SUPPLEMENTARY INFORMATION

Analysis of sheared genomic DNA

Three hundred ng of sheared DNA was electrophoresed in 2% agarose gel, and stained with ethidium bromide. The gel image was analyzed using ImageJ software (National Institutes of Health) to measure the relative density of DNA. Relative copy numbers of DNA molecules were calculated using a formula (relative density/molecular weight). 100 bp DNA ladder (TOYOBO, Osaka, Japan) was used for determination of DNA length.

Methylation-specific PCR (MSP)

Genomic DNA of each cell line was digested with *Bam*HI, and 1 μg of digested DNA was used for bisulfite modification, which converts unmethylated cytosine to uracil, as described previously (Yamashita et al. 2006). Modified DNA was re-suspended in 40 μl of 1 x TE, and 1 μl was used for PCR using primers specific to methylated or unmethylated DNA. The MSP primers used in this study (Table S6) were designed to detect DNA methylation statuses of 4-8 CpG sites within the primer sequences. Since DNA methylation statuses of CGIs, especially NFRs, are regulated as a region (the majority of CpG sites on a single DNA molecule are methylated or unmethylated) (Kaneda et al. 2002; Kaneda et al. 2004), MSP can detect the methylation status of a region by interrogating 4-8 CpG sites in the region.

Validation of the Me value by bisulfite sequencing

The accuracy of the Me value was validated by bisulfite sequencing of 40 samples (10 CGIs in four cancer cell lines with variable methylation levels) (Yamashita et al. in

press). Each CGI had 7 to 21 CpG sites within 200 bp of microarray probes, and 10 clones or more were analyzed for each sample. A methylation level of a sample was calculated as a fraction of methylated CpG sites among the total CpG sites. A correlation coefficient between the Me value and the methylation level obtained by bisulfite sequencing was 0.86.

Quantitative reverse transcription PCR (RT-PCR)

Total RNA of RWPE1 and HMEC was isolated using ISOGEN (Nippon Gene, Tokyo, Japan). cDNA was synthesized from 3 μg of total RNA using SuperScriptIII reverse transcriptase and an oligo (dT) $_{12\text{-}18}$ primer (Invitrogen), and 1/120 of the total cDNA was used for quantitative RT-PCR. Quantitative RT-PCR was performed as described previously (Nakajima et al. 2009), using standard DNA after purification and quantification using QIAxcel (Qiagen, Valencia, CA). Primers used in quantitative RT-PCR are listed in Table S7.

Supplementary Figure legends

Fig. S1

Determination of an appropriate cutoff Me value for methylated (unmethylated) CGIs. Methylation statuses of 29 CGIs (listed in Table S1) were analyzed by qualitative MSP in 5 cell lines (total 145 CGIs). We were able to amplify 143 CGIs out of 145 CGIs by MSP, and these were used as a reference (41 methylated, 14 partially methylated, and 88 unmethylated CGIs). Sensitivity and specificity for methylated (unmethylated) CGIs were calculated as [# of CGIs assessed as HM (UM) by MeDIP-CGI microarray among CGIs determined as methylated (unmethylated) by MSP] and [# of CGIs assessed as MM and UM (HM and MM) by MeDIP-CGI microarray among CGIs determined as partially methylated and unmethylated (partially methylated and methylated) by MSP], respectively. (A) Determination of cutoff Me value for methylated CGIs. Cutoff values between 0.3 and 0.8 were tested, and a value of 0.6 was selected as a value of reasonable sensitivity and specificity. (B) Determination of cutoff Me value for unmethylated CGIs. Cutoff values between 0.2 and 0.7 were tested, and a value of 0.4 was selected.

Fig. S2

DNA methylation levels (the average Me values) at various positions against the TSSs in the normal human mammary epithelial cells and three breast cancer cell lines. Average Me value of CGIs continuous from their NFRs are shown. The blue dotted rectangle indicates the NFRs. Methylation levels of the NFRs were also similar to those of upstream regions up to -800 bp and downstream regions up to +800 bp.

