of at least one "radiotherapeutic agent" to a subject having a tumor or cancer and refers to any manner of treatment of a tumor or cancer with a radiotherapeutic agent. A radiotherapeutic agent includes, for example, ionizing radiation including, for example, external beam radiotherapy, stereotatic radiotherapy, virtual simulation, 3-dimensional conformal radiotherapy, intensity-modulated radiotherapy, ionizing particle therapy and radioisotope therapy.

[0098] Radiotherapy is based on ionizing radiation delivered to a target area that results in death of reproductive tumor cells. Some examples of radiotherapy include the radiation of cesium, palladium, iridium, iodine, or cobalt and is usually delivered as ionizing radiation delivered from a linear accelerator or an isotopic source such as a cobalt source. Also variations on linear accelerators are Cyberkine and Tomotherapy. Particle radiotherapy from cyclotrons such as Protons or Carbon nuclei may be employed. Also radioisotopes delivered systemically

such as p32 or radium 223 may be used. The external radiotherapy may be systemic radiation in the form of sterotacktic radiotherapy total nodal radiotherapy or whole body radiotherapy but is more likely focused to a particular site, such as the location of the tumor or the solid cancer tissues (for example, abdomen, lung, liver, lymph nodes, head, etc.). The radiation dosage regimen is generally defined in terms of Gray or Sieverts time and fractionation, and must be carefully defined by the radiation oncologist. The amount of radiation a subject receives will depend on various consideration but the two important considerations are the location of the tumor in relation to other critical structures or organs of the body, and the extent to which the tumor has spread. One illustrative course of treatment for a subject undergoing radiation therapy is a treatment schedule over a 5 to 8 week period, with a total dose of 50 to 80 Gray (Gy) administered to the subject in a single daily fraction of 1.8 to 2.0 Gy, 5 days a week. A Gy is an abbreviation for Gray and refers to 100 rad of dose.

[0099] Radiotherapy can also include implanting radioactive seeds inside or next to an site designated for radiotherapy and is termed brachytherapy (or internal radiotherapy, endocurietherapy or sealed source therapy). For prostate cancer, there are currently two types of brachytherapy: permanent and temporary. In permanent brachytherapy, radioactive (iodine-125 or palladium-103) seeds are implanted into the prostate gland using an ultrasound for guidance. Illustratively, about 40 to 100 seeds are implanted and the number and placement are generally determined by a computer-generated treatment plan known in the art specific for each subject. Temporary brachytherapy uses a hollow source placed into the prostate gland that is filled with radioactive material (iridium-192) for about 5 to about 15 minutes, for example. Following treatment, the needle and radioactive material are removed. This procedure is repeated two to three times over a course of several days.

[00100] Radiotherapy can also include radiation delivered by external beam radiation therapy (EBRT), including, for example, a linear accelerator (a type of high-powered X-ray machine that produces very powerful photons that penetrate deep into the body); proton beam therapy where photons are derived from a radioactive source such as iridium-192, caesium-137, radium-226 (no longer used clinically), or colbalt-60; Hadron therapy; multi-leaf collimator (MLC); and intensity modulated radiation therapy (IMRT). During this type of therapy, a brief exposure to the radiation is given for a duration of several minutes, and treatment is typically

given once per day, 5 days per week, for about 5 to 8 weeks. No radiation remains in the subject after treatment. There are several ways to deliver EBRT, including, for example, three-dimensional conformal radiation therapy where the beam intensity of each beam is determined by the shape of the tumor. Illustrative dosages used for photon based radiation is measured in Gy, and in an otherwise healthy subject (that is, little or no other disease states present such as high blood pressure, infection, diabetes, etc.) for a solid epithelial tumor ranges from about 60 to about 80 Gy, and for a lymphoma ranges from about 20 to about 40 Gy. Illustrative preventative (adjuvant) doses are typically given at about 45 to about 60 Gy in about 1.8 to about 2 Gy fractions for breast, head, and neck cancers.

[00101] When radiation therapy is a local modality, radiation therapy as a single line of therapy is unlikely to provide a cure for those tumors that have metastasized distantly outside the zone of treatment. Thus, the use of radiation therapy with other modality regimens, including chemotherapy, have important beneficial effects for the treatment of metastasized cancers.

[00102] Radiation therapy has also been combined temporally with chemotherapy to improve the outcome of treatment. There are various terms to describe the temporal relationship of administering radiation therapy and chemotherapy, and the following examples are illustrative treatment regimens and are generally known by those skilled in the art and are provided for illustration only and are not intended to limit the use of other combinations. "Sequential" radiation therapy and chemotherapy refers to the administration of chemotherapy and radiation therapy separately in time in order to allow the separate administration of either chemotherapy or radiation therapy. "Concomitant" radiation therapy and chemotherapy refers to the administration of chemotherapy and radiation therapy on the same day. Finally, "alternating" radiation therapy and chemotherapy refers to the administration of radiation therapy on the days in which chemotherapy would not have been administered if it were given alone.

[00103] It should be noted that other therapeutically effective doses of radiotherapy can be determined by a radiation oncologist skilled in the art and can be based on, for example, whether the subject is receiving chemotherapy, if the radiation is given before or after surgery, the type and/or stage of cancer, the location of the tumor, and the age, weight and general health of the subject.

[00104] Compositions herein may be formulated for oral, rectal, nasal, topical (including buccal and sublingual), transdermal, vaginal, injection/injectable, and/or parental (including subcutaneous, intramuscular, intravenous, and intradermal) administration. Other suitable administration routes are incorporated herein. The compositions may be presented conveniently in unit dosage forms and may be prepared by any methods known in the pharmaceutical arts. Examples of suitable drug formulations and/or forms are discussed in, for example, Hoover, John E. Remington's Pharmaceutical Sciences, Mack Publishing Co., Eston, Pa.; 18.sup.th edition (1995); and Liberman, H. A. and Lachman, L. Eds., Pharmaceutical Dosage Forms, Marcel Decker, New York, N.Y., 1980. Illustrative methods include the step of bringing one or more active ingredients into association with a carrier that constitutes one or more accessory ingredients. In general, the compositions may be prepared by bringing into association uniformly and intimately one or more active ingredients with liquid carriers or finely divided solid carriers or both, and then, if necessary, shaping the product.

[00105] Pharmaceutical formulations may include those suitable for oral, intramuscular, rectal, nasal, topical (including buccal and sub-lingual), vaginal or parenteral (including intramuscular, subcutaneous and intravenous) administration or in a form suitable for administration by inhalation or insufflation. One or more of the compounds of the invention, together with a conventional adjuvant, carrier, or diluent, may thus be placed into the form of pharmaceutical compositions and unit dosages thereof, and in such form may be employed as solids, such as tablets or filled capsules, or liquids such as solutions, suspensions, emulsions, elixirs, or capsules filled with the same, all for oral use, in the form of suppositories for rectal administration; or in the form of sterile injectable solutions for parenteral (including subcutaneous) use. Such pharmaceutical compositions and unit dosage forms thereof may comprise conventional ingredients in conventional proportions, with or without additional active compounds or principles, and such unit dosage forms may contain any suitable effective amount of the active ingredient commensurate with the intended daily dosage range to be employed.

[00106] A salt may be a pharmaceutically suitable (i.e., pharmaceutically acceptable) salt including, but not limited to, acid addition salts formed by mixing a solution of the instant compound with a solution of a pharmaceutically acceptable acid. A pharmaceutically acceptable acid may be, for example, hydrochloric acid, methanesulphonic acid, fumaric acid, maleic acid,

succinic acid, acetic acid, benzoic acid, oxalic acid, citric acid, tartaric acid, carbonic acid or phosphoric acid.

[00107] Suitable pharmaceutically-acceptable salts may further include, but are not limited to salts of pharmaceutically-acceptable inorganic acids, including, for example, sulfuric, phosphoric, nitric, carbonic, boric, sulfamic, and hydrobromic acids, or salts of pharmaceutically-acceptable organic acids such propionic, butyric, maleic, hydroxymaleic, lactic, mucic, gluconic, benzoic, succinic, phenylacetic, toluenesulfonic, benezenesulfonic, salicyclic sulfanilic, aspartic, glutamic, edetic, stearic, palmitic, oleic, lauric, pantothenic, tannic, ascorbic, and valeric acids.

[00108] Various pharmaceutically acceptable salts include, for example, the list of FDA-approved commercially marketed salts including acetate, benzenesulfonate, benzoate, bicarbonate, bitartrate, bromide, calcium edetate, camsylate, carbonate, chloride, citrate, dihydrochloride, edetate, edisylate, estolate, esylate, fumarate, gluceptate, gluconate, glutamate, glycollylarsanilate, hexylresorcinate, hydrabamine, hydrobromide, hydrochloride, hydroxynaphthoate, iodide, isethionate, lactate, lactobionate, malate, maleate, mandelate, mesylate, methylbromide, methylnitrate, methylsulfate, mucate, napsylate, mitrate, pamoate, pantothenate, phosphate, diphosphate, polygalacturonate, salicylate, stearate, subacetate, succinate, sulfate, tannate, tartrate, teoclate, and triethiodide.

[00109] A hydrate may be a pharmaceutically suitable (i.e., pharmaceutically acceptable) hydrate that is a compound formed by the addition of water or its elements to a host molecule (for example, the free form version of the compound) including, but not limited to, monohydrates, dihydrates, etc. A solvate may be a pharmaceutically suitable (i.e., pharmaceutically acceptable) solvate, whereby solvation is an interaction of a solute with a solvent that leads to stabilization of the solute species in a solution, and whereby the solvated state is an ion in a solution complexed by solvent molecules. Solvates and hydrates may also be referred to as "analogues."

[00110] A prodrug may be a compound that is pharmacologically inert but is converted by enzyme or chemical action to an active form of the drug (i.e., an active pharmaceutical ingredient) at or near the predetermined target site. In other words, prodrugs are inactive compounds or partially active compounds that yield an active compound upon metabolism in the

body, which may or may not be enzymatically controlled. Prodrugs may also be broadly classified into two groups: bioprecursor and carrier prodrugs. Prodrugs may also be subclassified according to the nature of their action. Bioprecursor prodrugs are compounds that already contain the embryo of the active species within their structure, whereby the active species are produced upon metabolism.

[00111] Carrier prodrugs are formed by combining the active drug (e.g., active ingredient) with a carrier species forming a compound having desirable chemical and biological characteristics, whereby the link is an ester or amide so that the carrier prodrug is easily metabolized upon absorption or delivery to the target site. For example, lipophilic moieties may be incorporated to improve transport through membranes. Carrier prodrugs linked by a functional group to carrier are referred to as bipartite prodrugs. Prodrugs where the carrier is linked to the drug by a separate structure are referred to as tripartite prodrugs, whereby the carrier is removed by an enzyme-controlled metabolic process, and whereby the linking structure is removed by an enzyme system or by a chemical reaction. A hydroxy-protecting group includes, for example, a tert-butyloxy-carbonyl (t-BOC) and t-butyl-dimethyl-silyl (TBS). Other hydroxy protecting groups contemplated are known in the art.

[00112] In another embodiment, a dosage form and/or composition may include one or more active metabolites of the active ingredients in place of or in addition to the active ingredients disclosed herein.

[00113] Dosage form compositions containing the active ingredients may also contain one or more inactive pharmaceutical ingredients such as diluents, solubilizers, alcohols, binders, controlled release polymers, enteric polymers, disintegrants, excipients, colorants, flavorants, sweeteners, antioxidants, preservatives, pigments, additives, fillers, suspension agents, surfactants (for example, anionic, cationic, amphoteric and nonionic), and the like. Various FDA-approved topical inactive ingredients are found at the FDA's "The Inactive Ingredients Database" that contains inactive ingredients specifically intended as such by the manufacturer, whereby inactive ingredients can also be considered active ingredients under certain circumstances, according to the definition of an active ingredient given in 21 CFR 210.3(b)(7). Alcohol is a good example of an ingredient that may be considered either active or inactive depending on the product formulation.

[00114] As used herein, an oral dosage form may include capsules (a solid oral dosage form consisting of a shell and a filling, whereby the shell is composed of a single sealed enclosure, or two halves that fit together and which are sometimes sealed with a band and whereby capsule shells may be made from gelatin, starch, or cellulose, or other suitable materials, may be soft or hard, and are filled with solid or liquid ingredients that can be poured or squeezed), capsule or coated pellets (solid dosage form in which the drug is enclosed within either a hard or soft soluble container or "shell" made from a suitable form of gelatin; the drug itself is in the form of granules to which varying amounts of coating have been applied), capsule coated extended release (a solid dosage form in which the drug is enclosed within either a hard or soft soluble container or "shell" made from a suitable form of gelatin; additionally, the capsule is covered in a designated coating, and which releases a drug or drugs in such a manner to allow at least a reduction in dosing frequency as compared to that drug or drugs presented as a conventional dosage form), capsule delayed release (a solid dosage form in which the drug is enclosed within either a hard or soft soluble container made from a suitable form of gelatin, and which releases a drug (or drugs) at a time other than promptly after administration, whereby enteric-coated articles are delayed release dosage forms), capsule delayed release pellets (solid dosage form in which the drug is enclosed within either a hard or soft soluble container or "shell" made from a suitable form of gelatin); the drug itself is in the form of granules to which enteric coating has been applied, thus delaying release of the drug until its passage into the intestines), capsule extended release (a solid dosage form in which the drug is enclosed within either a hard or soft soluble container made from a suitable form of gelatin, and which releases a drug or drugs in such a manner to allow a reduction in dosing frequency as compared to that drug or drugs presented as a conventional dosage form), capsule film-coated extended release (a solid dosage form in which the drug is enclosed within either a hard or soft soluble container or "shell" made from a suitable form of gelatin; additionally, the capsule is covered in a designated film coating, and which releases a drug or drugs in such a manner to allow at least a reduction in dosing frequency as compared to that drug or drugs presented as a conventional dosage form), capsule gelatin coated (a solid dosage form in which the drug is enclosed within either a hard or soft soluble container made from a suitable form of gelatin; through a banding process, the capsule is coated with additional layers of gelatin so as to form a complete seal), and capsule liquid filled (a solid dosage form in which the drug is enclosed within a soluble, gelatin shell which is

plasticized by the addition of a polyol, such as sorbitol or glycerin, and is therefore of a somewhat thicker consistency than that of a hard shell capsule; typically, the active ingredients are dissolved or suspended in a liquid vehicle).

[00115] Oral dosage forms contemplated herein also include granules (a small particle or grain), pellet (a small sterile solid mass consisting of a highly purified drug, with or without excipients, made by the formation of granules, or by compression and molding), pellets coated extended release (a solid dosage form in which the drug itself is in the form of granules to which varying amounts of coating have been applied, and which releases a drug or drugs in such a manner to allow a reduction in dosing frequency as compared to that drug or drugs presented as a conventional dosage form), pill (a small, round solid dosage form containing a medicinal agent intended for oral administration), powder (an intimate mixture of dry, finely divided drugs and/or chemicals that may be intended for internal or external use), elixir (a clear, pleasantly flavored, sweetened hydroalcoholic liquid containing dissolved medicinal agents; it is intended for oral use), chewing gum (a sweetened and flavored insoluble plastic material of various shapes which when chewed, releases a drug substance into the oral cavity), or syrup (an oral solution containing high concentrations of sucrose or other sugars; the term has also been used to include any other liquid dosage form prepared in a sweet and viscid vehicle, including oral suspensions).

