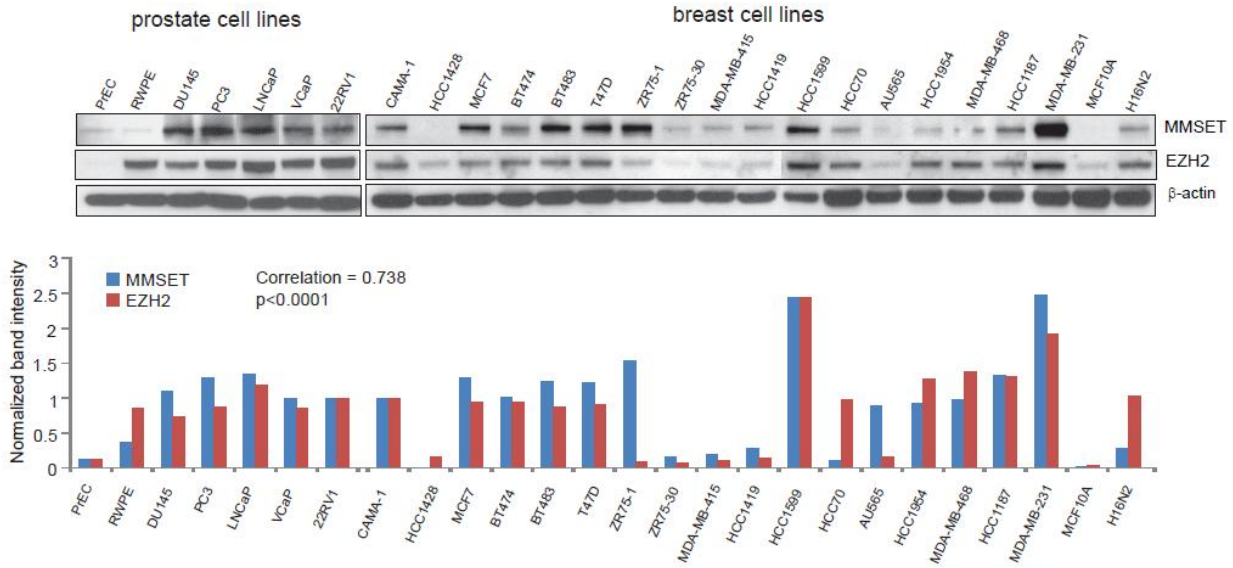
Supplemental Text and Figures

A



B

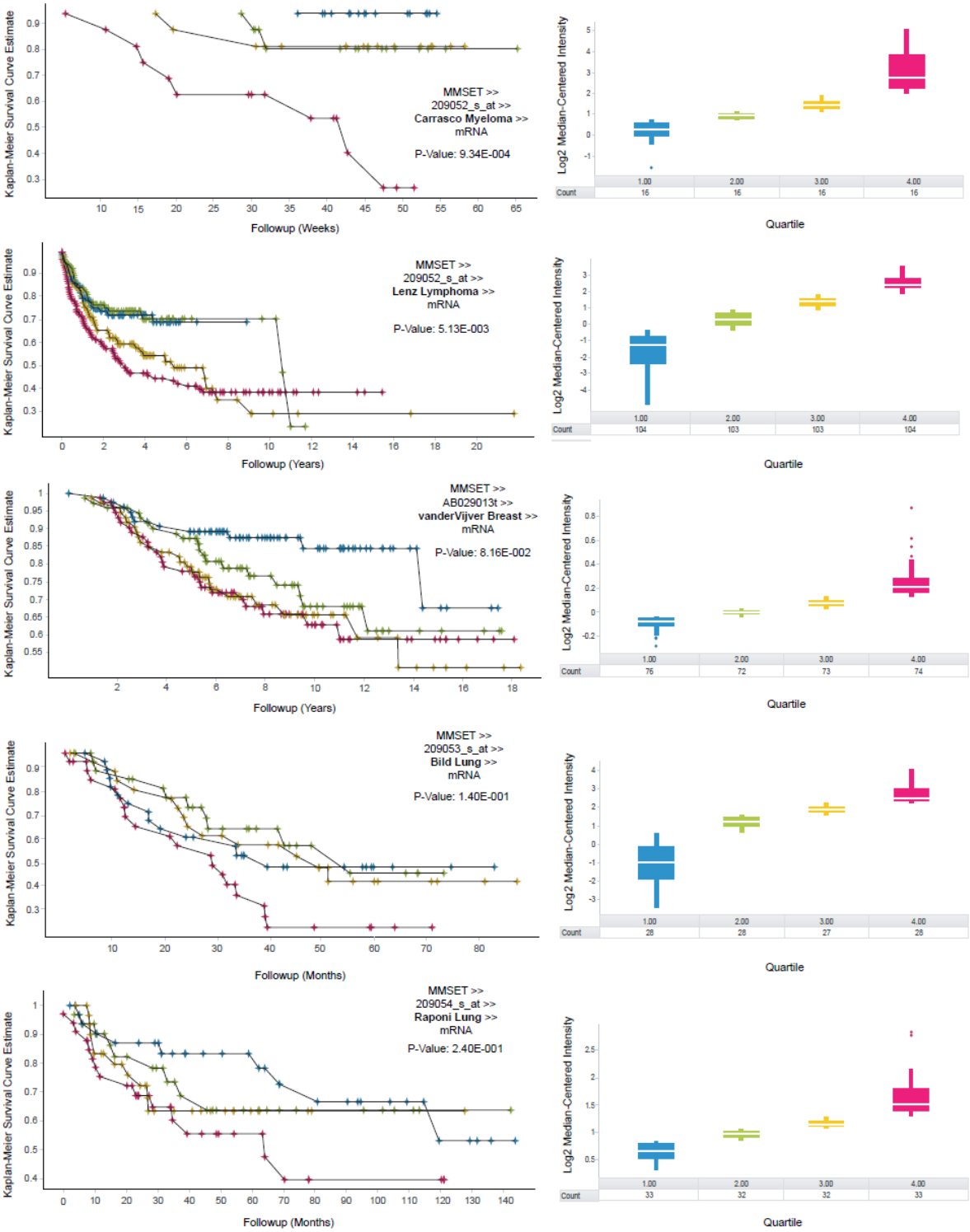

Kaplan-Meier Survival Curve Estimate
MMSET >>
209052_s_at >>
Carrasco Myeloma >>
mRNA
P-Value: 9.34E-004
Followup (Weeks)
Log2 Median-Centered Intensity
1.00 2.00 3.00 4.00
Count 16 16 16 16
Quartile
Kaplan-Meier Survival Curve Estimate
MMSET >>
209052_s_at >>
Lenz Lymphoma >>
mRNA
P-Value: 5.13E-003
Followup (Years)
Log2 Median-Centered Intensity
1.00 2.00 3.00 4.00
Count 104 103 103 104
Quartile
Kaplan-Meier Survival Curve Estimate
MMSET >>
AB029013t >>
vanderVijver Breast >>
mRNA
P-Value: 8.16E-002
Followup (Years)
Log2 Median-Centered Intensity
1.00 2.00 3.00 4.00
Count 76 72 73 74
Quartile
Kaplan-Meier Survival Curve Estimate
MMSET >>
209053_s_at >>
Bild Lung >>
mRNA
P-Value: 1.40E-001
Followup (Months)
Log2 Median-Centered Intensity
1.00 2.00 3.00 4.00
Count 28 28 27 28
Quartile
Kaplan-Meier Survival Curve Estimate
MMSET >>
209054_s_at >>
Raponi Lung >>
mRNA
P-Value: 2.40E-001
Followup (Months)
Log2 Median-Centered Intensity
1.00 2.00 3.00 4.00
Count 33 32 32 33
Quartile

Figure S1. MMSET and EZH2 protein levels in various cell lines, and MMSET expression associates with poor clinical outcome, **Related to Figure 1**.