Fig. S3

Validation of GeneChip oligonucleotide microarray results by quantitative RT-PCR of 30 genes. Average signal intensity measured by GeneChip oligonucleotide microarray (upper panel) and mRNA copy number normalized to *GAPDH* measured by quantitative RT-PCR (lower panel) are shown for RWPE1 (A) and HMEC (C). The means ± SD values ($n = 4$) are shown. Scatter plots show correlation between the signal intensity and mRNA copy number in RWPE1 (B) and HMEC (D). Strong correlation was observed in both cell lines (correlation coefficient = 0.95 and 0.97).

Fig. S4

The influence of DNA methylation on transcription levels within the same cells (RWPE1 and HMEC). Transcription levels of highly methylated (HM, black), moderately methylated (MM, gray), and unmethylated (UM, white) genes are shown, respectively. For the box plot, refer to the legend to Fig. 2B. Transcription levels of HM, and MM genes were compared to those of UM genes as described in Fig. 2B. A clear silencing effect of DNA methylation in the NFRs was observed.

Fig. S5

Correlation between DNA methylation frequency in cancer cell lines and high fraction of genes with low transcription in normal cell line (cells). Susceptible genes were divided into subclasses according to DNA methylation frequency in cancer cell lines (S1-S4 for the prostates; and S1-S3 for the mammary glands). The fractions of genes with high (blue), moderate (pink), and low (yellow) transcription are shown.

Fig. S6

Low transcription levels of DNA methylation-susceptible genes in the normal human mammary epithelial cells (HMEC). (A) The association between DNA methylation levels (Me value of the NFRs) in each of the three breast cancer cell lines (NCF7, ZR-75-1, and MDA-MB-468) and transcription levels in HMEC. Green dots represent genes highly methylated in a cancer cell line. Genes highly methylated in a cancer cell line had low transcription levels in normal cells. (B) Transcription levels of resistant (R), intermediate (Int), and susceptible (S1-S3) genes in HMEC. For the box plot and statistical methods, refer to the legend to Fig. 2B. Susceptible genes had significantly lower expression levels than resistant genes. (C) The fraction of genes with high (blue; signal intensity >1,000), moderate (pink; 250-1,000), and low (yellow; <250) transcription. Susceptible genes had a significantly larger fraction of genes with low transcription than the total genes.

Fig. S7

Validation of ChIP-on-chip results by quantitative ChIP-PCR. Twenty-three genes were used for validation. Each panel shows correlation between the data obtained by ChIP-on-chip and those obtained by quantitative ChIP-PCR in RWPE1. H3Ac, H3K4me3, H3K9me3, H3K27me3, and pol II binding levels had correlation coefficients of 0.80, 0.65, 0.80, 0.88, and 0.86, respectively.

Fig. S8

The association between the levels of candidate instructive factors in HMEC and DNA methylation susceptibility, among genes with low transcription in HMEC. (A) Histone modification levels of genes with different susceptibilities to DNA methylation.

For the box plot and statistical methods, refer to the legend to Fig. 2B. Active histone modifications were elevated in resistant genes, and H3K27me3 was elevated in susceptible genes. (B) The association between pol II binding and DNA methylation susceptibility. Pol II binding was associated with resistance even among genes with low transcription. (C) Levels of histone modifications and pol II binding at various positions against the TSSs in HMEC. Average levels of histone modifications and pol II binding of CGIs continuous from their NFRs are shown. The blue dotted rectangle indicates the NFRs. (D) The combination effect of one of the three active factors (the y-axis) and H3K27me3 (the x-axis) on resistance and susceptibility of genes with low transcription. Red and green dots represent DNA methylation-resistant and -susceptible genes, respectively, and were separated by any of the three combinations.