[00116] Oral dosage forms contemplated herein may further include a tablet (a solid dosage form containing medicinal substances with or without suitable diluents), tablet chewable (a solid dosage form containing medicinal substances with or without suitable diluents that is intended to be chewed, producing a pleasant tasting residue in the oral cavity that is easily swallowed and does not leave a bitter or unpleasant after-taste), tablet coated (a solid dosage form that contains medicinal substances with or without suitable diluents and is covered with a designated coating), tablet coated particles (a solid dosage form containing a conglomerate of medicinal particles that have each been covered with a coating), tablet delayed release (a solid dosage form which releases a drug or drugs at a time other than promptly after administration, whereby enteric-coated articles are delayed release dosage forms), tablet delayed release particles (a solid dosage form containing a conglomerate of medicinal particles that have been covered with a coating which releases a drug or drugs at a time other than promptly after administration, whereby enteric-coated articles are delayed release dosage forms), tablet

dispersible (a tablet that, prior to administration, is intended to be placed in liquid, where its contents will be distributed evenly throughout that liquid, whereby term 'tablet, dispersible' is no longer used for approved drug products, and it has been replaced by the term 'tablet, for suspension'), tablet effervescent (a solid dosage form containing mixtures of acids, for example, citric acid, tartaric acid, and sodium bicarbonate, which release carbon dioxide when dissolved in water, whereby it is intended to be dissolved or dispersed in water before administration), tablet extended release (a solid dosage form containing a drug which allows at least a reduction in dosing frequency as compared to that drug presented in conventional dosage form), tablet film coated (a solid dosage form that contains medicinal substances with or without suitable diluents and is coated with a thin layer of a water-insoluble or water-soluble polymer), tablet film coated extended release (a solid dosage form that contains medicinal substances with or without suitable diluents and is coated with a thin layer of a water-insoluble or water-soluble polymer; the tablet is formulated in such manner as to make the contained medicament available over an extended period of time following ingestion), tablet for solution (a tablet that forms a solution when placed in a liquid), tablet for suspension (a tablet that forms a suspension when placed in a liquid, which is formerly referred to as a 'dispersible tablet'), tablet multilayer (a solid dosage form containing medicinal substances that have been compressed to form a multiple-layered tablet or a tablet-within-a-tablet, the inner tablet being the core and the outer portion being the shell), tablet multilayer extended release (a solid dosage form containing medicinal substances that have been compressed to form a multiple-layered tablet or a tablet-within-a-tablet, the inner tablet being the core and the outer portion being the shell, which, additionally, is covered in a designated coating; the tablet is formulated in such manner as to allow at least a reduction in dosing frequency as compared to that drug presented as a conventional dosage form), tablet orally disintegrating (a solid dosage form containing medicinal substances which disintegrates rapidly, usually within a matter of seconds, when placed upon the tongue), tablet orally disintegrating delayed release (a solid dosage form containing medicinal substances which disintegrates rapidly, usually within a matter of seconds, when placed upon the tongue, but which releases a drug or drugs at a time other than promptly after administration), tablet soluble (a solid dosage form that contains medicinal substances with or without suitable diluents and possesses the ability to dissolve in fluids), tablet sugar coated (a solid dosage form that contains medicinal substances with or

without suitable diluents and is coated with a colored or an uncolored water-soluble sugar), and the like.

[00117] Injection and infusion dosage forms (i.e., parenteral dosage forms) include, but are not limited to, the following. Liposomal injection includes or forms liposomes or a lipid bilayer vesicle having phospholipids that encapsulate an active drug substance. Injection includes a sterile preparation intended for parenteral use. Five distinct classes of injections exist as defined by the USP. Emulsion injection includes an emulsion comprising a sterile, pyrogen-free preparation intended to be administered parenterally. Lipid complex and powder for solution injection are sterile preparations intended for reconstitution to form a solution for parenteral use.

[00118] Powder for suspension injection is a sterile preparation intended for reconstitution to form a suspension for parenteral use. Powder lyophilized for liposomal suspension injection is a sterile freeze dried preparation intended for reconstitution for parenteral use that is formulated in a manner allowing incorporation of liposomes, such as a lipid bilayer vesicle having phospholipids used to encapsulate an active drug substance within a lipid bilayer or in an aqueous space, whereby the formulation may be formed upon reconstitution. Powder lyophilized for solution injection is a dosage form intended for the solution prepared by lyophilization ("freeze drying"), whereby the process involves removing water from products in a frozen state at extremely low pressures, and whereby subsequent addition of liquid creates a solution that conforms in all respects to the requirements for injections. Powder lyophilized for suspension injection is a liquid preparation intended for parenteral use that contains solids suspended in a suitable fluid medium, and it conforms in all respects to the requirements for Sterile Suspensions, whereby the medicinal agents intended for the suspension are prepared by lyophilization.

[00119] Solution injection involves a liquid preparation containing one or more drug substances dissolved in a suitable solvent or mixture of mutually miscible solvents that is suitable for injection. Solution concentrate injection involves a sterile preparation for parenteral use that, upon addition of suitable solvents, yields a solution suitable for injections. Suspension injection involves a liquid preparation (suitable for injection) containing solid particles dispersed throughout a liquid phase, whereby the particles are insoluble, and whereby an oil phase is dispersed throughout an aqueous phase or vice-versa. Suspension liposomal injection is a liquid

preparation (suitable for injection) having an oil phase dispersed throughout an aqueous phase in such a manner that liposomes (a lipid bilayer vesicle usually containing phospholipids used to encapsulate an active drug substance either within a lipid bilayer or in an aqueous space) are formed. Suspension sonicated injection is a liquid preparation (suitable for injection) containing solid particles dispersed throughout a liquid phase, whereby the particles are insoluble. In addition, the product may be sonicated as a gas is bubbled through the suspension resulting in the formation of microspheres by the solid particles.

[00120] A parenteral carrier system may include one or more pharmaceutically suitable excipients, such as solvents and co-solvents, solubilizing agents, wetting agents, suspending agents, thickening agents, emulsifying agents, chelating agents, buffers, pH adjusters, antioxidants, reducing agents, antimicrobial preservatives, bulking agents, protectants, tonicity adjusters, and special additives.

[00121] Inhalation dosage forms include, but are not limited to, aerosol being a product that is packaged under pressure and contains therapeutically active ingredients that are released upon activation of an appropriate valve system intended for topical application to the skin as well as local application into the nose (nasal aerosols), mouth (lingual and sublingual aerosols), or lungs (inhalation aerosols). Inhalation dosage forms further include foam aerosol being a dosage form containing one or more active ingredients, surfactants, aqueous or nonaqueous liquids, and the propellants, whereby if the propellant is in the internal (discontinuous) phase (i.e., of the oil-in-water type), a stable foam is discharged, and if the propellant is in the external (continuous) phase (i.e., of the water-in-oil type), a spray or a quick-breaking foam is discharged. Inhalation dosage forms also include metered aerosol being a pressurized dosage form consisting of metered dose valves which allow for the delivery of a uniform quantity of spray upon each activation; powder aerosol being a product that is packaged under pressure and contains therapeutically active ingredients, in the form of a powder, that are released upon activation of an appropriate valve system; and aerosol spray being an aerosol product which utilizes a compressed gas as the propellant to provide the force necessary to expel the product as a wet spray and being applicable to solutions of medicinal agents in aqueous solvents.

[00122] Pharmaceutically suitable inhalation carrier systems may include pharmaceutically suitable inactive ingredients known in the art for use in various inhalation

dosage forms, such as (but not limited to) aerosol propellants (for example, hydrofluoroalkane propellants), surfactants, additives, suspension agents, solvents, stabilizers and the like.

[00123] A transdermal dosage form may include, but is not limited to, a patch being a drug delivery system that often contains an adhesive backing that is usually applied to an external site on the body, whereby the ingredients either passively diffuse from, or are actively transported from some portion of the patch, and whereby depending upon the patch, the ingredients are either delivered to the outer surface of the body or into the body; and other various types of transdermal patches such as matrix, reservoir and others known in the art. The "pharmaceutically suitable transdermal carrier system" includes pharmaceutically suitable inactive ingredients known in the art for use in various transdermal dosage forms, such as (but not limited to) solvents, adhesives, diluents, additives, permeation enhancing agents, surfactants, emulsifiers, liposomes, and the like.

[00124] Suitable dosage amounts and dosing regimens may be selected in accordance with a variety of factors, including one or more particular conditions being treated, the severity of the one or more conditions, the genetic profile, age, health, sex, diet, and weight of the subject, the route of administration alone or in combination with pharmacological considerations including the activity, efficacy, bioavailability, pharmacokinetic, and toxicological profiles of the particular compound employed, whether a drug delivery system is utilized and whether the drug is administered as part of a drug combination. Therefore, the dosage regimen to be employed may vary widely and may necessarily deviate from the dosage regimens set forth herein.

[00125] Contemplated dosage forms may include an amount of one or more expression inhibitors (or inhibitors of expression) ranging from about 1 to about 1200 mg, or about 5 to about 100 mg, or about 25 to about 800 mg, or about 100 to about 500 mg, or 0.1 to 50 milligrams (± 10%), or 10 to 100 milligrams (± 10%), or 5 to 500 milligrams (± 10%), or 0.1 to 200 milligrams (± 10%), or 1 to 100 milligrams (± 10%), or 5 to 50 milligrams (± 10%), or 30 milligrams (± 10%), or 20 milligrams (± 10%), or 10 milligrams (± 10%), or 5 milligrams (± 10%), per dosage form, such as, for example, a tablet, a pill, a bolus, and the like.

[00126] In another embodiment, a dosage form may be administered to a subject in need thereof once per day, or twice per day, or once every 6 hours, or once every 4 hours, or once every 2 hours, or hourly, or twice an hour, or twice a day, or twice a week, or monthly.

[00127] The phrase "therapeutically effective" is intended to qualify the amount that will achieve the goal of improvement in disease severity and/or the frequency of incidence over non-treatment, while limiting, reducing, or avoiding adverse side effects typically associated with disease therapies. A "therapeutic effect" relieves to some extent one or more of the symptoms of a cancer disease or disorder. In reference to the treatment of a cancer, a therapeutic effect refers to one or more of the following: 1) reduction in the number of cancer cells by, for example, killing the cancer cells; 2) reduction in tumor size; 3) inhibition (i.e., slowing to some extent, preferably stopping) of cancer cell infiltration into peripheral organs; 4) inhibition (i.e., slowing to some extent, preferably stopping) of tumor metastasis; 5) inhibition, to some extent, of tumor growth; 6) relieving or reducing to some extent one or more of the symptoms associated with the disorder; and/or 7) relieving or reducing the side effects associated with the administration of anticancer agents. "Therapeutic effective amount" is intended to qualify the amount required to achieve a therapeutic effect.

[00128] A therapeutically effective amount of an expression inhibitor (or inhibitors of expression) may be any amount that begins to improve cancer treatment in a subject. In one embodiment, an effective amount of an expression inhibitor used in the therapeutic regime described herein may be, for example, about 1 mg, or about 5 mg, or about 10 mg, or about 25 mg, or about 50 mg, or about 100 mg, or about 200 mg, or about 400 mg, or about 500 mg, or about 600 mg, or about 1000 mg, or about 1200 mg, or about 1400 mg, or from about 10 to about 60 mg, or about 50 mg to about 200 mg, or about 150 mg to about 600 mg per day. Further, another effective amount of an expression inhibitor used herein may be that which results in a detectable blood level of above about 1 ng/dL, 5, ng/dL, 10 ng/dL, 20, ng/dL, 35 ng/dL, or about 70 ng/dL, or about 140 ng/dL, or about 280 ng/dL, or about 350 ng/dL, or lower or higher.

[00129] The term "pharmaceutically acceptable" is used herein to mean that the modified noun is appropriate for use in a pharmaceutical product. Pharmaceutically acceptable cations include metallic ions and organic ions. Other metallic ions include, but are not limited to appropriate alkali metal salts, alkaline earth metal salts and other physiological acceptable metal ions. Exemplary ions include aluminium, calcium, lithium, magnesium, potassium, sodium and zinc in their usual valences. Organic ions include protonated tertiary amines and quaternary ammonium cations, including in part, trimethylamine, diethylamine, N,N'-

dibenzylethylenediamine, chloroprocaine, choline, diethanolamine, ethylenediamine, meglumine (N-methylglucamine) and procaine. Pharmaceutically acceptable acids include without limitation hydrochloric acid, hydrobromic acid, phosphoric acid, sulfuric acid, methanesulfonic acid, acetic acid, formic acid, tartaric acid, maleic acid, malic acid, citric acid, isocitric acid, succinic acid, lactic acid, gluconic acid, glucuronic acid, pyruvic acid oxalacetic acid, fumaric acid, propionic acid, aspartic acid, glutamic acid, benzoic acid, and the like.

[00130] It is further contemplated that one active ingredient may be in an extended release form, while an optional second, third, or fourth other active ingredient, for example, may or may not be, so the recipient experiences, for example, a spike in the second, third, or fourth active ingredient that dissipates rapidly, while the first active ingredient is maintained in a higher concentration in the blood stream over a longer period of time. Similarly, one of the active ingredients may be an active metabolite, while another may be in an unmetabolized state, such that the active metabolite has an immediate effect upon administration to a subject whereas the unmetabolized active ingredient administered in a single dosage form may need to be metabolized before taking effect in the subject.

[00131] Also contemplated are solid form preparations that include at least one active ingredient which are intended to be converted, shortly before use, to liquid form preparations for oral administration. Such liquid forms include solutions, suspensions, and emulsions. These preparations may contain, in addition to the active component, colorants, flavors, stabilizers, buffers, artificial and natural sweeteners, dispersants, thickeners, solubilizing agents, and the like. Solutions or suspensions may be applied topically and/or directly to the nasal cavity, respiratory tract, eye, or ear by conventional means, for example with a dropper, pipette or spray.

[00132] Alternatively, one or more of the active ingredients may be provided in the form of a dry powder, for example a powder mix of the compound in a suitable powder base such as lactose, starch, starch derivatives such as hydroxypropylmethyl cellulose and polyvinylpyrrolidone (PVP). Conveniently the powder carrier may form a gel in the nasal cavity. The powder composition may be presented in unit dose form, for example, in capsules or cartridges of, for example, gelatin, or blister packs from which the powder may be administered by means of an inhaler.

[00133] The pharmaceutical preparations may be in unit dosage forms. In such form, the preparation may be subdivided into unit doses containing appropriate quantities of the active component. The unit dosage form can be a packaged preparation, such as a kit or other form, the package containing discrete quantities of preparation, such as packeted tablets, capsules, liquids or powders in vials or ampoules. Also, the unit dosage form can be a capsule, tablet, cachet, or lozenge, or it can be the appropriate number of any of these in packaged form.

EXAMPLES

[00134] MATERIALS AND METHODS

[00135] **Affymetrix Gene Arrays.** A total of 24 samples including 3 biological replicates of each [1833 cells expressing control scrambled shRNA (SCR sh) or high mobility group AT-hook 2 (HMGA2) shRNA (shHMGA2), MDA-MB-436 cells expressing SCR sh or shHMGA2, xenograft 1833 tumor cells expressing inducible ten–eleven translocation 1 (Tet1) or homeobox A9 (HOXA9) with or without induction of Tet1 or HOXA9] were analyzed by using Affymetrix GeneChip Human Gene 1.0 ST Array. The RNA quality control, cRNA amplification, hybridization, and image scan were conducted in the Functional Genomics Facility at the University of Chicago. The quantified signals were normalized by using Robust Multiarray Average (RMA) (1). R (Version 2.11) (2) and related packages from Bioconductor (Version 2.4) (3) were used for the analysis of the normalized data. Differential expression was defined as fold ≥ 1.25, $P < 0.05$, and false discovery rate (FDR) < 0.05. We also performed median-centering gene across all arrays for heatmap illustration. The microarray data have been deposited in the Gene Expression Ominibus repository (accession no. GSE43741).

[00136] **Cell Culture and Generation of Cell Lines.** Cell lines (1833, MDAMB- 231, andMDA-MB-436) were cultured in a complete medium consisting of DMEM supplemented with 10% (vol/vol) FBS, 50 U/mL penicillin, and 50 μg/mL streptomycin. For inducible cell lines, Tet System Approved FBS was used instead. HMGA2 depletion and control were achieved by transducing the cells with HMGA2 shRNA or scrambled control in a pLKO.1 lentiviral vector (Open Biosystems). After transduction, cells were selected and maintained in 0.5 μg/mL puromycin. The HMGA2 expressing cell line was generated by transfection with

pH3HX–HMGA2 plasmid using Attractene transfection reagent (Qiagen). Cells were selected in 500 μg/mL geneticin (G418). The TET1 expressing cell line was generated by transfection with pMSCV–Flag–TET1- puro plasmid using Attractene (Qiagen). Cells were selected in 0.5 μg/mL puromycin. The inducible Tet1- or HOXA9-expressing cell line was generated by cotransducing the cells with pLVX– TRE3G–Flag–Tet1 (expressing inducible Tet1) or pLVX–TRE3G–HOXA9 (expressing inducible HOXA9) and pCMV–Tet3G (expressing Tet-On 3G element) in lentiviral vectors (Clontech).Cells were selected and maintained in 0.5 μg/mL puromycin and 500 μg/mL G418.

[00137] Transient Transfection with siRNA. Cell lines were transiently transfected by using HiPerFect transfection reagent (Qiagen) following the manufacturer's protocols. TET1, HOXA7, or HOXA9 siRNA On-TARGET plus SMARTpool and the relevant control were from Dharmacon RNAi Technologies.