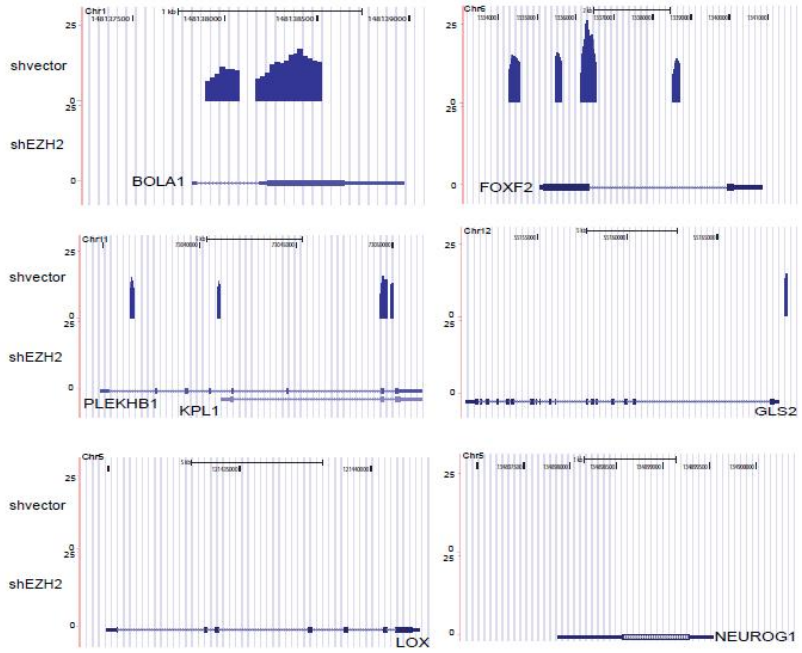
(A) Immunoblot analysis (top panel) demonstrates high correlation between MMSET and EZH2 protein levels in prostate and breast cell lines. β-Actin served as a loading control. Normalized band intensity (ImageJ) is represented in the bar graph (bottom panel). **(B)** MMSET expression is associated with poor clinical outcome in multiple cancer microarray profiling data sets. Samples of each dataset were classified into four quartiles based on the transcript levels of MMSET (right panel) and survival curves were obtained by Kaplan-Meier (KM) analysis (left panel). KM plots are shown for Carrasco et al., 2006, for multiple myeloma, Lenz et al., 2008 for lymphoma, van der Vijver et al., 2002 for breast cancer, and Bild et. al., 2006, Raponi et. al., 2006 for lung cancer data set. The number of cases per dataset, follow-up period, reporter ID, and p values are provided for each study. Analyses were carried out using Oncomine power tool beta.

A

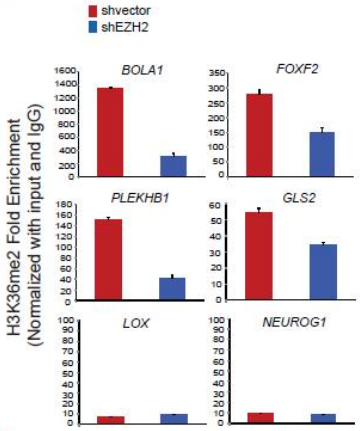
B

File	# of lanes	# of total sequenced reads (millions)	# of reads uniquely mapped (millions)	# of enriched regions
shvector H3K36me2	1	13.05	10.68	42,480 covering 7.81 MB
shEZH2 H3K36me2	1	25.68	21.95	11,789 covering 1.81MB
shvector IgG	1	21.86	21.85	2362 covering 0.60 MB

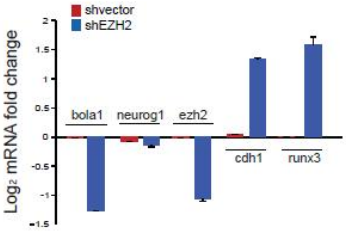
C



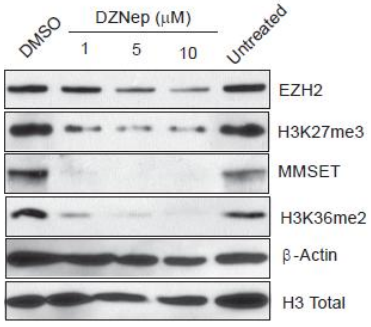
D



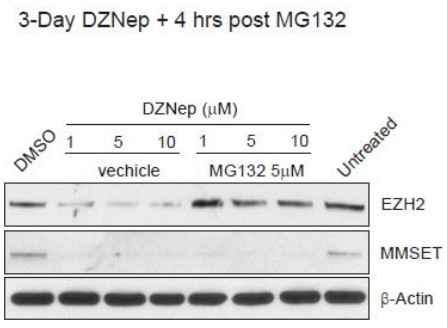
E



F



G



H

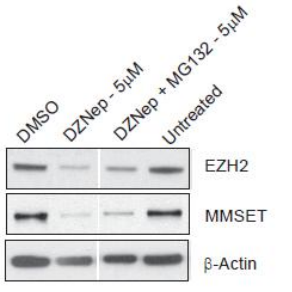
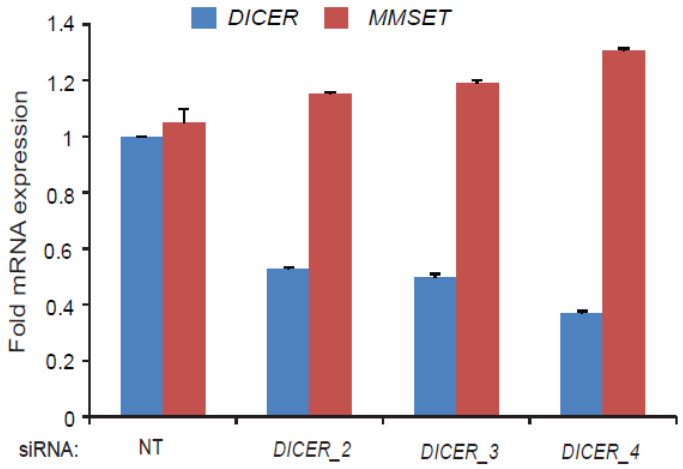