Fig. S9

Levels of histone modifications of genes with high, moderate, and low transcription at various positions against the TSSs in RWPE1 and HMEC. Average levels of histone modifications of CGIs continuous from their NFRs are shown. Genes with low transcription were further divided by DNA methylation status of the NFRs. Low (HM) and Low (UM) represent genes highly methylated and unmethylated, respectively. The blue dotted rectangle indicates the NFRs. Genes with high and low transcription had elevated active and inactive histone modifications, respectively.

Fig. S10

The association between pol II binding and DNA methylation resistance in the total 6,857 genes in the mammary gland, regardless of transcription levels. (A) Classification of genes by pol II status and H3K27me3 in the human mammary epithelial cells. We

were able to analyze transcription levels for 4,506 of 5,430 resistant, 1,690 of 1,913 intermediate, and 661 of 733 susceptible genes (total 6,857 of 8,076 genes) due to a difference in microarray platforms. Genes with active, stalled, and low pol II were classified as described in Fig. 4A, and their numbers are shown. (B) The fractions of resistant (red), intermediate (light green), and susceptible (green) genes according to the pol II and H3K27me3 statuses. Genes with either active or stalled pol II had a larger fraction of resistant genes, and genes with low pol II had a larger fraction of susceptible and intermediate genes.

Fig. S11

Specificity of antibodies used in ChIP-on-chip. Quantitative ChIP-PCR was performed for an actively transcribed gene (*GAPDH*), a target of H3K9me3 (satellite II), and two targets of H3K27me3 (*KCNA1*, and *CNR1*). Thirty μg of sheared chromatin extracted from RWPE1 was immunoprecipitated, and quantitative PCR was performed using primer sets shown in Table S5 to obtain the IP /WCE (%). It was confirmed that each antibody enriched its target sequences.

Fig. S12

Analysis of the size of DNA sheared for MeDIP. (A) Gel image of sheared DNA. Three hundred ng of sheared genomic DNA was electrophoresed in 2 % agarose gel and DNA was visualized by ethidium bromide. The length of the size marker is shown at the left of the panel. (B) Analysis of DNA density. Relative DNA density was measured using ImageJ software, and plotted. (C) Relative copy number of sheared DNA. DNA density was normalized by molecular weight of DNA, and plotted. Most DNA was sheared to a length around 300 bp.

Supplementary Reference

Kaneda, A., M. Kaminishi, K. Yanagihara, T. Sugimura, and T. Ushijima. 2002. Identification of silencing of nine genes in human gastric cancers. *Cancer Res* **62:** 6645-6650.

Kaneda, A., T. Tsukamoto, T. Takamura-Enya, N. Watanabe, M. Kaminishi, T. Sugimura, M. Tatematsu, and T. Ushijima. 2004. Frequent hypomethylation in multiple promoter CpG islands is associated with global hypomethylation, but not with frequent promoter hypermethylation. *Cancer Sci* **95:** 58-64.

Nakajima, T., S. Yamashita, T. Maekita, T. Niwa, K. Nakazawa, and T. Ushijima. 2009. The presence of a methylation fingerprint of *Helicobacter pylori* infection in human gastric mucosae. *Int J Cancer* **124:** 905-910.

Yamashita, S., K. Hosoya, K. Gyobu, H. Takeshima, and T. Ushijima. in press. Development of a novel output value for quantitative assessment in methylated DNA immunoprecipitation-CpG island microarray analysis. *DNA Res*.

Yamashita, S., Y. Tsujino, K. Moriguchi, M. Tatematsu, and T. Ushijima. 2006. Chemical genomic screening for methylation-silenced genes in gastric cancer cell lines using 5-aza-2'-deoxycytidine treatment and oligonucleotide microarray. *Cancer Sci* **97:** 64-71.