[00138] RNA Isolation and Quantitative RT-PCR Analysis. Total RNA was isolated from cells by using miRNeasy Mini Kit (Qiagen) following the manufacturer's instructions. Quantitative RT-PCR (qRT-PCR) was performed as described (4). The qPCR primers for human or murine HMGA2, TET1, HOXA genes and GAPDH were Taqman and were purchased from Applied Biosystems.

[00139] Immunoblotting. Cells were lysed in complete lysis buffer (pH 7.5) by using Nuclear Extract Kit (Active Motif) following the manufacturer's protocols. Proteins were measured, separated, and probed as described (4). Antibodies specific for each protein were HMGA2 (61042; Active Motif), TET1 (sc-163446; Santa Cruz Biotechnologies), HOXA7 (09-086; Millipore), HOXA9 (09-178; Millipore), GAPDH (ab9484; Abcam), and Flag-M1 (F3040; Sigma).

[00140] Genomic DNA Isolation and Analysis of DNA Methylation by Methylation-Specific Digestion Combined with qPCR or by Bisulfite Modification Sequencing. Genomic DNA was isolated from cells with the DNeasy Blood and Tissue Kit (Qiagen) following the manufacturer's instructions. For detection of TET1 and HOX gene promoter CpG island methylation, genomic DNA was subjected to four digestions (mock, methylation-sensitive, methylationdependent, or both) by using the EpiTect Methyl DNA Restriction Kit (SABiosciences) following the manufacturer's instructions. Products from the digestion were

quantified for levels of methylation by qPCR using either the Human HOX Genes DNA Methylation PCR Array or the EpiTect Methyl qPCR Primer Assay for TET1 (SABiosciences). Results were analyzed by software from SABiosciences (www.sabiosciences.com/dna_methylation_data_analysis.php). For bisulfite sequencing assay, genomic DNA was subjected to sodium bisulfite treatment by using the EpiTect Bisulfite Kits (Qiagen). Products from the treatment were amplified by PCR using specific primers designed with MethPrimer software (5):

SEQ ID 1 TET1 Forward (5′–3′): TTATGTAGTTTTATTTGTTTTTTTATTGTG

SEQ ID 2 TET1 Reverse (5′–3′): CAACTCCAAACCTACACCAAC

SEQ ID 3 HOXA7 Forward (5′–3′): TATAATTTTGATTTGTGATTTGTTGTT

SEQ ID 4 HOXA7 Reverse (5′–3′): AAACCTCTTACCCTTCCATTCTAAA

SEQ ID 5 HOXA9 Forward (5′–3′): TTGGGAATTTTGATTGTTAGTTGA

SEQ ID 6 HOXA9 Reverse (5′–3′): TACCAAAACACTCCAAACAAAAAC

PCR products were purified by using a PCR purification kit (Qiagen). Purified products were subcloned by using a TA Cloning Kit (Invitrogen), and individual inserts from 10 or more randomly selected clones were sequenced.

[00141] 5-Hydroxymethylcytosine Labeling Reaction and Dot-Blot Assay. The 5-hydroxymethylcytosine (5-hmC) labeling reactions and dot-blot assays were performed as described (6). Briefly, 600 ng of genomic DNA samples were spotted and measured for levels of 5-hmC. Quantification was calculated by using a working curve generated by 1–8 ng of 32-bp synthetic biotin-5-N3-gmC–containing DNA.

[00142] ChIP Assay. The ChIP assay was performed with the Champion- ChIP One-Day Kit (SABiosciences) following the manufacturer's instructions. Briefly, cells were fixed and cross-linked with 1% formaldehyde. Cross-linked chromatin was sheared by using a sonicator. Antibody used for immunoprecipitation was anti-Tet1 (Santa Cruz Biotechnologies), anti-H3K4Me3, or anti-IgG (Abcam). Precipitated DNA was purified and then analyzed by qPCR with primers specific for the TET1, HOXA7, or HOXA9 region:

SEQ ID 7 TET1 Site-1 Forward (5′–3′): TTTGGGAACCGACTCCTCACCT

SEQ ID 8 TET1 Site-1 Reverse (5′–3′): TCGGGCAAACTTTCCAACTCGC

SEQ ID 9 TET1 Site-2 Forward (5′–3′): ACGCTGGGCATTTCTGATCCACTA

SEQ ID 10 TET1 Site-2 Reverse (5′–3′): TATTGTGCAGCTCGTTTAGTGCCC

SEQ ID 11 TET1 Site-3 Forward (5′–3′): ACTTTGACCTCCCAAAGTGCTGGA

SEQ ID 12 TET1 Site-3 Reverse (5′–3′): ACCTGAGTGATGCTGAGACTTCCT

SEQ ID 13 HOXA7 Site-1 Forward (5′–3′): AAAGCGCGTTCACATAATAC

SEQ ID 14 HOXA7 Site-1 Reverse (5′–3′):GTTATCATATATCACTCTACCTCGT

SEQ ID 15 HOXA7 Site-2 Forward (5′–3′): CATTCCTGCTCCGGTTT

SEQ ID 16 HOXA7 Site-2 Reverse (5′–3′): GGTCCATAAAGGCCGAAG

SEQ ID 17 HOXA7 Site-3 Forward (5′–3′): CCACCCTGCCTTGTTTCAACATCA

SEQ ID 18 HOXA7 Site-3 Reverse (5′–3′): ACCAAGTTGTCAGTGAGCCTTCCA

SEQ ID 19 HOXA9 Site-1 Forward (5′–3′): TTCATCCTCACCAGCAGTTCCAGT

SEQ ID 20 HOXA9 Site-1 Reverse (5′–3′): GGGCCATTTCGGAGTTCATTGTGT

SEQ ID 21 HOXA9 Site-2 Forward (5′–3′): CCACCCTGCCTTGTTTCAACATCA

SEQ ID 22 HOXA9 Site-2 Reverse (5′–3′): ACCAAGTTGTCAGTGAGCCTTCCA

For the input control, 1% of sonicated DNA was directly purified and analyzed by qPCR with the same primers.

[00143] Demethylation Treatment. Cells were subjected to 5-azacytidine (1 or 3.6 μM), decitabine (5-aza-dC; 220 or 440 nM), or mock treatment by addition into cell culture at indicated concentration. Cells were treated daily. RNA and protein were isolated respectively.

[00144] Cloning of pMSCV–Flag–TET1–puro, Inducible Tet1, or Inducible HOXA9 Plasmid. The DNA sequence containing a Flag sequence encoding DYKDDDDK and the mouse Tet1 gene C-terminal 673 amino acids including the catalytic domain (GU079948) was synthesized by GenScript, and then was inserted into a MSCVpuro plasmid at XhoI/EcoRI sites.

For cloning of inducible Tet1, the Tet1 sequence fragment was fused into the pLVX–TRE3G, a Tet-On 3G inducible lentiviral vector (Clontech) at BamHI/EcoRI sites. A similar strategy was applied for cloning of inducible HOXA9 plasmid. The DNA sequence for HOXA9 coding sequence was obtained from the pMSCVPIG–HOXA9 plasmid, a gift from Jianjun Chen (The University of Chicago, Chicago). All constructs were confirmed by sequencing.

[00145] **Cell Proliferation and in Vitro Cell Invasion Assays.** Cell proliferation assays were performed by using the CellTiter-Blues assay (Promega) as described (4). Invasion assays were performed as described (4) with modifications. Briefly, the inserts were coated with Matrigel basement membrane matrix (BD Biosciences). To assess the cell invasion ability, 105 cells were seeded on top of the polymerized Matrigel in serum-free medium, and complete medium (10% FBS) was placed in the lower compartment. After 24 h, cells on the lower part of the insert were stained with BD Calcein AM Fluorescent Dye. The inserts with the invaded cells were incubated in dissociation buffer (Travigen) with gentle shaking. Fluorescence measurements were used to record data with 485 nm for excitation and 530 nm for emission.

[00146] **Animal Studies.** Treatment of mice was done in accordance with a protocol approved by the Institutional Animal Care and Use Committee of The University of Chicago. Xenograft breast tumor growth and bone metastasis assay have been described (4). For the induction of Tet1 or HOXA9, 1833 Tet1 or HOXA9 inducible cells were plated in the presence of 1 μg/mL doxycycline. Twenty-four hours later, 106 cells were orthotopically injected into the second mammary fat pad of mice for tumor growth assay, or 105 cells were injected into the left ventricle of mice for bone metastasis assay. Mice were administered drinking water containing 4% sucrose only or 2 mg/mL doxycycline and 4% sucrose. Mice were imaged for luciferase activity after 3 wk. After 6 wk, tumor tissues were dissected, fixed, and embedded. For intravasation assays, the mouse blood was taken from heart and lysed by the addition of red blood cell lysis buffer (pH 7.2) (STEMCELL). Cells were collected by centrifugation, and total RNA isolated from cells was analyzed for human (tumor) and mouse (control) GAPDH transcripts by qRT-PCR.

[00147] **MMTV–Wnt1 Hmga2 Knockout Mice.** Wnt1 transgenic mice in the Hmga2 wild-type (Hmga2+/+), heterozygous (Hmga2+/−), or null (Hmga2−/−) genetic backgrounds have been described (7). Briefly, C57BL/6J–Hmga2+/− female mice were mated with

C57BL/6J–Wnt1 male mice (Jackson Laboratory). F1 Wnt1 transgenic Hmga2+/− male mice were then mated with F1 Wnt1 transgenic Hmga2+/− female littermates to obtain the F2 transgenic mice in the Hmga2+/+, Hmga2+/−, and Hmga2−/− genetic backgrounds. Wnt and Hmga2 loci have been confirmed by PCRbased genotyping.

[00148] **Immunostaining.** Immunostaining for paraffin-embedded tumor samples was performed by the Human Tissue Resource Centre Core Facility at the University of Chicago. Section of the samples was stained with hematoxylin and eosin (H&E), anti-Tet1 (Abcam), anti-Hoxa9 (Abcam), anti–5-hydroxy-methylcytosine (Active Motif), or anti-Ki67 (Thermo Scientific) antibody.

[00149] **Statistical Analysis of Experimental Results.** Samples were analyzed by using the two-sample Student t test assuming equal variances (two-tailed). P values were calculated for samples from three independent experiments unless otherwise indicated. Gene annotation enrichment analysis was performed by using Database for Annotation, Visualization and Integrated Discovery software (8). Gene set enrichment analysis was performed by using GSEA software (9).

[00150] **Subject Data and Kaplan–Meier Analysis.** Gene expression array data (10–13) and relevant clinical information for >800 breast cancer subject samples were downloaded either from the relative publication Web sites or from the www.ncbi.nlm.nih.gov/geo repository. The data were organized into two sets based on the platform on which the arrays were performed. Set one includes 101 microarrays for the breast cancer subject samples; set two includes 735 microarrays. Kaplan–Meier analysis was performed by using survival package in R (2).

[00151] **References for this Section**

[00152] 1. Irizarry RA, et al. (2003) Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res 31(4):e15.

[00153] 2. R Development Core Team (2009) R: A language and environment for statistical computing (R Foundation for Statistical Computing, Vienna).

[00154] 3. Gentleman RC, et al. (2004) Bioconductor: Open software development for computational biology and bioinformatics. Genome Biol 5(10):R80.

[00155] 4. Dangi-Garimella S, et al. (2009) Raf kinase inhibitory protein suppresses a metastasis signalling cascade involving LIN28 and let-7. EMBO J 28(4):347–358.

[00156] 5. Li LC, Dahiya R (2002) MethPrimer: Designing primers for methylation PCRs. Bioinformatics 18(11):1427–1431.

[00157] 6. Song CX, et al. (2011) Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol 29(1):68–72.

[00158] 7. Morishita A, et al. (2013) HMGA2 is an in vivo driver of tumor metastasis. Cancer Res, in press.

[00159] 8. Huang W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4(1):44–57.

[00160] 9. Subramanian A, et al. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 102(43):15545–15550.

[00161] 10. Enerly E, et al. (2011) miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. PLoS ONE 6(2):e16915.

[00162] 11. Miller LD, et al. (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci USA 102(38):13550–13555.

[00163] 12. Desmedt C, et al.; TRANSBIG Consortium (2007) Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. Clin Cancer Res 13(11): 3207–3214.

[00164] 13. Wang Y, et al. (2005) Gene-expression profiles to predict distant metastasis of lymphnode- negative primary breast cancer. Lancet 365(9460):671–679.

[00165] *HMGA2* depleted cell lines were generated by lentiviral transduction with *HMGA2* shRNA; *TET1* and *HOXA* gene knock-down were generated by transfection with relative siRNA; *TET* and *HOXA9* inducible expression cells were generated by lentiviral transduction with relative inducible expression vector.

[00166] Example 1. TET1 and HOX gene expression are dramatically induced upon depletion of HMGA2 in both invasive human breast cancer cells and *MMTV-Wnt1* mouse breast tumors

[00167] Gene expression arrays were conducted using two invasive human breast cancer cell lines expressing either HMGA2 shRNA or control scrambled shRNA: 1833, a bone-tropic derivative of the human breast cancer cell line MDA-MB-231 (23), and MDA-MB-436. We found a dramatic induction of Homeobox (HOX) genes particularly at the *HOXA* loci in the more invasive 1833 cells (Figs. 1*A-C* and S1). The HOX genes are comprised of four clusters (A, B, C, and D) located on different human chromosomes. This transcriptional factor family with 39 members controls posterior-anterior patterning during embryogenesis and the development of specific organs (reviewed in Ref.(24)). We validated the induction of *HOXA* gene expression including *HOXA4*, *HOXA5*, *HOXA6*, *HOXA7*, *HOXA9* and *HOXA11* in *HMGA2*-depleted 1833 cells at both mRNA (Figs. 1*D* and *F*) and protein (Fig. 1*G*) levels.

[00168] Among the other genes induced by *HMGA2* depletion was *TET1* (Figs. 1*A* and S1). We confirmed that HMGA2 is a negative regulator of TET1 in 1833 cells by qRT-PCR and immunoblotting (Figs. 1*E* and *G*). Consistent with increased TET1 expression, we observed elevated 5hmC levels in *HMGA2*-depleted 1833 cells (Fig. 1*H*). We also observed similar induction of *TET1* and *HOXA* gene expression following *HMGA2* depletion by shRNA in MDA-MB-436 cells, although the effects were not as robust, consistent with their relatively higher basal levels of TET1 and HOXA protein and the less invasive phenotype of these cells (Figs. S2*A-D*). Consistent with these observations, analysis of gene expression in a cohort of 75 human breast tumors (25) showed a significantly negative correlation between *HMGA2* and *TET1* gene expression (Fig. S2*E*), and this relationship existed in both ER-negative (Fig. S2*F*) and ER-positive (Fig. S2*G*) subject subpopulations.

[00169] To validate the regulation of *TET1* and *HOXA* genes by HMGA2 *in vivo*, we used an *MMTV-Wnt1* transgenic mouse model for breast cancer. Deletion of *Hmga2* by crossing *MMTV- Wnt1* mice with *Hmga2*-specific knock-out mice reduced tumor incidence (26) and decreased tumor cell proliferation as assessed by immunostaining of Ki67 (Fig. S3). Analysis by qRT-PCR or immunohistochemistry showed a strong induction of *Tet1*, 5hmC and *Hoxa* gene expression including *Hoxa9* and *Hoxa7* in tumors from the *MMTV-Wnt1*/Hmga2$^{-/-}$ mice (Figs.

1*I* and *J*). Moreover, expression of HMGA2 in both 1833 and MDA-MB-436 cells inhibited *TET1* expression (Fig. S4). These results indicate that induction of *TET1* by depletion of *HMGA2* is not an off-target effect, and raise the possibility that loss of *HMGA2* suppresses breast tumor growth by inducing *TET1* and *HOXA* genes.