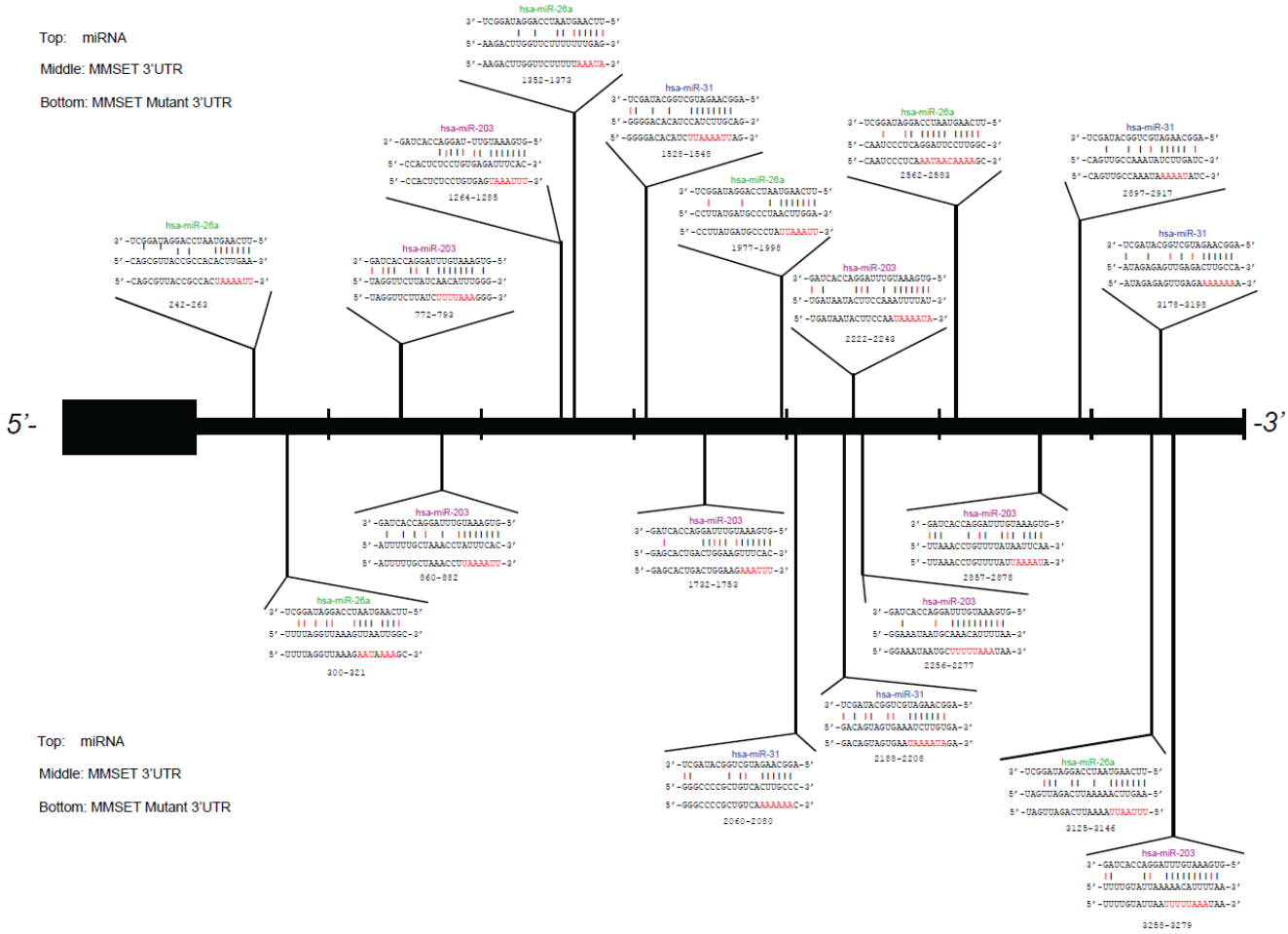
Figure S3. EZH2 depletion leads to diminished global transcription activation mark H3K36me2, Related to Figure 2.

(A) Immunoblot analysis of stable knockdown of EZH2 in MBA-MB-231 cells showing reduced levels of EZH2, MMSET, H3K27me3 and H3K36me2. β-Actin was used as a loading control. **(B)** Table showing high-throughput sequencing read information and enrichment (H peak) carried out for ChIP libraries of H3K36me2 or IgG in control and EZH2 knockdown cells. **(C)** UCSC genome browser view of ChIP-seq H3K36me2 binding peaks on the TSS and gene body of candidate genes in control and EZH2 knockdown cells. LOX and NEUROG1 were used as examples of genes with no enrichment. **(D)** ChIP-qPCR validation of the genes shown in E. **(E)** qRT-PCR for mRNA expression of bola1 and neurog1 in control and EZH2 knockdown cells. The cdh1 and runx3 genes known to be upregulated upon EZH2 knockdown were used as positive control (n=3 ± SEM). **(F)** DZNep treatment reduces MMSET expression and its substrate, H3K36me2 mark in DU145 cells. EZH2 and its substrate H3K27me3 served as controls. **(G)** DZNep treatment reduces EZH2 and not MMSET protein level through proteasome mediated degradation. Immunoblot showing reduction in the EZH2 and MMSET protein upon 3-day DZNep treatment, and a specific induction of EZH2 alone and not MMSET with 4hr incubation of the cells with proteasome inhibitor MG132. **(H)** Longer incubation (48hr) of the cells with MG132 shows slight increase in both EZH2 and MMSET level. β-Actin were used as a loading control.

A



B


Top: miRNA
Middle: MMSET 3'UTR
Bottom: MMSET Mutant 3'UTR
5'-
-3'
hsa-miR-26a
hsa-miR-203
hsa-miR-31
Top: miRNA
Middle: MMSET 3'UTR
Bottom: MMSET Mutant 3'UTR