[00170] Example 2. TET1 is involved in an auto-regulation in human breast cancer cells

[00171] Since the TET1 protein may bind directly to its own promoter region as suggested by the ChIP-seq data for *Tet1* in mouse ES cells (9, 10, 27, 28), we investigated whether TET1 was also involved in regulating its own expression in human breast cancer cells. Conventional ChIP assays using 1833 cells expressing *TET1* or vector control showed that TET1 bound to its own promoter region (Fig. 2*A*). Consistent with increased expression of TET1, ChIP assays also showed that 1833 cells expressing TET1 exhibited increased binding of H3K4Me3, a histone marker for transcriptional activation, to the *TET1* promoter region (Fig. 2*A*). Furthermore, because TET proteins are typically involved in DNA demethylation pathways (2-5, 29, 30), we analyzed the effect of TET1 on the methylation status of its own promoter regions. Our DNA methylation-specific digestion combined with qPCR showed that about 70% of the *TET1* promoter region within ± 1 kilobase from the transcription start site (TSS) in the parent 1833 cells contained methylated CpG islands, whereas the fraction decreased to 9% following *HMGA2* depletion (Fig. 2*B*). Similarly, bisulfite sequencing within the same region of the *TET1* promoter showed an increase in demethylated CpGs from 38% in 1833 cells to 86% in *HMGA2*- depleted cells (Fig. 2*C*). 1833 cells treated with 5-azacytidine, a demethylation reagent that inhibits DNMTs, also showed an increase in TET1 expression (Fig. 2*D*). Together, our results suggest that *HMGA2* depletion in 1833 cells causes extensive demethylation of the *TET1* promoter, and therefore results in a robust induction of *TET1* expression.

[00172] Example 3. TET1 directly induces *HOXA* gene expression in breast cancer cells through binding to the promoter regions of *HOXA* genes and contributing to local demethylation

[00173] Previous ChIP-seq analyses of TET1 in mouse ES cells also implied that *Hoxa* genes might be downstream targets of TET1 as their promoter regions are enriched with TET1

protein binding (9, 10, 27, 28). To investigate whether TET1 is an upstream regulator of the *HOXA* cluster in human breast cancer cells, we transfected HMGA2-depleted 1833 cells with *TET1* siRNA and observed a significant decrease in *HOXA* gene expression and 5hmC levels along with decreased *TET1* expression (Figs. 3*A* and *B*). Conversely, ectopic expression of *TET1* in the parent 1833 cells dramatically increased *HOXA9* expression and 5hmC levels (Figs. 3*C* and *D*).

[00174] To test whether TET1 can regulate *HOXA* gene expression *in vivo*, we stably transduced 1833 cells with an inducible *TET1* expression vector. Cells were orthotopically injected into the mammary fat pad of mice. QRT-PCR and immunohistochemistry analysis of the mouse xenografts after six weeks of treatment with doxycycline showed a dramatic induction of TET1 and HOXA9 expression (Figs. 3*E* and *F*) as well as increased 5hmC levels (Fig. 3*F*). Gene expression analysis in a cohort of 54 human breast tumors (25) also showed a strong positive correlation between *TET1* and *HOXA* gene expression (Fig. 3*G*).

[00175] We next investigated whether the induction of *HOXA* gene expression can also be attributed to direct binding of TET1 and subsequent demethylation of the *HOXA* promoter regions. Our ChIP-qPCR assays confirmed that TET1 and H3K4Me3 also bound to the *HOXA* gene promoter regions (Figs. 4*A* and *B*). Our DNA methylation-specific digestion combined with qPCR assay showed that, in the parent 1833 cells, only a small fraction of the promoter region of the *HOXA4-A11* genes lacked methylation (Fig. 4*C*), accounting for the low expression of *HOXA* transcripts; by contrast, *HMGA2* depletion caused dramatic loss of methylation at the *HOXA* gene loci (Fig. 4*C*). Bisulfite sequencing of the 14 CpGs near the *HOXA7* transcription start site showed only 4% were unmethylated, whereas *HMGA2* depletion caused over 80% demethylation (Fig. 4*D*). Similarly, bisulfite sequencing of the 15 CpGs in the upstream 1kb locus of the *HOXA9* promoter showed that only 5% of the CpG sites were unmethylated, while unmethylated CpGs increased to 91% upon loss of *HMGA2* in 1833 cells (Fig. 4*E*). 1833 cells treated with 5-azacytidine to demethylate DNA also showed a dramatic increase in *HOXA7* and *HOXA9* expression (Figs. 4*F* and *G*). These results suggest that TET1 binds directly to the *HOXA* promoter regions and contributes to local demethylation, inducing activating histone binding and gene transcription in breast cancer cells.

[00176] Example 4. HMGA2/TET1/HOXA pathway regulates breast cancer cell invasion

[00177] To assess the pathological significance of this HMGA2/TET1/HOXA signaling cascade, we determined the effect of manipulating these genes on cell invasion. *HMGA2* depletion in 1833 cells decreased cell invasion (Fig. 5*A*); this effect was reversed in part by siRNA depletion of *TET1*, *HOXA9* or *HOXA7* (Figs. 5*B*, 5*C* and S5*I-K*). 1833 cells treated by demethylation reagent 5-azacytidine or decitabine showed a similar decreased cell invasion and a partial rescue in invasion followed by siRNA depletion of *HOXA9* (Figs. 5*D*, S5*A-C* and S5*E-G*). These data are consistent with a previous study showing that HOXA9 is a breast cancer inhibitor (31). Taken together, our results reveal a signaling cascade whereby HMGA2 promotes breast cancer cell invasion in part through inhibition of TET1-mediated demethylation and *HOXA* gene expression.

[00178] Example 5. Both TET1 and its downstream target, HOXA9, suppress breast tumor growth, intravasation and metastasis

[00179] To determine whether TET1 or HOXA9 can reverse the tumorigenic phenotype in breast cancer cells transformed by HMGA2, we injected 1833 cells expressing inducible *TET1* or *HOXA9* into the mammary fat pad of mice followed by doxycycline treatment and tested their effect on xenograft tumor growth. Consistent with our *in vitro* observation (Figs. S5*D*, S6*A-C*), induced expression of *TET1* (Figs. 3*E* and *F*) or *HOXA9* (Figs. S6*D* and *E*) significantly suppressed xenograft tumor growth (Figs. 5*E-G*) and tumor cell proliferation (Figs. 5*H* and *I*).

[00180] To test TET1 or HOXA9 regulation of invasion *in vivo*, we determined their effect on tumor cell intravasation from a primary site in a murine orthotopic model. The 1833 cells expressing inducible *TET1* or *HOXA9* were injected into the mammary fat pad of mice. After 6 weeks of treatment with doxycycline, cells isolated from the blood were lysed and analyzed for human (tumor) or mouse (control) *GAPDH* transcripts. QRT-PCR analysis showed that both TET1 and HOXA9 significantly inhibited tumor cell intravasation (Figs. 5*J* and *K*).

[00181] Since *HMGA2* depletion suppresses breast tumor cell invasion and bone metastasis (21, 22), we determined whether its downstream effecters TET1 and HOXA9 similarly inhibit tumor metastasis. Luciferase-labeled 1833 cells expressing inducible *TET1* or

HOXA9 were injected into the left ventricle of mice that were subsequently treated with doxycycline. After 3 weeks, mice were imaged for luciferase activity. TET1 or HOXA9 expression caused a dramatic decrease in bone metastasis (Figs. 5*L* and *M*), and a significant increase in overall survival rate (Figs. 5*N* and S6*F-H*).

[00182] Example 6. HMGA2/TET1/HOXA9 regulate a common set of important genes and encompass a prognostic signature for subject survival

[00183] To identify and compare target genes of HMGA2, TET1 and HOXA9, we performed additional microarray assays for cells expressing induced *TET1* and *HOXA9*. Compared to the parental 1833 cells, there were 1012, 7220 and 7132 genes differentially expressed ($p < 0.05$, FDR <.05, and fold change > 1.25) upon *HMGA2* depletion, *TET1* induction or *HOXA9* induction, respectively (Fig. 6*A*). Interestingly, over 60% of the genes differentially regulated by TET1 or HOXA9 were the same (4510 genes, Fig. 6*A*), indicating that HOXA9 is a major downstream effecter of TET1. There were 214 genes that overlapped among all three sets including 144 up- regulated and 70 down-regulated genes (Fig. 6*A*; Table S1). Gene annotation enrichment analysis (DAVID) (32) indicated that the 144 up-regulated set was enriched in genes that have functions such as binding, catalytic activity, transcription regulator activity, and developmental processes, whereas the 70 down-regulated set was enriched in genes related to epithelial cell proliferation (p= 0.041), and the extracellular matrix (p = 0.0077) (Table S2). Gene set enrichment analysis (GSEA) (33) indicated that the down-regulated set was also enriched in genes that promote tumor growth and comprise metastasis signatures, such as CCL2, EFEMP1, IL7R, PPAP2B and STX3 (34). This pattern of gene regulation is consistent with a role for TET1 through its effecter HOXA9 in the suppression of breast tumor growth and metastasis.

[00184] These data illustrate a novel signaling cascade in human breast cancer progression, by which expression of the oncogene *HMGA2* leads to TET1 suppression. Since TET1 binds and demethylates itself as well as *HOXA* genes including *HOXA7* and *HOXA9*, decreased TET1 causes further inhibition of *TET1* as well as loss of *HOXA* gene expression. Suppression of TET1 and HOXA9 then enables expression of genes that promote breast tumor growth and metastasis (Fig. 6*B*). When considered individually, neither gene expression of *HMGA2*, *TET1*, *HOXA7* nor *HOXA9* significantly predicts survival in a heterogeneous group

of breast cancer subjects (Fig. 6*C*, left panel; Fig. S7, left panel). By contrast, Kaplan-Meier analysis using the complete HMGA2-TET1-HOXA pathway (*HMGA2* high, and *TET1* low, *HOXA9/7* low versus *HMGA2* low, and *TET1* high, *HOXA9/7* high) or a combination of *HMGA2* and *HOXA* genes was able to stratify subjects and predict survival (Fig. 6*C*, right panel; Fig. S7, right panel). There are no significant differences in the composition of cancer subtypes between the two stratified groups of subjects (Table S3), suggesting that this regulatory mechanism exists in a variety of breast cancer subtypes. These results indicate that the individual genes are not predictive alone but together define a relevant signaling environment that can be used to identify subjects for targeted DNA methylation-based therapy.

[00185] TABLES

[00186] Table S1. List of the 214 genes commonly regulated by HMGA2, TET1 and HOXA9 in 1833 cells.

SYMBOL	Depleted HMGA2 vs. control	Induced TET1 vs. control	Induced HOXA9 vs. control	Tet1 target in mES
ABCG2	down	down	down	no
ACVR1C	up	up	up	yes
ADAMTS12	down	down	down	no
AK5	down	down	down	yes
AKAP12	up	up	up	yes
APOC1	up	up	up	no
ARL6	up	up	up	no
ASAM	down	down	down	no
BAI3	up	up	up	yes
BCL6B	up	up	up	yes
BMP4	down	down	down	no
BRWD1	up	up	up	yes
C15orf51	up	up	up	no
C1orf91	up	up	up	no
C3orf28	up	up	up	no
C6orf211	up	up	up	no
C9orf86	up	up	up	no
CA2	up	up	up	no
CALCRL	up	up	up	no
CAMKV	up	up	up	yes
CASKIN1	up	up	up	yes
CCDC120	up	up	up	no

CCL2	down	down	down	no
CDA	down	down	down	yes
CDC37L1	up	up	up	yes
CENTA2	up	up	up	no
CLCN4	down	down	down	no
CMBL	down	down	down	no
CNTN1	up	up	up	no
COX7C	up	up	up	no
CREB3L2	down	down	down	yes
CSRP1	down	down	down	yes
CYP4V2	down	down	down	no
DDAH1	down	down	down	yes
DEPDC2	up	up	up	no
DIXDC1	up	up	up	yes
DKFZp434H1419	down	down	down	no
DLX2	up	up	up	yes
DNAJC6	down	down	down	yes
DSC2	up	up	up	no
DTX3L	down	down	down	yes
DUSP10	down	down	down	yes
DYNC2LI1	up	up	up	yes
DYSF	down	down	down	yes
EFEMP1	down	down	down	yes
EIF5A2	down	down	down	yes
EML1	up	up	up	yes
EPHA3	up	up	up	yes
EPHA7	up	up	up	no
EPHB1	up	up	up	yes
EPM2AIP1	up	up	up	yes
ERC2	up	up	up	yes
ETV5	down	down	down	yes
EVI2A	down	down	down	no
FABP6	up	up	up	no
FAM130A2	up	up	up	no
FAM83B	up	up	up	no
FBXO27	down	down	down	yes
FKBP10	up	up	up	yes
FLJ32810	down	down	down	no
FLJ37396	up	up	up	no
FLJ37453	up	up	up	no
FLJ43315	up	up	up	no

FLJ44253	up	up	up	no
FLRT3	up	up	up	no
FUZ	up	up	up	no
GALNT3	down	down	down	yes
GALNTL1	up	up	up	yes
GFOD2	up	up	up	yes
GFPT1	down	down	down	yes
GJC1	up	up	up	yes
GPM6B	up	up	up	no
GPRIN3	up	up	up	yes
GSTM4	down	down	down	yes
GSTO2	up	up	up	no
GULP1	down	down	down	yes
H3F3A	up	up	up	yes
HEY1	up	up	up	yes
HIST1H2BJ	up	up	up	yes
HMGN3	up	up	up	yes
HNRNPA1	up	up	up	no
HOXD13	up	up	up	yes
HSPC105	down	down	down	no
IFFO	up	up	up	no
IFT57	up	up	up	yes
IFT81	up	up	up	yes
IL7R	down	down	down	no
INADL	up	up	up	yes
JARID1A	up	up	up	no
JHDM1D	up	up	up	yes
JMJD1A	up	up	up	no
JUB	down	down	down	yes
KAL1	down	down	down	no
KCNJ2	up	up	up	no
KCNQ3	down	down	down	yes
KIAA1199	down	down	down	no
KLHDC8B	up	up	up	yes
KLHL3	up	up	up	no
LGR4	down	down	down	no
LIPH	down	down	down	yes
LMBR1L	up	up	up	yes
LOC286297	up	up	up	no
LOC389833	up	up	up	no
LOC440737	up	up	up	no

LOC646934	up	up	up	no
LOC728220	up	up	up	no
LOC728914	up	up	up	no
LOC729530	up	up	up	no
LPHN2	up	up	up	no
LPHN3	up	up	up	yes
LRP1B	up	up	up	no
LZTFL1	up	up	up	yes
MAMLD1	down	down	down	no
MARCH3	down	down	down	no
MFGE8	down	down	down	yes
MID1	down	down	down	no
MOAP1	up	up	up	yes
MPND	up	up	up	yes
MTIF3	up	up	up	yes
NCOA5	down	down	down	no
NDUFA1	up	up	up	yes
NDUFB4	up	up	up	no
NEBL	up	up	up	yes
NEK3	up	up	up	yes
NEXN	down	down	down	yes
NPHP3	up	up	up	yes
NR2F1	up	up	up	yes
NR3C2	down	down	down	yes
NRK	down	down	down	no
NSF	down	down	down	no
NUP62CL	up	up	up	yes
OAF	down	down	down	yes
OAS3	down	down	down	no
PAK1	up	up	up	no
PAXIP1	down	down	down	yes
PCDH17	up	up	up	no
PCDH18	up	up	up	yes
PCDH19	up	up	up	yes
PCDH7	up	up	up	yes
PCDH9	up	up	up	yes
PDE5A	up	up	up	yes
PDE7B	down	down	down	no
PES1	up	up	up	no
PHACTR1	down	down	down	yes
PLK2	down	down	down	yes

PLTP	up	up	up	yes
PPAP2B	down	down	down	yes
PPARA	down	down	down	yes
PTPRB	up	up	up	yes
RAB40A	up	up	up	no
RANBP9	down	down	down	yes
RAPGEF4	up	up	up	no
RARB	up	up	up	no
RASGEF1B	up	up	up	yes
RBM3	up	up	up	no
RBM34	up	up	up	yes
RNF128	up	up	up	yes
RNU2	up	up	up	no
RNU2B	up	up	up	no
ROR1	down	down	down	yes
RPL3	up	up	up	no
RPL35	up	up	up	no
RPL36A	up	up	up	no
RPL39	up	up	up	no
SCARNA8	up	up	up	no
SCARNA9	up	up	up	no
SECTM1	down	down	down	no
SEPP1	up	up	up	no
SERINC2	down	down	down	yes
SERPINF1	up	up	up	no
SERTAD4	down	down	down	yes
SESN3	up	up	up	yes
SLAMF7	down	down	down	no
SLC13A4	up	up	up	no
SLC16A2	down	down	down	no
SLC22A3	down	down	down	yes
SLC29A4	up	up	up	yes
SLC44A5	up	up	up	yes
SNCA	up	up	up	yes
SNORA14B	up	up	up	no
SNORA50	up	up	up	no
SNORD15A	up	up	up	no
SNORD15B	up	up	up	no
SNTB1	down	down	down	yes
SP140	down	down	down	no
SPON1	up	up	up	yes

SPPL2B	up	up	up	no
SPRED2	down	down	down	yes
SPRY2	down	down	down	yes
ST3GAL6	down	down	down	yes
STK38L	up	up	up	yes
STX3	down	down	down	no
SYT1	up	up	up	yes
TCF4	up	up	up	yes
TGDS	up	up	up	no
THOC3	up	up	up	yes
TIMP3	down	down	down	no
TMEM67	up	up	up	yes
TNFRSF21	down	down	down	yes
TSC22D3	up	up	up	no
TSPAN7	up	up	up	no
UXT	up	up	up	no
VAV1	up	up	up	yes
VKORC1	up	up	up	yes
WDR48	up	up	up	yes
WDR63	up	up	up	no
WNT3	down	down	down	yes
YPEL2	up	up	up	yes
ZBTB26	up	up	up	yes
ZNF484	up	up	up	no
ZNF74	up	up	up	no
ZNF773	up	up	up	no
ZNF84	up	up	up	no
ZSCAN23	up	up	up	no

[00187] Table S2. The list of gene categories in which the 214 genes commonly regulated by HMGA2, TET1 and HOXA9 in 1833 cells were enriched (*P* Value is hypergeometric probability).