C

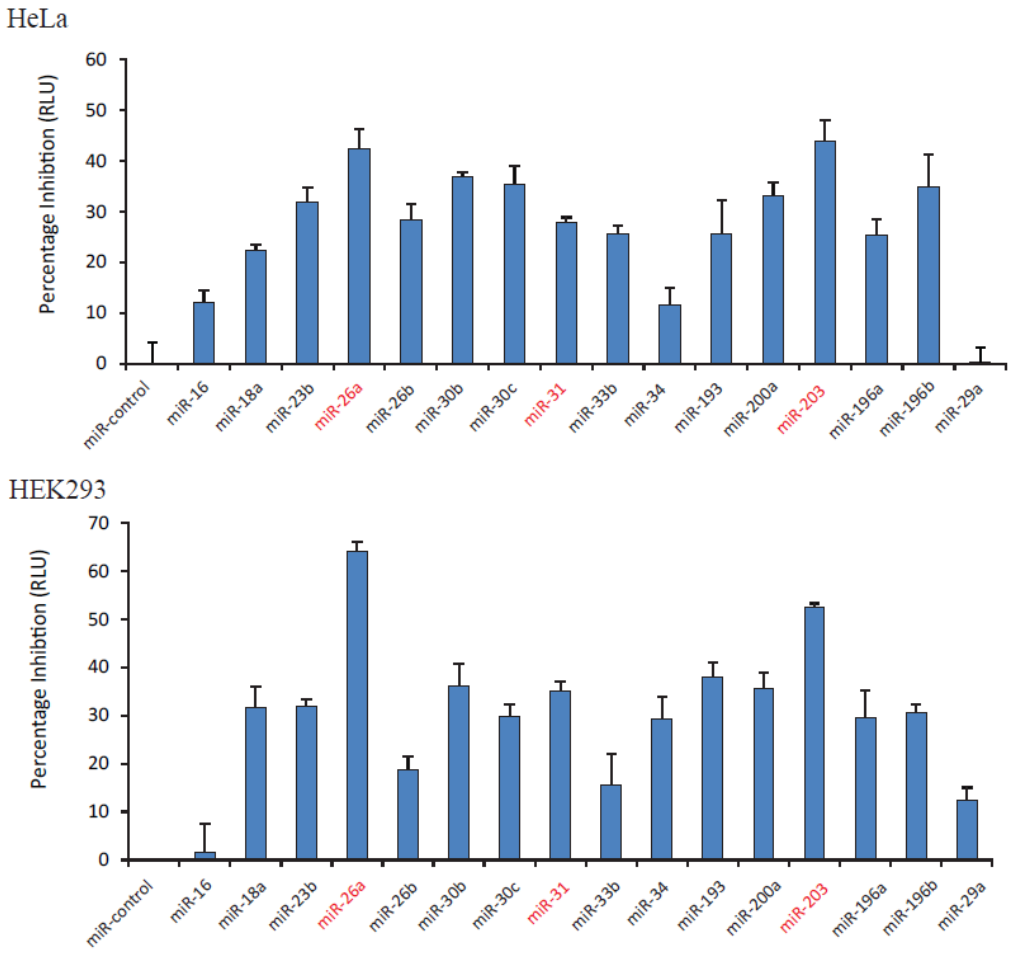


Figure S3. MMSET regulation by microRNAs, Related to Figure 3.

(A) Dicer knockdown has no effect MMSET transcript levels. Dicer and MMSET transcript levels following knockdown of Dicer using three different siRNA duplexes in PC3 prostate cancer cells as determined by qRT-PCR. **(B)** Predicted microRNA binding sites in MMSET 3'UTR. Schematic depicting multiple putative binding sites for miR-203, miR-26a and miR-31 in the 3.3 kb 3'UTR of MMSET predicted by PICTAR, TargetScan and PITA. Mutated sequence of the miR binding sites are shown in red. The full length MMSET wild-type or mutant 3'UTR was cloned downstream to a luciferase reporter in pEZX-MT01 vector for reporter assays. **(C)** Inhibition of MMSET by microRNAs. Luciferase Reporter assays with full length 3'UTR (3.3 kb) of MMSET co-transfected with 15 miRs predicted to bind MMSET, a non-binding miR-29a and a negative control miR in HeLa (top panel) and HEK293 cells (bottom panel). The data is presented as percent inhibition of luciferase activity ±SEM. miR-26a and miRs displaying greatest inhibition are shown in red (miR-31 and miR-203).

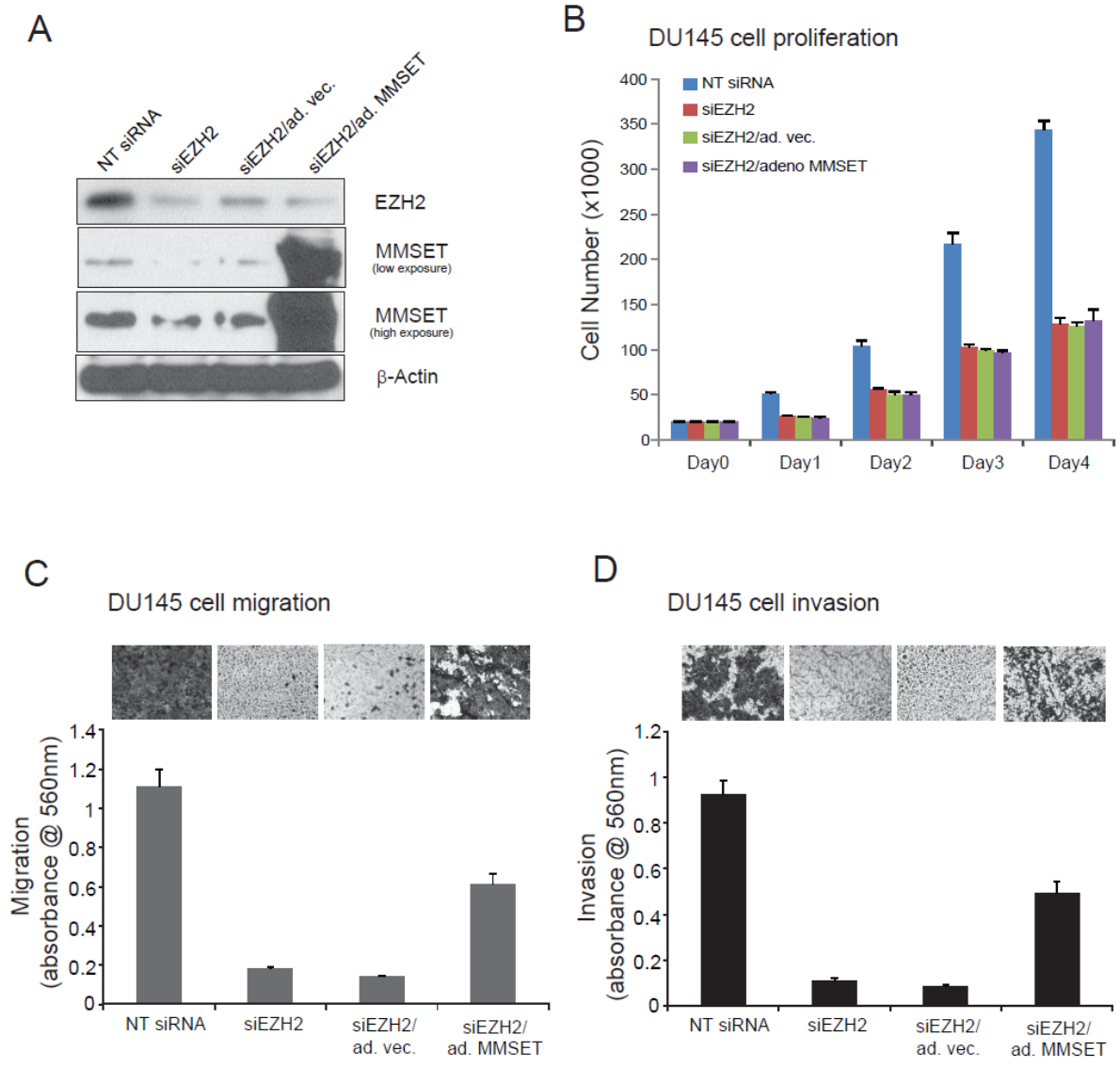
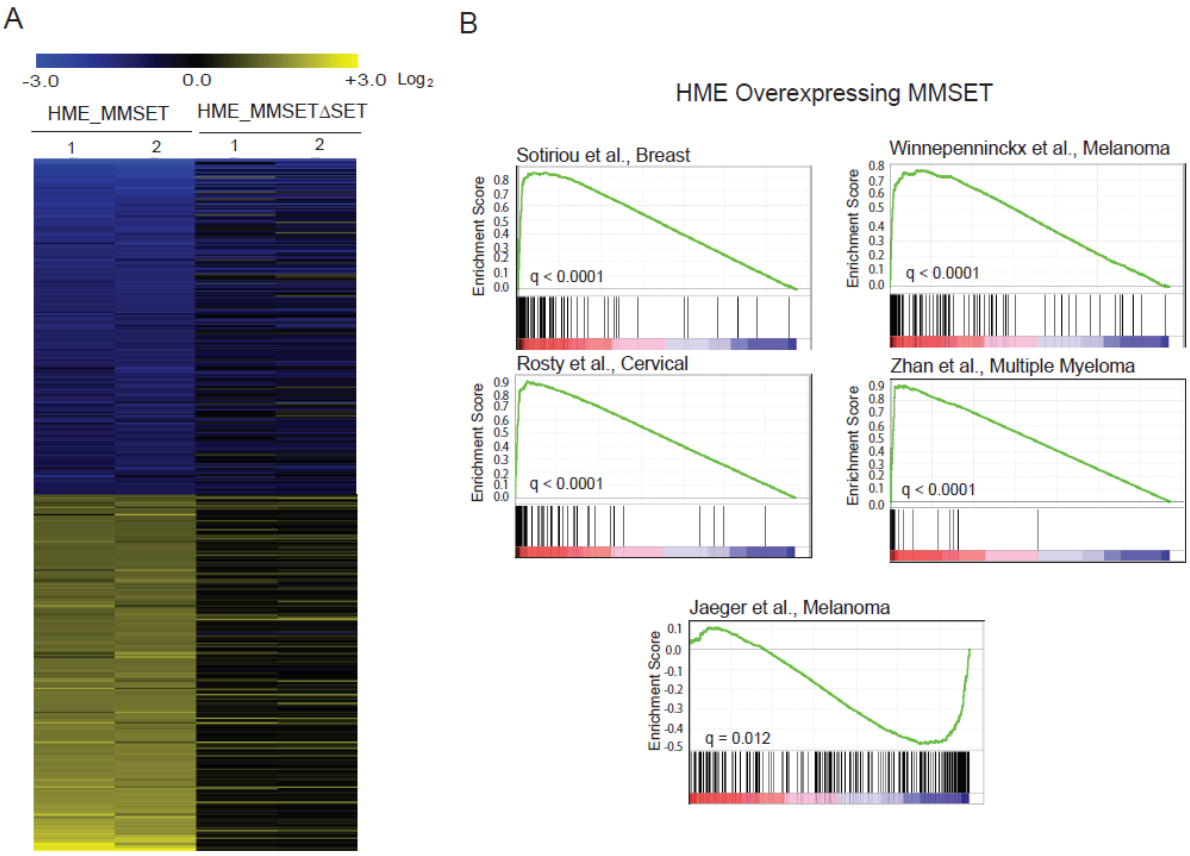