Term	Count	Fold Enrichment	*P* Value	Genes	Regulation
Homophilic cell adhesion	6	6.8	0.002	*PCDH9, DSC2, PCDH7, PCDH17, PCDH19, PCDH18*	up

Cell adhesion	9	3.4	0.005	*FLRT3, CNTN1, PCDH9, DSC2, PCDH7, PCDH17, PCDH19, PCDH18, SPON1*	up
Plasma membrane	35	1.4	0.011	*SYT1, SNCA, AKAP12, TSPAN7, KCNJ2, EPHB1, ACVR1C, GJC1, RAB40A, CAMKV, LPHN2, LPHN3, SLC29A4, BAI3, CALCRL, PAK1, FLRT3, PTPRB, DIXDC1, INADL, PCDH9, PCDH7, PCDH17, VAV1, PCDH19, PCDH18, EPHA3, TMEM67, LMBR1L, EPHA7, CNTN1, DSC2, CA2, SLC13A4, ERC2*	up
Cell junction	7	2.8	0.038	*SYT1, DIXDC1, INADL, DSC2, ERC2, PAK1, GJC1*	up
Methylation	6	4.0	0.017	*RPL36A, LOC728914, RBM3, HIST1H2BJ, H3F3A, HNRNPA1, RAB40A*	up
Phosphorylation	11	2.0	0.039	*CAMKV, NDUFB4, EPHA7, NEK3, SNCA, PAK1, NDUFA1, STK38L, EPHB1, ACVR1C, EPHA3*	up
Translational elongation	4	5.9	0.030	*RPL36A, RPL35, RPL3, RPL39*	up
Negative regulation of transcription from RNA polymerase II promoter	6	3.4	0.032	*DLX2, HEY1, BCL6B, RARB, TCF4, NR2F1*	up
Branching morphogenesis of a nerve	2	99.1	0.020	*DLX2, EPHA7*	up
Proteinaceous extracellular matrix	6	5.1	0.006	*BMP4, WNT3, KAL1, EFEMP1, ADAMTS12, TIMP3*	down
Extracellular matrix	6	4.7	0.008	*BMP4, WNT3, KAL1, EFEMP1, ADAMTS12, TIMP3*	down

Response to extracellular stimulus	5	5.4	0.013	*BMP4, PPARA, CCL2, SLC22A3, TIMP3*	down
Plasma membrane part	15	1.9	0.020	*PHACTR1, JUB, STX3, ASAM, MFGE8, NEXN, IL7R, SLC16A2, DYSF, KCNQ3, SNTB1, SPRED2, ROR1, SLC22A3, EIF5A2*	down
Extracellular region part	9	2.5	0.020	*BMP4, SECTM1, WNT3, CCL2, KAL1, EFEMP1, MFGE8, ADAMTS12, TIMP3*	down
Membrane organization	6	3.7	0.020	*STX3, DYSF, GULP1, DNAJC6, MFGE8, MARCH3*	down
Membrane	31	1.4	0.024	*GALNT3, JUB, TNFRSF21, ASAM, NR3C2, IL7R, LGR4, SERINC2, SPRY2, DYSF, KCNQ3, EVI2A, ST3GAL6, SNTB1, SPRED2, CREB3L2, SLC22A3, PPAP2B, SECTM1, STX3, MFGE8, SLAMF7, CYP4V2, ABCG2, MARCH3, SLC16A2, PAXIP1, ROR1, LIPH, EIF5A2, CLCN4*	down
Signal	19	1.7	0.027	*BMP4, SECTM1, TNFRSF21, CCL2, ASAM, EFEMP1, MFGE8, SLAMF7, OAF, IL7R, TIMP3, LGR4, WNT3, KIAA1199, EVI2A, KAL1, ROR1, LIPH, ADAMTS12*	down
Epithelial cell proliferation	2	47.5	0.041	*BMP4, LGR4*	down

[00188] Table S3. The composition of breast cancer subtypes for two groups of subjects stratified by the complete HMGA2/TET1/HOXA pathway in Fig. 6C.

PATHWAY	High HMGA2 and Low	Low HMGA2 and High

		TET1/HOXA9 (N = 34)	TET1/HOXA9 (N = 35)
ER negative		10	11
HER2 negative		26	27
Tumor subtype	Basal	2	7
	ERBB2+	3	6
	Lum A	17	13
	Lum B	5	3
	Normal-like	2	3
	Others	5	3

[00189] References for the Examples

1. Hansen KD, *et al.* (2011) Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* 43(8):768-775.

2. Guo JU, Su Y, Zhong C, Ming GL, Song H (2011) Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* 145(3):423-434.

3. Ito S, *et al.* (2011) Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5- carboxylcytosine. *Science* 333(6047):1300-1303.

4. He YF, *et al.* (2011) Tet-Mediated Formation of 5-Carboxylcytosine and Its Excision by TDG in Mammalian DNA. *Science* 333:1303-1307.

5. Tahiliani M, *et al.* (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324(5929):930-935.

6. Inoue A, Zhang Y (2011) Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* 334(6053):194.

7. Ito S, *et al.* (2010) Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* 466(7310):1129-1133.

8. Koh KP, *et al.* (2011) Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* 8(2):200-213.

9. Xu Y, *et al.* (2011) Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell* 42(4):451-464.

10. Williams K, *et al.* (2011) TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature* 473(7347):343-348.

11. Haffner MC, *et al.* (2011) Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget* 2(8):627-637.

12. Kudo Y, *et al.* (2012) Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Sci* 103(4):670-676.

13. Yang H, *et al.* (2013) Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* 32(5):663-669.

14. Lian CG, *et al.* (2012) Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* 150(6):1135-1146.

15. Fusco A, Fedele M (2007) Roles of HMGA proteins in cancer. *Nature reviews. Cancer* 7(12):899-910.

16. Rogalla P, *et al.* (1997) Expression of HMGI-C, a member of the high mobility group protein family, in a subset of breast cancers: relationship to histologic grade. *Mol Carcinog* 19(3):153-156.

17. Abe N, *et al.* (2003) An increased high-mobility group A2 expression level is associated with malignant phenotype in pancreatic exocrine tissue. *Br J Cancer* 89(11):2104-2109.

18. Miyazawa J, Mitoro A, Kawashiri S, Chada KK, Imai K (2004) Expression of mesenchyme-specific gene HMGA2 in squamous cell carcinomas of the oral cavity. *Cancer Res* 64(6):2024-2029.

19. Meyer B, *et al.* (2007) HMGA2 overexpression in non-small cell lung cancer. *Mol Carcinog* 46(7):503-511.

20. Benson KF, Chada K (1994) Mini-mouse: phenotypic characterization of a transgenic insertional mutant allelic to pygmy. *Genet Res* 64(1):27-33.

21. Yun J, *et al.* (2011) Signalling pathway for RKIP and Let-7 regulates and predicts metastatic breast cancer. *EMBO J* 30(21):4500-4514.

22. Dangi-Garimella S, *et al.* (2009) Raf kinase inhibitory protein suppresses a metastasis signalling cascade involving LIN28 and let-7. *EMBO J* 28(4):347-358.

23. Kang Y, *et al.* (2003) A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* 3(6):537-549.

24. Shah N, Sukumar S (2010) The Hox genes and their roles in oncogenesis. *Nat Rev Cancer* 10(5):361-371.

25. Enerly E, *et al.* (2011) miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS One* 6(2):e16915.

26. Morishita A, *et al.* (2013) HMGA2 is an in vivo driver of tumor metastasis. *Cancer Res* In press.

27. Wu H, *et al.* (2011) Dual functions of Tet1 in transcriptional regulation in mouse embryonic stem cells. *Nature* 473(7347):389-393.

28. Wu H, Zhang Y (2011) Tet1 and 5-hydroxymethylation: A genome-wide view in mouse embryonic stem cells. *Cell Cycle* 10(15):2428-2436.

29. Bhutani N, Burns DM, Blau HM (2011) DNA demethylation dynamics. *Cell* 146(6):866-872.

30. Cortellino S, *et al.* (2011) Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146(1):67-79.

31. Gilbert PM, *et al.* (2010) HOXA9 regulates BRCA1 expression to modulate human breast tumor phenotype. *The Journal of clinical investigation* 120(5):1535-1550.

32. Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4(1):44-57.

33. Subramanian A, *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(43):15545-15550.

34. Han HJ, Russo J, Kohwi Y, Kohwi-Shigematsu T (2008) SATB1 reprogrammes gene expression to promote breast tumour growth and metastasis. *Nature* 452(7184):187-193.

35. Hsu CH, *et al.* (2012) TET1 suppresses cancer invasion by activating the tissue inhibitors of metalloproteinases. *Cell reports* 2(3):568-579.

36. Ono R, *et al.* (2002) LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer research* 62(14):4075-4080.

37. Faber J, *et al.* (2009) HOXA9 is required for survival in human MLL-rearranged acute leukemias. *Blood* 113(11):2375-2385.

38. Novak P, *et al.* (2006) Epigenetic inactivation of the HOXA gene cluster in breast cancer. *Cancer Res* 66(22):10664-10670.

39. Dreszer TR, *et al.* (2012) The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res* 40(Database issue):D918-923.

[00190] The invention has been described in an illustrative manner and it is to be understood the terminology used is intended to be in the nature of description rather than of limitation. All patents and other references cited herein are incorporated herein by reference in their entirety. It is also understood that many modifications, equivalents, and variations of the present disclosure are possible in light of the above teachings. Therefore, it is to be understood that within the scope of the appended claims, the invention may be practiced other than as specifically described.

CLAIMS

We claim:

1. A method of diagnosing a subject, the method comprises:

analyzing a biological sample from the subject for expression of HMGA2, TET1, HOXA7, and HOXA9;

comparing expression levels of the HMGA2, TET1, HOXA7, and HOXA9 in the biological sample to a respective expression reference level of HMGA2, TET1, HOXA7, and HOXA9 in a control sample; and

diagnosing the subject with at least one of a decreased survival rate, a poor prognosis, a faster progression of a cancer, and a higher risk of relapse of a cancer, when the expression level of HMGA2 in the biological sample is higher than that of the respective control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are lower than the respective expression reference level of the control sample.

2. The method of claim 1 further comprising administering a therapeutically-effective amount of at least one of a chemotherapeutic agent or a radiotherapeutic agent to the diagnosed subject.

3. The method of claim 2, wherein the chemotherapeutic agent is a DNA demethylation agent.

4. The method of claim 3, wherein the DNA demethylation agent is at least one of azacitidine and decitabine.

5. The method of claim 2, wherein the subject exhibits at least one of an increased chance of survival, a better prognosis, a slower progression of the cancer, and a lower risk of relapse of the cancer as compared to the diagnosis before the administering of the therapeutically-effective amount of at least one of the chemotherapeutic agent or the radiotherapeutic agent.

6. The method of claim 1 further comprising diagnosing the subject with at least one of an increase survival rate, a better prognosis, a slower progression of the cancer, and a lower risk of relapse of the cancer if the expression level of HMGA2 in the biological sample is lower than that of the respective control sample and the expression levels of TET1, HOXA7, and

HOXA9 in the biological sample are higher than the respective expression reference level of the control sample.

7. The method of claim 1, wherein the cancer is breast cancer.

8. The method of claim 1, wherein the expression levels are determined by quantifying at least one of respective expression of a mRNA encoding HMGA2, TET1, HOXA7, or HOXA9; and a respective quantity of a nucleic acid of at least one of HMGA2, TET1, HOXA7, and HOXA9; or a respective functional fragment or variant thereof.

9. The method of claim 8, wherein the expression levels are determined immunochemically and based on an antibody-based detection system.

10. The method of claim 9, wherein the antibody binds specifically to a protein of at least one of HMGA2, TET1, HOXA7, and HOXA9, or a fragment thereof.

11. The method of claim 1, wherein the control sample is obtained from the subject.

12. The method of claim 1, wherein the control sample is obtained from a tissue not diagnosed with the cancer.

13. The method of claim 1, wherein the HMGA2 expression in the biological sample is greater than about 20% of the expression reference level of HMGA2 in the control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are about 20% lower than that of the respective expression reference level of the control sample.

14. A method of treating cancer in a subject in need thereof, the method comprises:

obtaining a biological sample from the subject;

measuring expression levels of HMGA2, TET1, HOXA7, and HOXA9 in the biological sample;

comparing the expression levels of HMGA2, TET1, HOXA7, and HOXA9 in the biological sample to a respective reference level of HMGA2, TET1, HOXA7, and HOXA9 in a control sample; and

administering a therapeutically-effective amount of at least one of a chemotherapeutic agent or a radiotherapeutic agent to the subject when the expression level of HMGA2 in the biological sample is higher than that of the respective control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are lower than the respective expression reference level of the control sample.

15. The method of claim 14, wherein the expression level of HMGA2 in the biological sample is at least 20% higher than that of the respective control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are at least 20% lower than the respective expression reference level of the control sample.

16. The method of claim 14, wherein the therapeutically-effective amount of at least one of the chemotherapeutic agent or the radiotherapeutic agent decreases the expression of HMGA2 in the subject as compared to the respective expression reference level of the control sample.

17. The method of claim 14, wherein the therapeutically-effective amount of at least one of the chemotherapeutic agent and the radiotherapeutic agent increases expression of at least one of TET1, HOXA7, and HOXA9 in the subject as compared to the respective expression reference level of the control sample.

18. The method of claim 14, wherein the cancer is breast cancer.

19. A method of determining a prognosis for survival of a subject diagnosed with a cancer, the method comprises:

analyzing a biological sample from the subject for expression of HMGA2, TET1, HOXA7, and HOXA9;

comparing expression levels of the HMGA2, TET1, HOXA7, and HOXA9 in the biological sample to a respective reference level of HMGA2, TET1, HOXA7, and HOXA9 in a control sample; and

diagnosing the subject with a poor prognosis if the expression level of HMGA2 in the biological sample is higher than that of the respective control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are lower than the respective expression reference level of the control sample; or a better prognosis if the expression level of HMGA2 in the biological sample is lower than that of the respective control sample and the expression levels of TET1, HOXA7, and HOXA9 in the biological sample are higher than the respective expression reference level of the control sample.

20. The method of claim 19 further comprising at least one of administering to the subject diagnosed with the poor prognosis a therapeutically-effective amount of at least one of a chemotherapeutic agent or a radiotherapeutic agent; or a kit comprising at least one agent to detect the expression levels of at least one of the HMGA2, TET1, HOXA7, and HOXA9 in the biological sample, and, optionally, instructions for use.