Figure S4. MMSET expression is required for EZH2 mediated migration and invasion but not cell proliferation, Related to Figure 4.

DU145 cells were transfected with contorl siRNA or siEZH2 along with either control adenovirus or MMSET adenovirus infection. **(A)** Immunoblot analysis showing reduction in EZH2 and MMSET levels in siEZH2 transfected cells, and overexpression of MMSET in adeno-MMSET infected cells. β-Actin served as a loading control.**(B-D)** cell proliferation assay by counting cells using Coulter counter (B), cell migration as monitered by ability of the cells to migrate through 0.8 micron membrane (Boyden Chamber) (C), and cell invasion as monitored by modified Boyden chamber-Matrigel invasion assay (D) with representative images of migrated and invaded cells shown in the inset.



C

Gene Set	n	HME MMSET NES	HME MMSET FDR q-value	HME MMSETΔSET NES	HME MMSETΔSET FDR q-value
SOTIRIOU_BREAST_CANCER_GRADE_1_VS_3_UP	123	3.15	<0.0001	0.88	1.00
ROSTY_CERVICAL_CANCER_PROLIFERATION_CLUSTER_UP	114	3.30	<0.0001	0.71	1.00
WINNEPENNINCKX_MELANOMA_METASTASIS_UP	121	2.88	<0.0001	0.77	1.00
KAUFFMANN_MELANOMA_RELAPSE_UP	55	2.75	<0.0001	0.53	1.00
SHEDDEN_LUNG_CANCER_POOR_SURVIVAL_UP	352	3.09	<0.0001	1.10	0.76
SENGUPTA_NASOPHARYNGEAL_CARCINOMA_UP	218	2.22	<0.0001	0.97	1.00
BIDUS_ENDOMETRIAL_METASTASIS_UP	255	2.75	<0.0001	1.32	0.60
ZHAN_MULTIPLE_MYELOMA_PR_UP	31	2.73	<0.0001	0.93	1.00
RODRIGUES_THYROID_CARCINOMA_POORLY_DIFFERENTIATED_UP	490	2.60	<0.0001	-	-
JAEGER_MELANOMA_METASTASIS_DN	223	-1.87	0.012	-1.71	0.10
BIDUS_ENDOMETRIAL_METASTASIS_DN	119	-1.82	0.019	-1.48	0.24
BARRIER_COLO_CANCER_RECURRENCE_DN	18	-1.94	0.007	-1.82	0.06

Figure S5. Genes regulated by enforced MMSET expression significantly overlap with genes associated with different cancers, Related to Figure 5.

(A) Cluster diagram of genes significantly up and down regulated by MMSET overexpression in HME cells. **(B)** The plots represent the observed enrichment score profile of genes induced and repressed by MMSET overexpression in HME cells. Enrichment was determined by Gene Set Enrichment Analysis (GSEA). **(C)** Table showing "Gene sets" enriched among genes upregulated and downregulated in MMSET or MMSETΔSET overexpression in HME cells. n- number of genes in each set, NES-

normalized enrichment score, FDR q value- statistical significance. Note the complete lack of any significant concepts in MMSETΔSET data set.