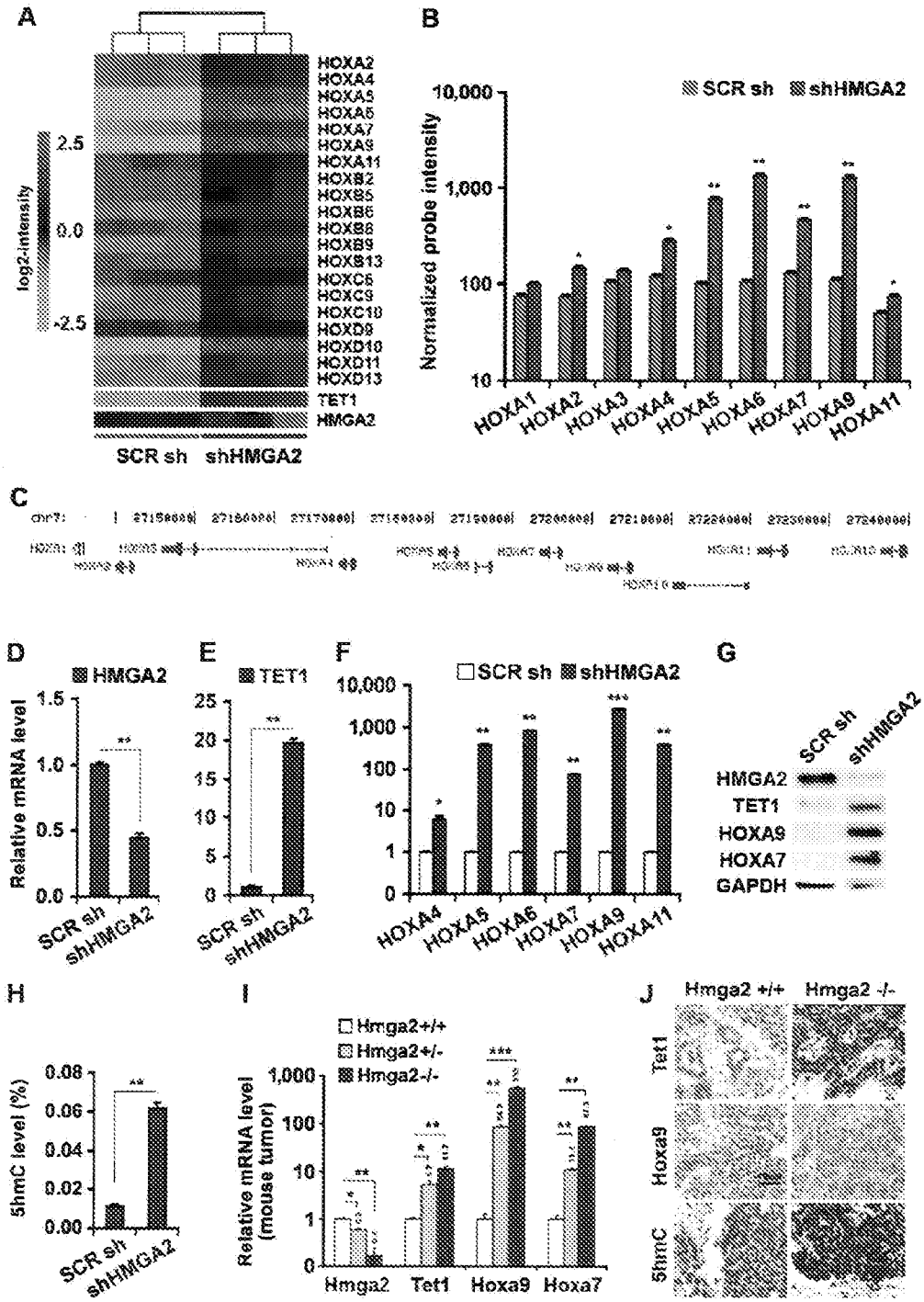
A
HOXA2
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA11
HOXB2
HOXB5
HOXB6
HOXB8
HOXB9
HOXB13
HOXC5
HOXC9
HOXC10
HOXD9
HOXD10
HOXD11
HOXD13
TET1
HMGA2
2.5
0.0
-2.5
log2-intensity
SCR sh
shHMGA2
B
SCR sh
shHMGA2
Normalized probe intensity
10,000
1,000
100
10
HOXA1
HOXA2
HOXA3
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA11
C
D
HMGA2
Relative mRNA level
1.5
1.0
0.5
0.0
SCR sh
shHMGA2
E
TET1
25
20
15
10
5
0
SCR sh
shHMGA2
F
SCR sh
shHMGA2
10,000
1,000
100
10
1
0
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA11
G
SCR sh
shHMGA2
HMGA2
TET1
HOXA9
HOXA7
GAPDH
H
5hmC level (%)
0.08
0.06
0.04
0.02
0.00
SCR sh
shHMGA2
I
Hmga2+/+
Hmga2+/-
Hmga2-/-
Relative mRNA level (mouse tumor)
1,000
100
10
1
0
Hmga2
Tet1
Hoxa9
Hoxa7
J
Hmga2 +/+
Hmga2 -/-
Tet1
Hoxa9
5hmC

FIG. 1

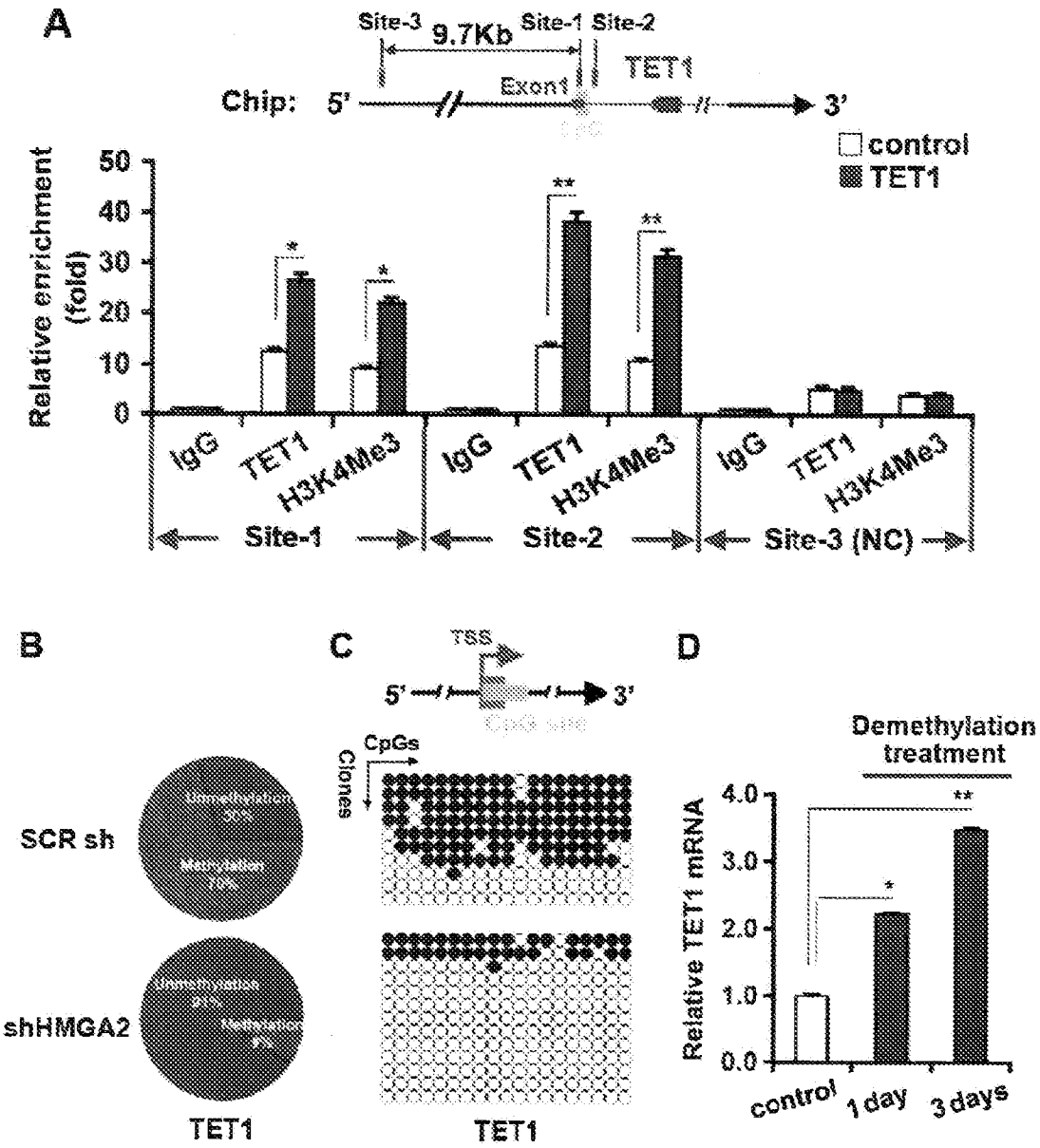
A
Site-3
9.7Kb
Site-1
Site-2
TET1
Exon1
Chip:
5'
3'
control
TET1
Relative enrichment (fold)
50
40
30
20
10
0
IgG
TET1
H3K4Me3
Site-1
Site-2
Site-3 (NC)
B
SCR sh
shHMGA2
TET1
C
TSS
5'
3'
CpG site
CpGs
Clones
TET1
D
Demethylation treatment
Relative TET1 mRNA
4.0
3.0
2.0
1.0
0.0
control
1 day
3 days

FIG. 2

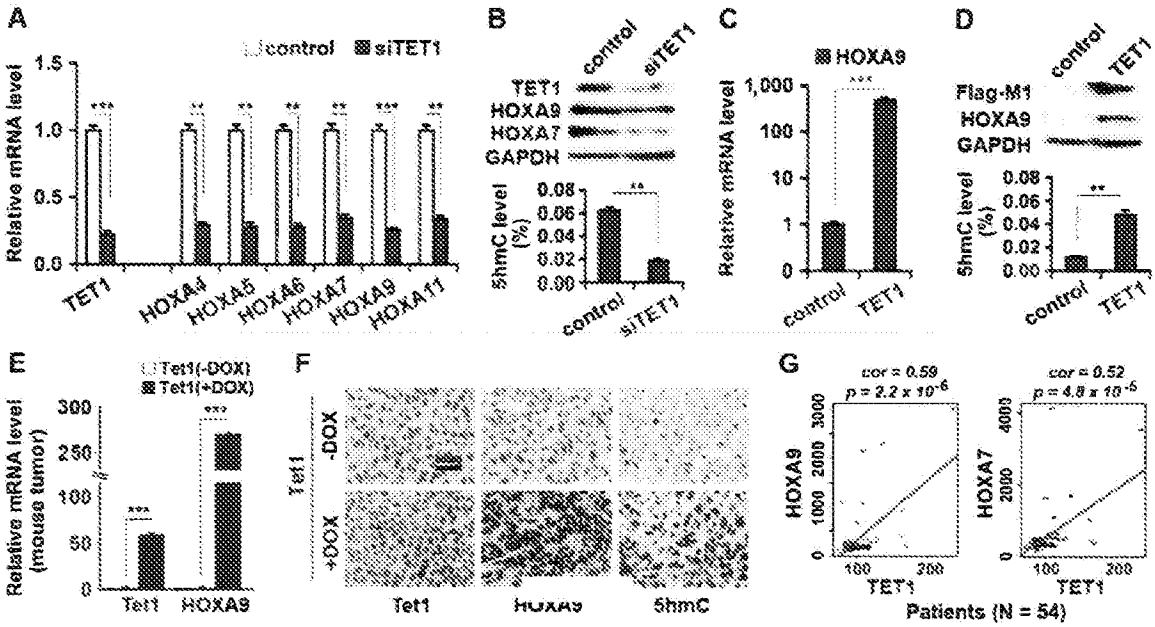
A
control
siTET1
Relative mRNA level
TET1
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA11
B
control
siTET1
TET1
HOXA9
HOXA7
GAPDH
5hmC level (%)
C
HOXA9
Relative mRNA level
D
control
TET1
Flag-M1
HOXA9
GAPDH
5hmC level (%)
E
Tet1(-DOX)
Tet1(+DOX)
Relative mRNA level (mouse tumor)
Tet1
HOXA9
F
Tet1
-DOX
+DOX
Tet1
5hmC
G
cor = 0.59
p = 2.2 x 10^-6
cor = 0.52
p = 4.8 x 10^-5
HOXA9
HOXA7
TET1
TET1
Patients (N = 54)

FIG. 3

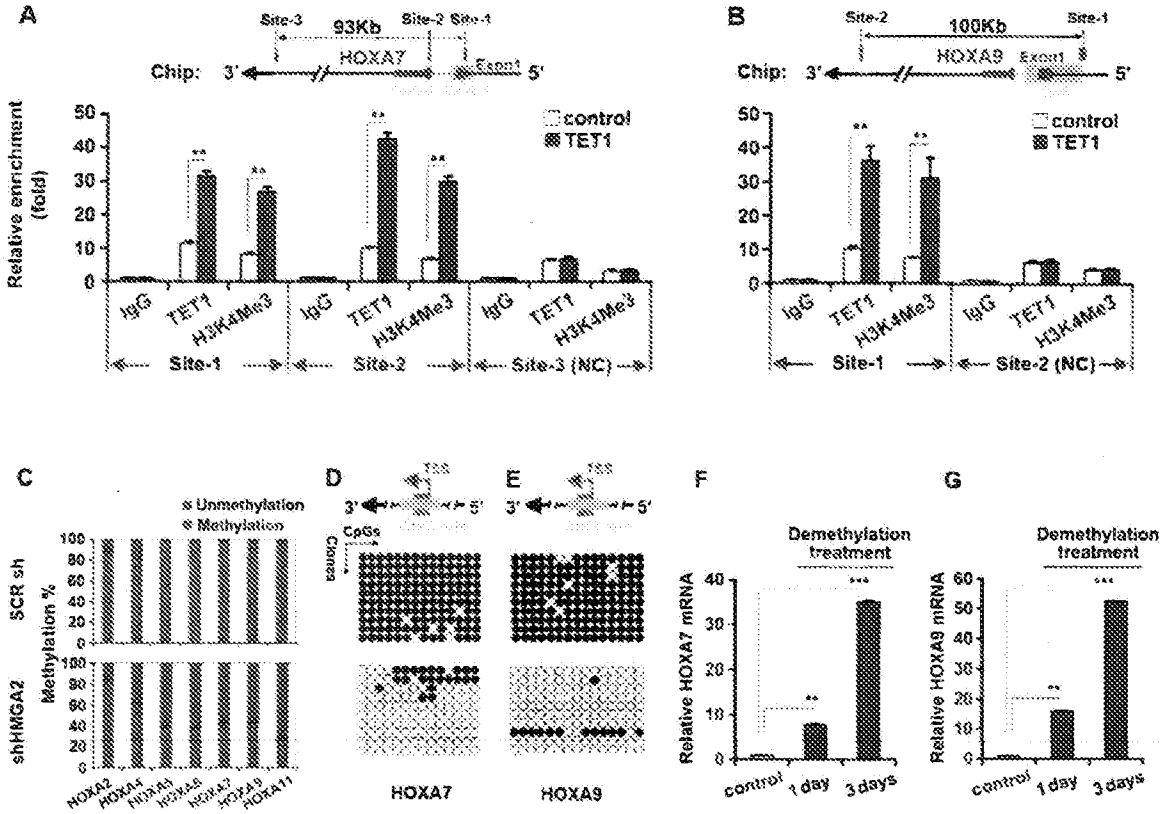
A
Site-3
93Kb
Site-2
Site-1
HOXA7
Exon1
Chip: 3'
5'
Relative enrichment (fold)
control
TET1
IgG
TET1
H3K4Me3
Site-1
Site-2
Site-3 (NC)
B
Site-2
100Kb
Site-1
HOXA9
Exon1
Site-2 (NC)
C
Unmethylation
Methylation
SCR sh
shHMGA2
Methylation %
HOXA2
HOXA4
HOXA5
HOXA6
HOXA7
HOXA9
HOXA11
D
E
CpGs
Clones
HOXA7
HOXA9
F
Demethylation treatment
Relative HOXA7 mRNA
control
1 day
3 days
G
Demethylation treatment
Relative HOXA9 mRNA
control
1 day
3 days

FIG. 4

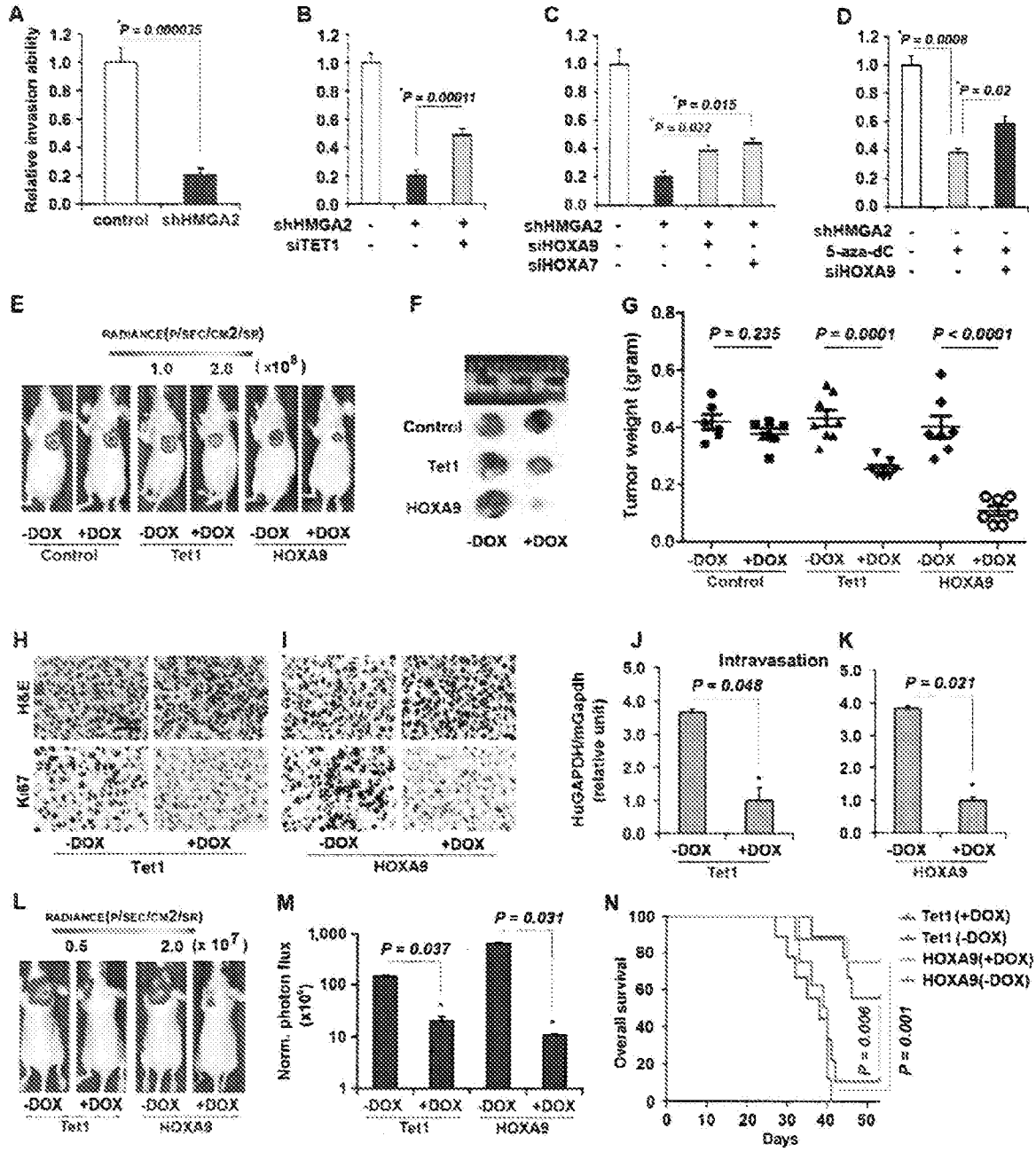
A
Relative invasion ability
P = 0.000035
control
shHMGA2
B
P = 0.00011
shHMGA2
siTET1
C
P = 0.015
P = 0.032
shHMGA2
siHOXA9
siHOXA7
D
P = 0.0008
P = 0.02
shHMGA2
5-aza-dC
siHOXA9
E
RADIANCE(P/SEC/CM2/SR)
1.0
2.8
(x10^6)
-DOX +DOX -DOX +DOX -DOX +DOX
Control
Tet1
HOXA9
F
Control
Tet1
HOXA9
-DOX +DOX
G
Tumor weight (gram)
P = 0.235
P = 0.0001
P < 0.0001
-DOX +DOX -DOX +DOX -DOX +DOX
Control
Tet1
HOXA9
H
H&E
Ki67
-DOX +DOX
Tet1
I
-DOX +DOX
HOXA9
J
Intravasation
P = 0.048
HuGAPDH/mGapdh (relative unit)
-DOX +DOX
Tet1
K
P = 0.021
-DOX +DOX
HOXA9
L
RADIANCE(P/SEC/CM2/SR)
0.5
2.0 (x 10^7)
-DOX +DOX -DOX +DOX
Tet1
HOXA9
M
P = 0.037
P = 0.031
Norm. photon flux (x10^4)
-DOX +DOX -DOX +DOX
Tet1
HOXA9
N
Overall survival
Days
Tet1(+DOX)
Tet1(-DOX)
HOXA9(+DOX)
HOXA9(-DOX)
P = 0.006
P = 0.001

FIG. 5

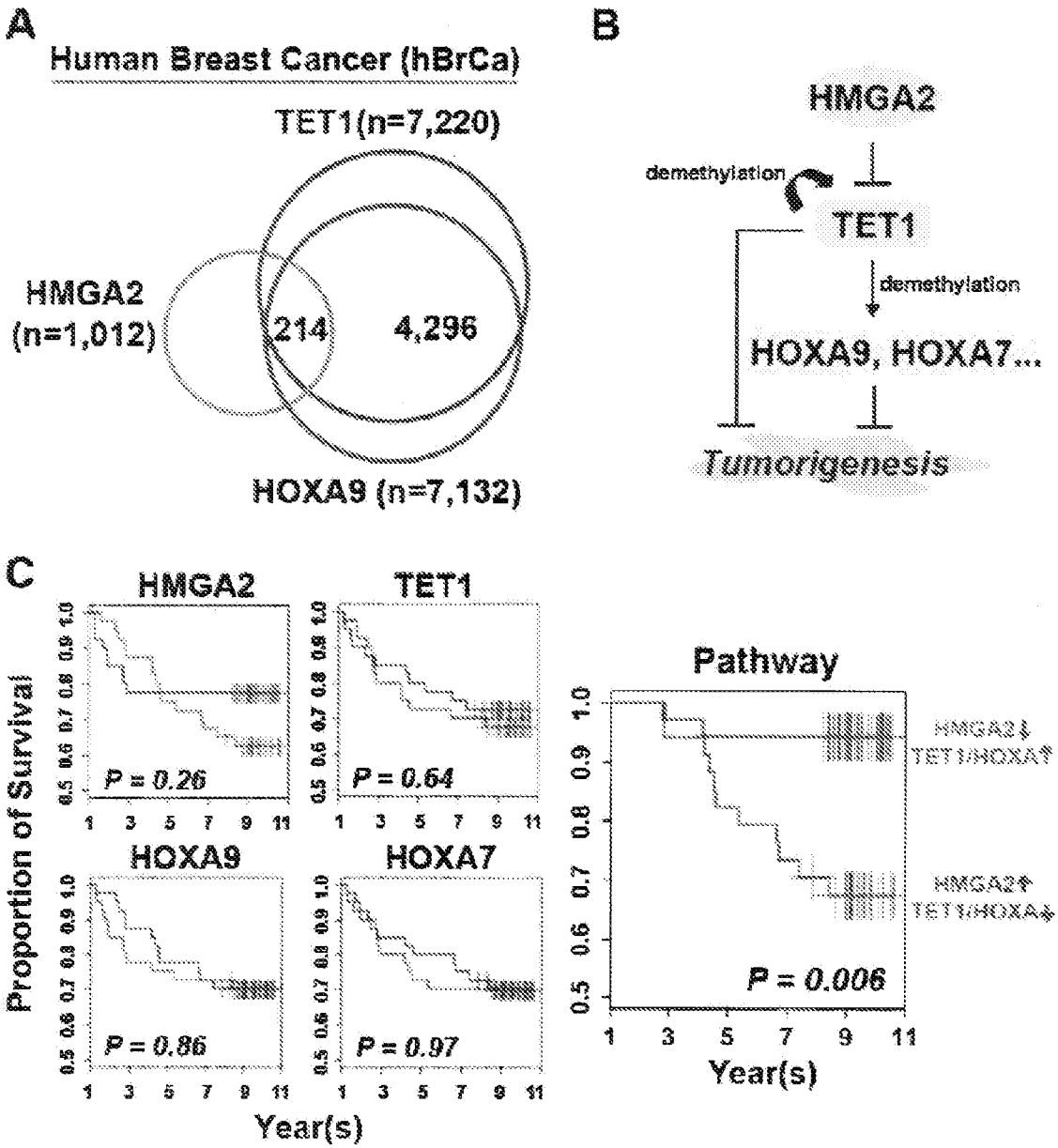
A
Human Breast Cancer (hBrCa)
TET1(n=7,220)
HMGA2
(n=1,012)
214
4,296
HOXA9 (n=7,132)
B
HMGA2
demethylation
TET1
demethylation
HOXA9, HOXA7...
Tumorigenesis
C
HMGA2
TET1
HOXA9
HOXA7
P = 0.26
P = 0.64
P = 0.86
P = 0.97
Proportion of Survival
Year(s)
Pathway
HMGA2↓
TET1/HOXA↑
HMGA2↑
TET1/HOXA↓
P = 0.006
Year(s)

FIG. 6

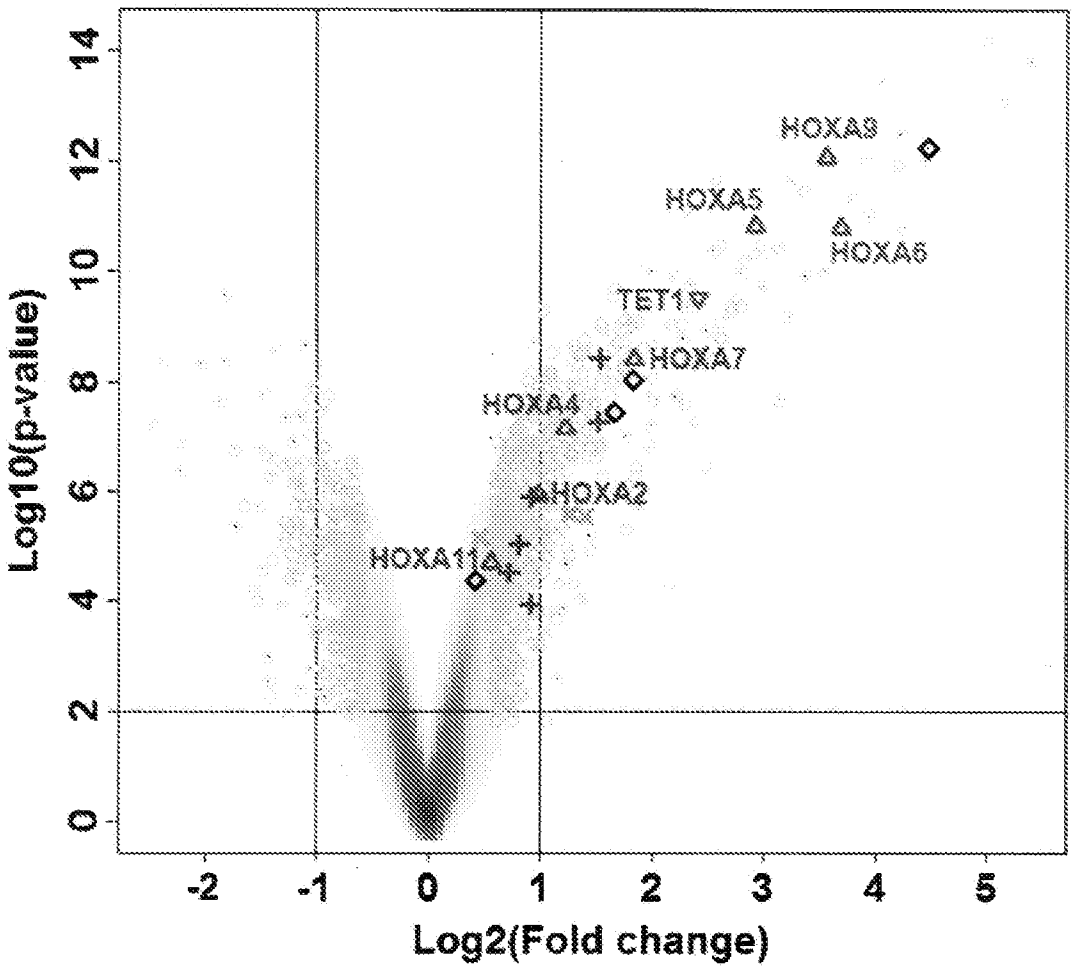
HOXA9
HOXA5
HOXA6
TET1
HOXA7
HOXA4
HOXA2
HOXA11
Log10(p-value)
Log2(Fold change)
0
2
4
6
8
10
12
14
-2
-1
0
1
2
3
4
5

FIG. S1

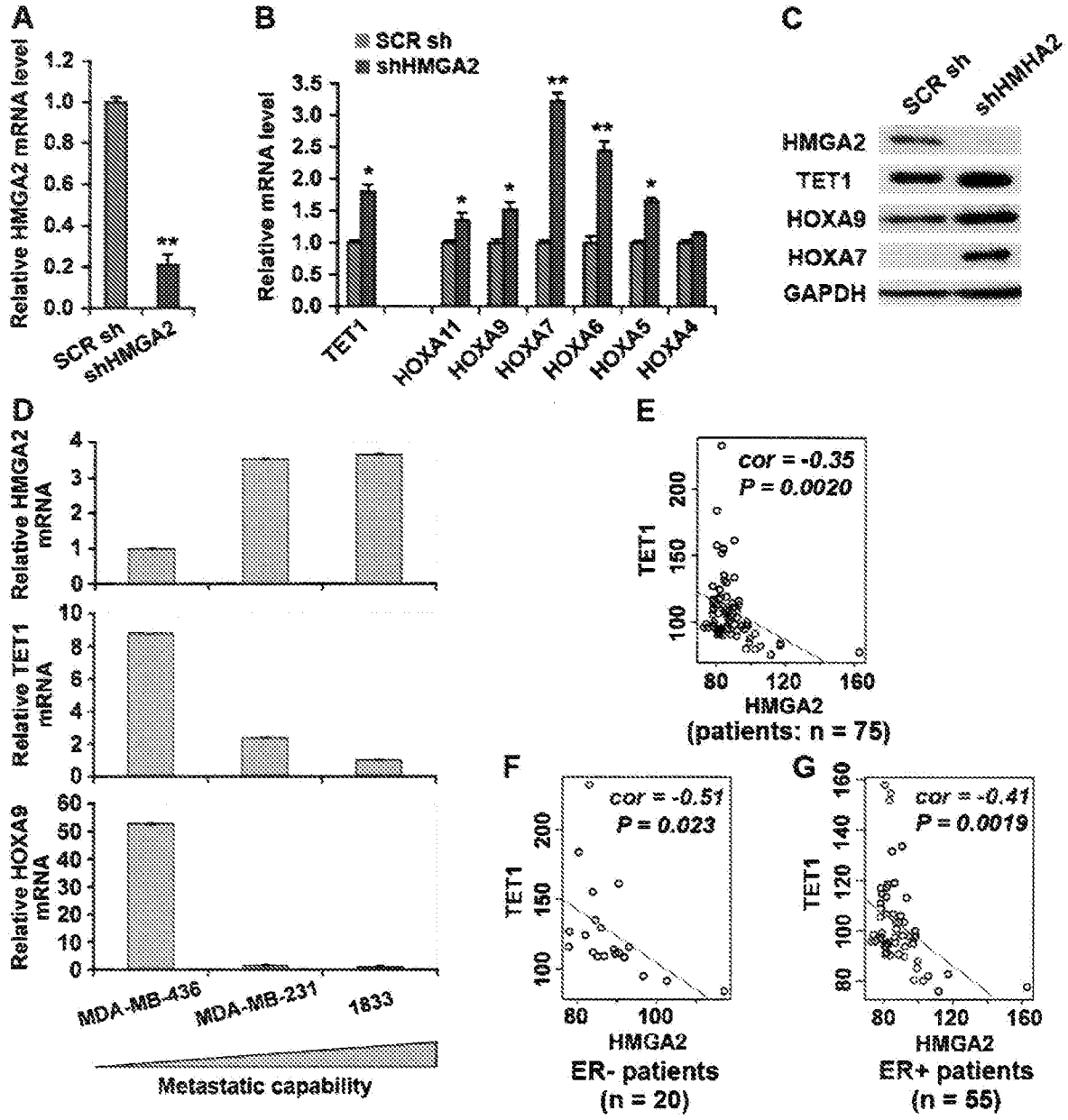
A
Relative HMGA2 mRNA level
SCR sh
shHMGA2
B
SCR sh
shHMGA2
Relative mRNA level
TET1
HOXA11
HOXA9
HOXA7
HOXA6
HOXA5
HOXA4
C
SCR sh
shHMHA2
HMGA2
TET1
HOXA9
HOXA7
GAPDH
D
Relative HMGA2 mRNA
Relative TET1 mRNA
Relative HOXA9 mRNA
MDA-MB-436
MDA-MB-231
1833
Metastatic capability
E
cor = -0.35
P = 0.0020
TET1
HMGA2
(patients: n = 75)
F
cor = -0.51
P = 0.023
TET1
HMGA2
ER- patients
(n = 20)
G
cor = -0.41
P = 0.0019
TET1
HMGA2
ER+ patients
(n = 55)