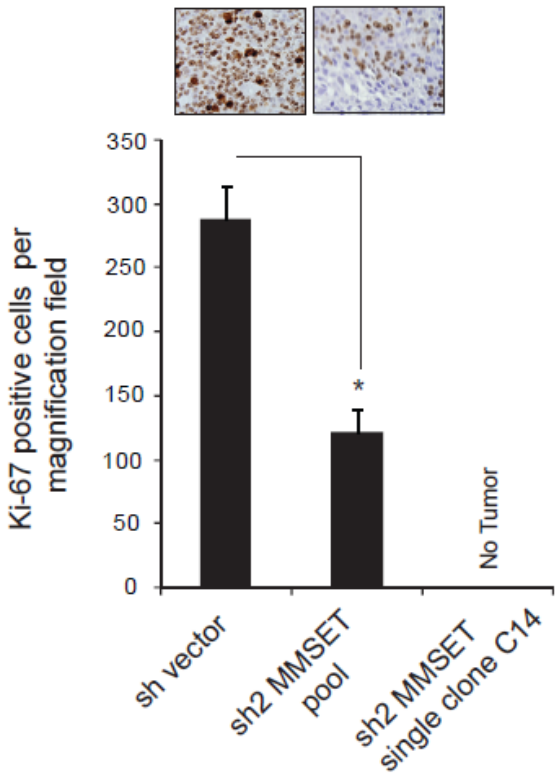


Figure S6. Related to Figure 6.

Reduced tumor proliferation index of MMSET knockdown DU145 xenograft tissue shown by Ki-67 immunohistochemical staining. Inset shows representative images.

Table S1

Gene (HMTase)	Spearman Correlation with EZH2
EZH2	1.00
MMSET	0.79
SUV39H1	0.68
PRMT3	0.68
SUV39H2	0.64
PRMT5	0.62
NSD1	0.59
PRMT1	0.50
CARM1	0.49
PRDM15	0.49
WHSC1L1	0.47
SMYD2	0.47
SMYD5	0.46
DOT1L	0.46
PRDM4	0.44
PRDM10	0.44
SMYD3	0.41
SETD1A	0.41
EHMT1	0.39
MLL	0.39
SETDB1	0.39
PRDM12	0.39
ASH1L	0.35
PRDM13	0.32
PRMT6	0.31
SETD2	0.29
SMYD4	0.28
MLL2	0.26
MLL3	0.25
SETD5	0.25
MLL4	0.19
SETD8	0.18
PRDM9	0.15
PRMT7	0.14
SUV420H1	0.14
SETD3	0.13
SETD7	0.13
PRDM7	0.13
SUV420H2	0.12
PRDM11	0.06
PRDM2	0.01
PRDM14	-0.01
PRMT8	-0.02
MLL5	-0.03
SETDB2	-0.06
SMYD1	-0.08
PRDM5	-0.12
PRDM1	-0.20
SETMAR	-0.21
PRMT2	-0.22
PRDM16	-0.24
EZH1	-0.39

Table S2

Cancer Type	Study	Sample Size	Probe	MMSET	EZH1	PRMT1	SETD5	SETDB1
Adrenal	Giordano	19	MMSET / 38988_at	0.23	-0.14	0.12	0	0.26
Bladder	Sanchez-Carbayo	157	MMSET / 209053_s_at	0.57	-0.02	0.16	0.28	0.35
Bladder	Dyrskjot	60	MMSET / 209053_s_at	0.71	0.28	0.28	0.39	0.45
Brain	Sun	180	MMSET / 209053_s_at	0.75	-0.12	0.46	0.64	0.58
Brain	French	33	MMSET / 209053_s_at	0.75	-0.2	0.45	0.73	0.76
Breast	Richardson	47	MMSET / 209053_s_at	0.71	0.14	0.23	0.3	0.01
Colon	Ki	123	MMSET / AI635529	0.42	0.13	0.71	0.11	-0.02
Colon	Sabates-Bellver	64	MMSET / 209053_s_at	0.73	0.13	0.79	0.63	0.61
Gastric	Derrico	69	MMSET / 209053_s_at	0.83	0.39	0.8	0.43	0.43
Kidney	Beroukhim	70	MMSET / 209053_s_at	0.76	0.13	0.12	-0.12	0.34
Liver	Wurmbach	75	MMSET / 209053_s_at	0.82	0.36	0.55	0.39	0.67
Lung	Bhattacharjee	203	MMSET / 38988_at	0.44	0.12	0.25	0	0.25
Lung	Su	66	MMSET / 209053_s_at	0.71	0.12	0.42	0.14	0.3
Lung	Stearman	39	MMSET / 38988_at	0.49	-0.76	0.3	0	0.32
Lung	Landi	107	MMSET / 209053_s_at	0.71	-0.2	0.41	0.18	0.44
Melanoma	Riker	87	MMSET / 209053_s_at	0.41	-0.01	0.06	0.09	-0.007
Mesothelioma	Gordon	54	MMSET / 209053_s_at	0.51	0.27	0.34	0.37	0.18
Pancreas	Pei	52	MMSET / 209053_s_at	0.71	0.3	0.69	0.37	0.47
Prostate	Varambally	19	MMSET / 209053_s_at	0.9	0.4	0.52	0.86	0.78
Sarcoma	Detwiller	54	MMSET / 209053_s_at	0.56	-0.16	-0.03	0.38	0.17
Sarcoma	Barretina	158	MMSET / 209053_s_at	0.56	0.23	0.12	0.43	0.43
Vulva	Santegoets	19	MMSET / 209053_s_at	0.64	0.28	-0.21	0.2	0.11
		1755	**Samples**					

Table S1 and S2. Correlation between EZH2 and other HMTases expression in RNA-seq and microarray data, Related to Figure 1.

(S1) RNA-seq based correlation values between EZH2 and 51 different histone methyltransferases. Normalized RPKM values from 474 tumor and cell line transcriptomes were used to calculate the correlation between EZH2 and each given histone methyl transferase gene. MMSET, shown in red, was the top ranking candidate with a correlation value of 0.79 ($p<0.0001$). **(S2)** Correlation of EZH2

expression with histone methyltransferase genes MMSET, EZH1, PRMT1, SETD5 and SETDB1 across 22 cancer profiling studies originating from 15 tumor types. Only studies displaying overexpression of EZH2 greater than 2 fold (and p-value <0.0001) in tumor samples relative to normal were included. The probe used for measuring MMSET expression in each study is shown. For studies of remaining genes using platforms with multiple probes, the probe with the maximum correlation score was selected to represent the isoform with the greatest coexpression to EZH2. The gene with the maximum correlation score for each study is shown in red.

Supplemental Experimental Procedure

Expression Analysis

Expression analysis of transcriptome sequencing (RNA-seq) data was conducted by comparing the normalized expression values (reads per kilobase per million (RPKM)) between two independent genes. Expression analysis across cancer profiling studies was carried out using Oncomine (www.oncomine.org) (Rhodes et al., 2004). Co-expression analysis of histone methyltransferase genes was done by exporting raw data from Oncomine for selected studies. To enrich for profiling studies that display activated EZH2 expression in tumor samples relative to normal, we selected only studies where EZH2 was differentially expressed with at least a 2 fold-change and a p-value < 0.0001. Correlation values between two genes were determined by iterating through each probe associated with gene X and comparing it to each probe associated with gene Y. As each probe may represent a different transcript isoform, the maximum correlation score was selected to represent the isoforms having the greatest positive co-expression. All microarray expression studies for MMSET and EZH2 were downloaded from the Oncomine database and normalized and log2 base values were plotted using R. Kaplan-Meier survival analyses for MMSET were carried out using Oncomine power tools beta.