FIG. S2

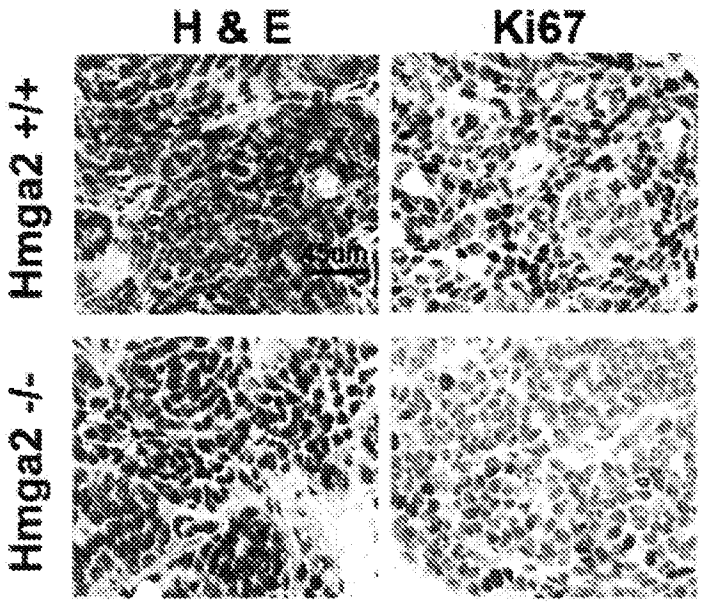

FIG. S3

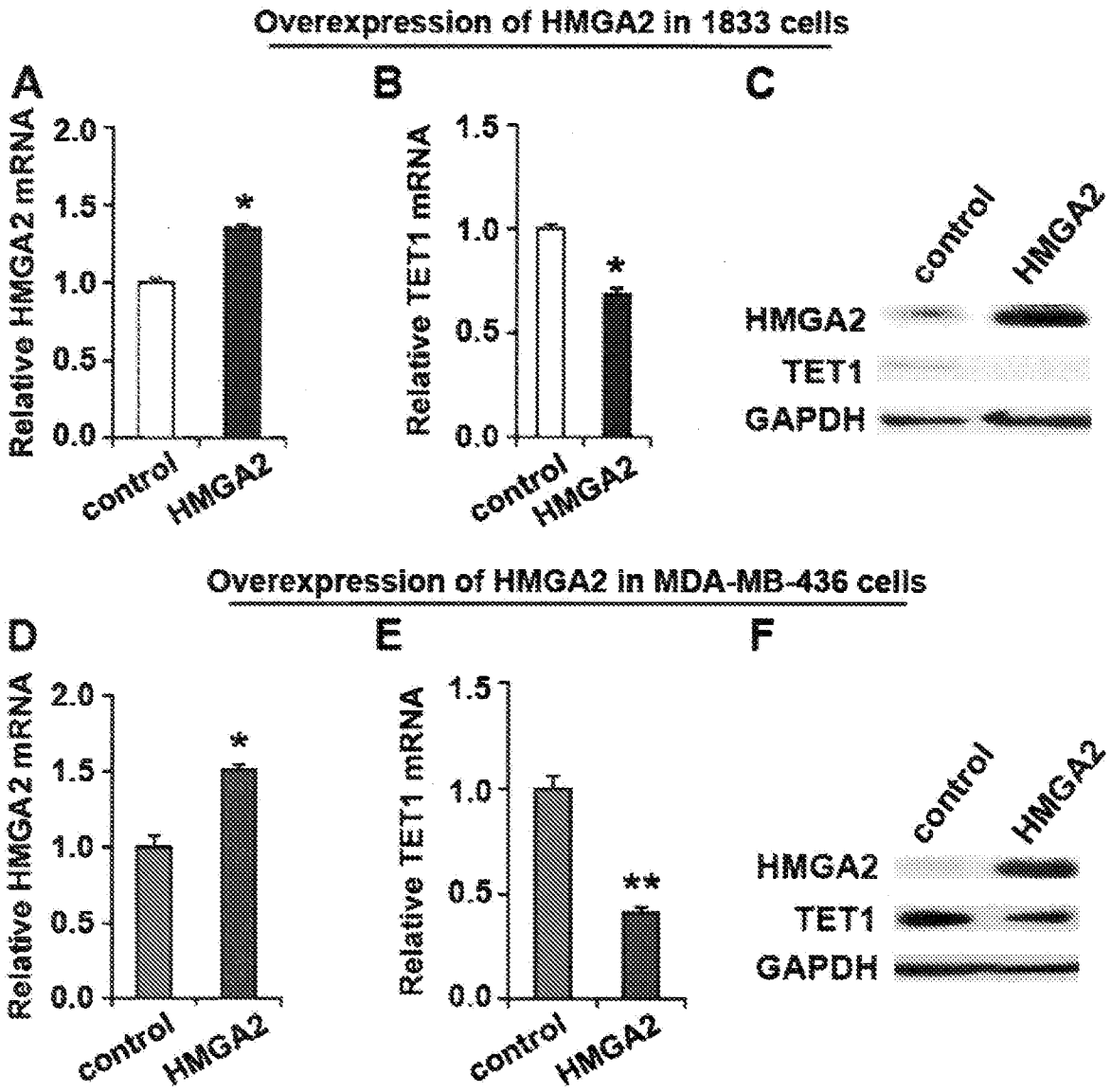
Overexpression of HMGA2 in 1833 cells
A
Relative HMGA2 mRNA
2.0
1.5
1.0
0.5
0.0
*
control
HMGA2
B
Relative TET1 mRNA
1.5
1.0
0.5
0.0
*
control
HMGA2
C
control
HMGA2
HMGA2
TET1
GAPDH
Overexpression of HMGA2 in MDA-MB-436 cells
D
Relative HMGA2 mRNA
2.0
1.5
1.0
0.5
0.0
*
control
HMGA2
E
Relative TET1 mRNA
1.5
1.0
0.5
0.0
**
control
HMGA2
F
control
HMGA2
HMGA2
TET1
GAPDH

FIG. S4

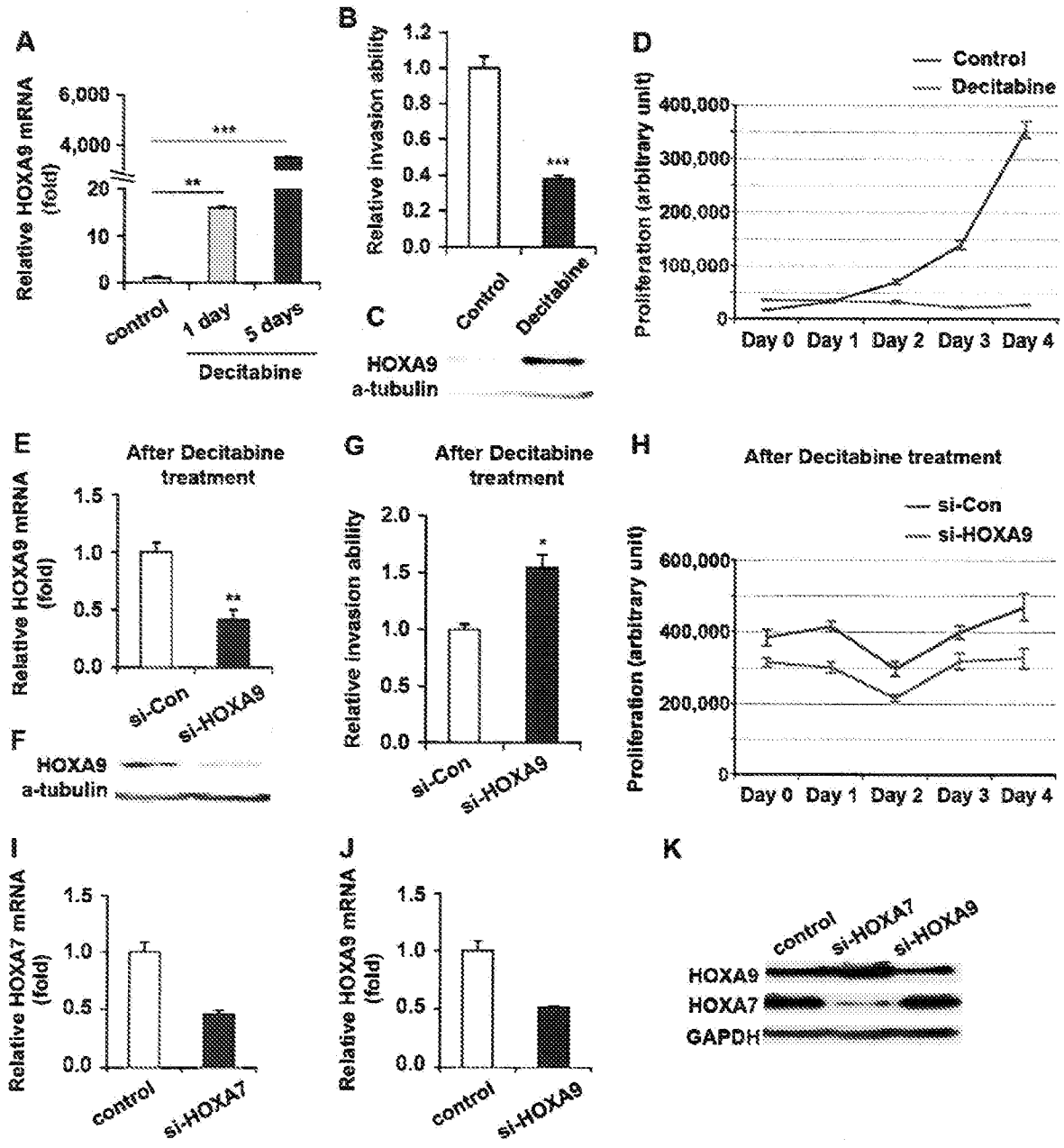
A
Relative HOXA9 mRNA (fold)
6,000
4,000
20
10
0
***
**
control
1 day
5 days
Decitabine
B
Relative invasion ability
1.2
1.0
0.8
0.6
0.4
0.2
0.0
***
Control
Decitabine
C
HOXA9
a-tubulin
D
Control
Decitabine
Proliferation (arbitrary unit)
400,000
300,000
200,000
100,000
0
Day 0 Day 1 Day 2 Day 3 Day 4
E
After Decitabine treatment
Relative HOXA9 mRNA (fold)
1.5
1.0
0.5
0.0
**
si-Con
si-HOXA9
F
HOXA9
a-tubulin
G
After Decitabine treatment
Relative invasion ability
2.0
1.5
1.0
0.5
0.0
*
si-Con
si-HOXA9
H
After Decitabine treatment
si-Con
si-HOXA9
Proliferation (arbitrary unit)
600,000
400,000
200,000
0
Day 0 Day 1 Day 2 Day 3 Day 4
I
Relative HOXA7 mRNA (fold)
1.5
1.0
0.5
0.0
control
si-HOXA7
J
Relative HOXA9 mRNA (fold)
1.5
1.0
0.5
0.0
control
si-HOXA9
K
control
si-HOXA7
si-HOXA9
HOXA9
HOXA7
GAPDH

FIG. S5

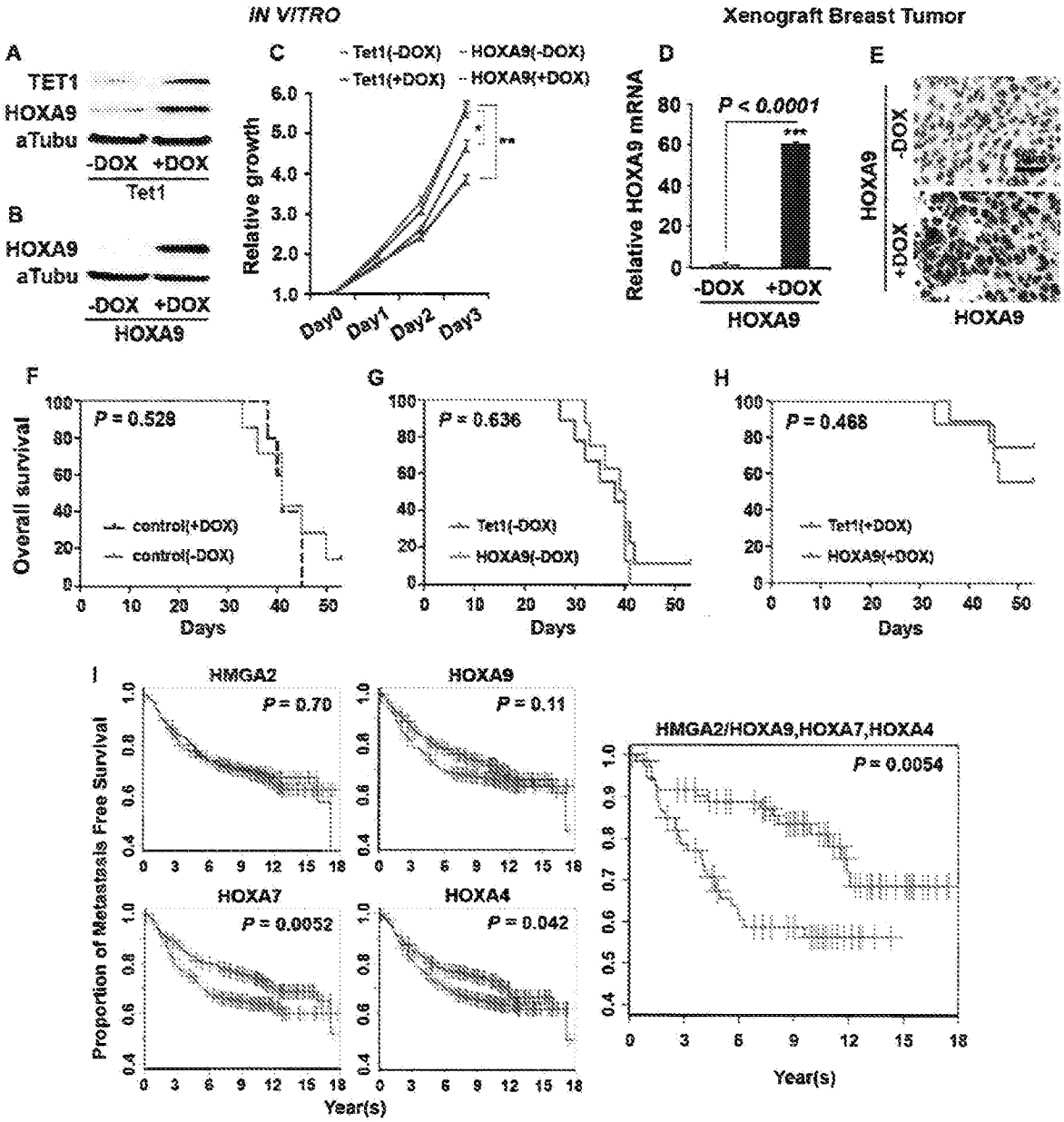
IN VITRO
Xenograft Breast Tumor
A
TET1
HOXA9
aTubu
-DOX +DOX
Tet1
B
HOXA9
aTubu
-DOX +DOX
HOXA9
C
Tet1(-DOX)
Tet1(+DOX)
HOXA9(-DOX)
HOXA9(+DOX)
Relative growth
Day0 Day1 Day2 Day3
D
P < 0.0001
Relative HOXA9 mRNA
-DOX +DOX
HOXA9
E
HOXA9
-DOX
+DOX
HOXA9
F
P = 0.528
Overall survival
control(+DOX)
control(-DOX)
Days
G
P = 0.636
Tet1(-DOX)
HOXA9(-DOX)
Days
H
P = 0.468
Tet1(+DOX)
HOXA9(+DOX)
Days
I
HMGA2
P = 0.70
HOXA9
P = 0.11
HOXA7
P = 0.0052
HOXA4
P = 0.042
Proportion of Metastasis Free Survival
Year(s)
HMGA2/HOXA9,HOXA7,HOXA4
P = 0.0054
Year(s)

FIG. S6