Prostate tissue specimens

Prostate tissues from clinically localized prostate cancer and androgen-independent metastatic prostate cancer patients were obtained from the radical prostatectomy series and Rapid Autopsy Program

respectively through the UM Prostate SPORE Tissue Core. All tissue samples were collected with informed consent under IRB approved protocols at the University of Michigan.

Tissue microarray analysis

The clinically stratified prostate cancer tissue microarrays used in this study have been described (Yu et al., 2007). MMSET and EZH2 protein expression were evaluated across the prostate cancer spectrum. Standard immunohistochemistry was performed with antibodies against MMSET and EZH2 on three high-density tissue microarrays representing benign prostate, clinically localized prostate cancer and metastatic prostate cancer. Product score of staining percentage and intensity was used as an overall measure. Intensity was scored as negative (score = 1), weak (score = 2), moderate (score = 3), or strong (score = 4) which was multiplied by staining percentage to produce the product score for each core. In total, 57 benign, 170 localized prostate cancer, and 38 metastatic prostate cancer tissue cores were evaluable out of 3 tissue microarrays each for MMSET and EZH2 expression.

Chromatin immunoprecipitation (ChIP) and ChIP-seq analysis

The ChIP assay for H3K36me2 from stable EZH2 knockdown and vector control MDA-MB-231 cells was performed using HighCell ChIP kit (Diagenode) according to manufacturer's protocol. Briefly, the cells were cross-linked for 10 minutes with 1% formaldehyde. Cross-linking was terminated by the addition of 1/10 volume 1.25M glycine for 5min at room temperature followed by cell lysis and sonication, resulting in an average chromatin fragment size of 200bp. DNA was isolated from samples by incubation with the antibody at 4°C followed by wash and reversal of cross-linking. The ChIP-seq sample preparation for sequencing was performed according to the manufacturer's instructions (Illumina). ChIP-enriched DNA samples (10-20 ng) were converted to blunt-ended fragments using a combination of T4 DNA polymerase, E.coli DNA polymerase I large fragment (Klenow polymerase) and T4 polynuleotide kinase (New England BioLabs, NEB). A single A-base was added to fragment ends by Klenow fragment (3' to 5' exo minus; NEB) followed by ligation of Illumina adaptors (Quick ligase, NEB). The adaptor-

modified DNA fragments were enriched by PCR using the Illumina PE1.0/PE2.0 primers and Phusion DNA polymerase (NEB). PCR products were size selected using 3% NuSieve agarose gels (Lonza) followed by gel extraction using QIAEX II reagents (QIAGEN). Libraries were quantified using the Bioanalyzer 2100 (Agilent) and each sample was sequenced in a single lane on the Illumina Genome Analyzer II (40 nucleotide read length). HPeak, a Hidden Markov model (HMM)-based peak-calling software designed for the identification of protein-interactive genomic regions, was employed for ChIP-seq data analysis. On the basis of probabilistic models with a novel weighting scheme on read coverage, HPeak features a more rigorous statistical inference and therefore recognizes enriched regions more precisely (Qin et al., 2010). The heatmap of histone H3K36me2 marks was generated using python-based script on raw data and visualized using JavaTreeView. The overlap with gene promoter regions (±5,000 bp from TSS) is examined using 100bp window size. To provide a global landscape of overall enrichment for H3K36me2 between shvector and shEZH2, we divided each *binding* gene body into 20 bins of equal sizes and subsequently calculated and plotted the normalized average read count per bin across all genes. The primer sequences used for ChIP-qPCR validation are provided in **Table S5**.

GSK EZH2 lead inhibitor compounds and DZNep treatment

DU156 cells were seeded in 10cm plate and treated with varying concentrations of GSK (GlaxoSmithKline) compounds or DZNep for four days. Proteasome inhibitor MG132 was added 3 days post-DZNep treatment and processed 4hrs and 48hrs later. Total protein was extracted and was used for immunoblot analysis.

RNA interference and microRNA transfections

Knockdowns of specific genes were accomplished by RNA interference using commercially available siRNA duplexes for EZH2, MMSET and Dicer (Dharmacon, Lafayette, CO). At least 4 independent siRNAs were screened for knockdown efficiency against each target and the most effective siRNA was selected. Precursor microRNAs were purchased from Ambion (Austin, TX). Transfections were

performed with OptiMEM (Invitrogen) and oligofectamine (Invitrogen) as previously described (Varambally et al., 2008).

miR Reporter Luciferase Assays

The MMSET 3.3kb full length 3′ UTR Luciferase reporter pEZX-MT01 plasmid was purchased from GeneCopoeia (Rockville, MD). Mutant full length 3'UTR of MMSET was synthesized (IDT, Iowa) and subcloned into pEZX-MT01 vector. HeLa and HEK293 cells were cultured in 24-well plates and co-transfected with the reporter construct and 50nM precursor miR or negative control miR. The cells were transferred to a 96-well plate 12hr post-transfection and cultured for another 24hr. DU145 and PC3 cells were cotransfected with reporter plasmids and NT siRNA or siEZH2 using Fugene in a 6-well plate and transferred to a 96-well plate 24hr post-transfection and cultured for another 24hr. Luciferase activities were measured using Luc-Pair miR Luciferase Assay kit (GeneCopoeia, Rockville, MD) and normalized by Renilla luciferase activity from matched wells.

Stable overexpression and knockdowns

Stable overexpression of miR-203 in DU145 and PC3 cells was achieved through shRNA lentivirus constructs purchased from OpenBiosystem (Huntsville, AL). Stable knockdown of EZH2 in MDA-MB-231 cells was achieved through shRNA lentivirus constructs purchased from System Biosciences (Mountain View, CA). Generation of DU145 EZH2 knockdown single clones has been described earlier (Yu et al., 2007b). Stable knockdown of MMSET in DU145 and PC3 cells was achieved through two independent custom made shRNA lentivirus constructs (System Biosciences). Overexpression construct of full length MMSET or its SET domain mutant was achieved by PCR amplification of respective cDNA fragments and cloning into pLenti4V5 (Invitrogen) plasmid. Lentivirus transduced cells were grown in the presence of selection antibiotics prior to use in experiments. Lentiviruses and adenoviruses were generated by the University of Michigan Vector Core. The two shRNA target sequence for MMSET are given below:

sh1MMSET – GGUCCAAAGUCUCGGGUUA; sh2MMSET – UGUCAGUGGAGGAGCGGAA

RNA isolation and Quantitative Real-Time PCR Assays

Total RNA was isolated from tissues using Trizol and from cells using RNeasy Mini Kit (Qiagen). Quantitative PCR (QPCR) and miRNA Taqman qPCR (Applied Biosystems,USA) were performed as described (Asangani et al., 2008). All primers were designed using Primer 3 (http://frodo.wi.mit.edu/primer3/) and synthesized by Integrated DNA Technologies (Coralville, IA). QPCRs were performed in duplicate or triplicate. The primer sequences for the transcript analyzed by SYBR green qPCR are provided in the **Table S5**.

Cell Proliferation Assay

For proliferation assays, 20,000 cells/well were seeded in 24-well plates (n=3), and cells were harvested and counted at the indicated time points by Coulter counter (Beckman Coulter, Fullerton, CA).

Wound healing Assay

Cells were grown to confluency in 6-well plates. A scratch was made through the cell monolayer using a pipette tip; cells were then washed and fresh culture medium was added. Photographs of the wounded area were taken immediately after making the scratch (0hr time point) and after 24hr to monitor the migration of cells into the wounded area.

Matrigel Invasion Assays

Normal prostate epithelial cells (PrEC) and normal breast epithelial cells (H16N2), and DU145 prostate cancer cells were infected with adenovirus along with siRNA transfection. Forty-eight hours post-infection/transfection, cells were seeded in a transwell chamber pre-coated with Matrigel (BD Biosciences). Medium containing 10% FBS in the lower chamber served as chemoattractant. After 48hr, the non-invading cells and EC matrix were gently removed with a cotton swab and invasive cells located on the lower side of

the chamber were stained with crystal violet, air dried and photographed. For colorimetric assays, the inserts were treated with 150µl of 10% acetic acid and the absorbance measured at 560nm using a spectrophotometer (GE Healthcare). A similar protocol was followed for invasion assay of DU145 and PC3 cells with stable MMSET knockdown and HME and H16N2 cells overexpressing MMSET.

Prostatosphere Culture

Spheres culture was performed as described (Dontu et al., 2003; Yu et al., 2007a). Briefly, cells (1000 cells/mL) were cultured in suspension in serum-free DMEM-F12 (BioWhittaker) supplemented with B27 (1:50, Invitrogen), 20 ng/mL EGF (BD Biosciences), 0.4% bovine serum albumin (Sigma), and 4 µg/mL insulin (Sigma). To propagate spheres, spheres were collected by gentle centrifugation, dissociated to single cells and then cultured to generate prostatospheres of the next generation. Spheres larger than 100 µm were counted.

Gene expression profiling

Expression profiling was performed using the Agilent Whole Human Genome Oligo Microarray (Santa Clara, CA) according to the manufacturer's protocol. All samples were run in technical duplicates against control. Over- and under-expressed gene sets were generated by filtering to include only features with 2-fold average over- or underexpression (Log ratio with $p<0.01$) in all hybridization and were analyzed for enrichment of biological themes using DAVID bioinformatics platform (Dennis et al., 2003).

Chicken Chorio-Allantoic Membrane (CAM) assays

The CAM assay for local cell invasion, intravasation, metastasis and tumor (or xenograft) formation was performed as previously described (Brenner et al., 2011). Briefly, fertilized eggs were incubated in a rotary humidified incubator at 38°C for 10 days. CAM was released by applying mild amount of low pressure to the hole over the air sac and cutting a 1 cm^2 window encompassing a second hole near the

allantoic vein. Approximately 2 million cells in 50µl of media were implanted in each egg, windows were sealed and the eggs were returned to a stationary incubator.

For local invasion and intravasation experiments, the upper and lower CAM were isolated after 72hr. The upper CAM were processed and stained for chicken collagen IV (immunofluorescence) or human cytokeratin (immunohistochemistry) as previously described (Brenner et al., 2011).

For metastasis assay, the embryonic livers were harvested on day 18 of embryonic growth and analyzed for the presence of tumor cells by quantitative human Alu-specific PCR. Genomic DNA from lower CAM and livers were prepared using Puregene DNA purification system (Qiagen) and quantification of human-Alu was performed as described (van der Horst et al., 2004). Fluorogenic TaqMan qPCR probes were generated as described above and used to determine DNA copy number.

For xenograft growth assay with H16N2 and HME cells, the embryos were sacrificed on day 18 and the extra-embryonic xenograft were excised and weighed.

Murine Prostate Tumor Xenograft Model

Five-week-old male nude athymic BALB/c nu/nu mice (Charles River Laboratory, Wilmington, MA) were anesthetized using a cocktail of xylazine (80 mg/kg IP) and ketamine (10 mg/kg IP) and $5x10^6$ of MMSET knockdown prostate cancer cells suspended in 100µl of saline with 20% Matrigel (BD Biosciences) were implanted subcutaneously into the dorsal flank on both sides of the mice (shvector n=7, shMMSET pool n=7 and shMMSET single clone C14 n=5). Tumor growth was recorded weekly using digital calipers and tumor volumes were estimated using the formula $(\pi/6)$ $(L \times W^2)$, where L = length of tumor and W = width. At the end of the studies, mice were euthanized and various organs and bone marrow cells were isolated to examine spontaneous metastasis by human-Alu PCR quantification. All procedures involving mice were approved by the University Committee on Use and Care of Animals (UCUCA) at the University of Michigan and conform to all regulatory standards.

Supplemental References

Asangani, I.A., Rasheed, S.A., Nikolova, D.A., Leupold, J.H., Colburn, N.H., Post, S., and Allgayer, H. (2008). MicroRNA-21 (miR-21) post-transcriptionally downregulates tumor suppressor Pdcd4 and stimulates invasion, intravasation and metastasis in colorectal cancer. Oncogene *27*, 2128-2136.

Brenner, J.C., Ateeq, B., Li, Y., Yocum, A.K., Cao, Q., Asangani, I.A., Patel, S., Wang, X., Liang, H., Yu, J., *et al.* (2011). Mechanistic rationale for inhibition of poly(ADP-ribose) polymerase in ETS gene fusion-positive prostate cancer. Cancer Cell *19*, 664-678.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev *17*, 1253-1270.

Qin, Z.S., Yu, J., Shen, J., Maher, C.A., Hu, M., Kalyana-Sundaram, S., and Chinnaiyan, A.M. (2010). HPeak: an HMM-based algorithm for defining read-enriched regions in ChIP-Seq data. BMC Bioinformatics *11*, 369.

Rhodes, D.R., Yu, J., Shanker, K., Deshpande, N., Varambally, R., Ghosh, D., Barrette, T., Pandey, A., and Chinnaiyan, A.M. (2004). ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia *6*, 1-6.

van der Horst, E.H., Leupold, J.H., Schubbert, R., Ullrich, A., and Allgayer, H. (2004). TaqMan-based quantification of invasive cells in the chick embryo metastasis assay. Biotechniques *37*, 940-942, 944, 946.

Varambally, S., Cao, Q., Mani, R.S., Shankar, S., Wang, X., Ateeq, B., Laxman, B., Cao, X., Jing, X., Ramnarayanan, K., *et al.* (2008). Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer. Science *322*, 1695-1699.

Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J., *et al.* (2007a). let-7 regulates self renewal and tumorigenicity of breast cancer cells. Cell *131*, 1109-1123.

Yu, J., Cao, Q., Mehra, R., Laxman, B., Tomlins, S.A., Creighton, C.J., Dhanasekaran, S.M., Shen, R., Chen, G., Morris, D.S., *et al.* (2007b). Integrative genomics analysis reveals silencing of beta-adrenergic signaling by polycomb in prostate cancer. Cancer Cell *12*, 419-431.