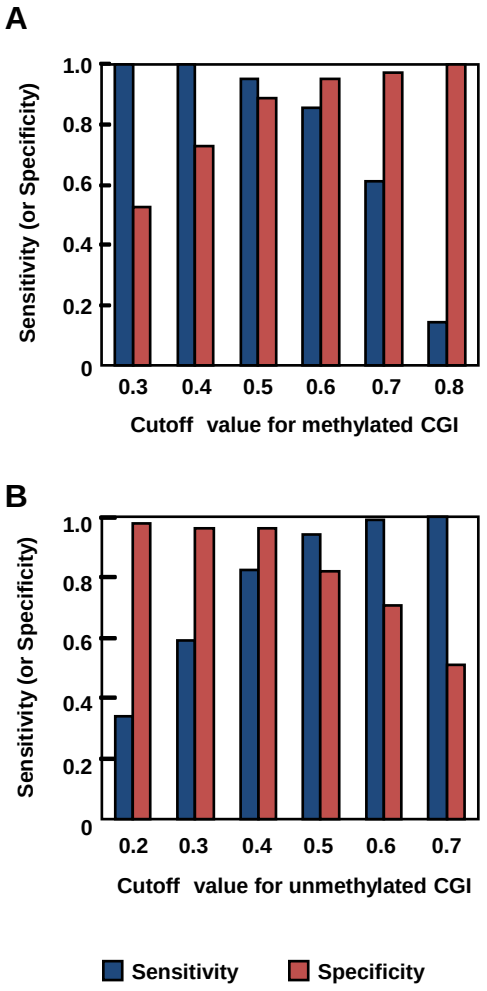
A
1.0
0.8
0.6
0.4
0.2
0
Sensitivity (or Specificity)
0.3
0.4
0.5
0.6
0.7
0.8
Cutoff value for methylated CGI
B
1.0
0.8
0.6
0.4
0.2
0
Sensitivity (or Specificity)
0.2
0.3
0.4
0.5
0.6
0.7
Cutoff value for unmethylated CGI
Sensitivity
Specificity

Fig. S1

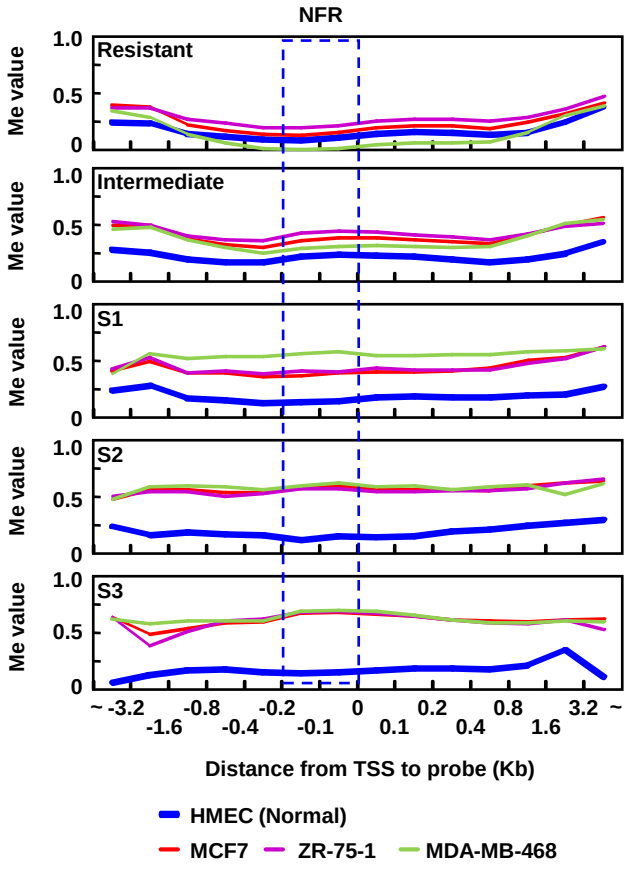
NFR
Resistant
Intermediate
S1
S2
S3
Me value
1.0
0.5
0
~ -3.2
-1.6
-0.8
-0.4
-0.2
-0.1
0
0.1
0.2
0.4
0.8
1.6
3.2 ~
Distance from TSS to probe (Kb)
HMEC (Normal)
MCF7
ZR-75-1
MDA-MB-468

Fig. S2

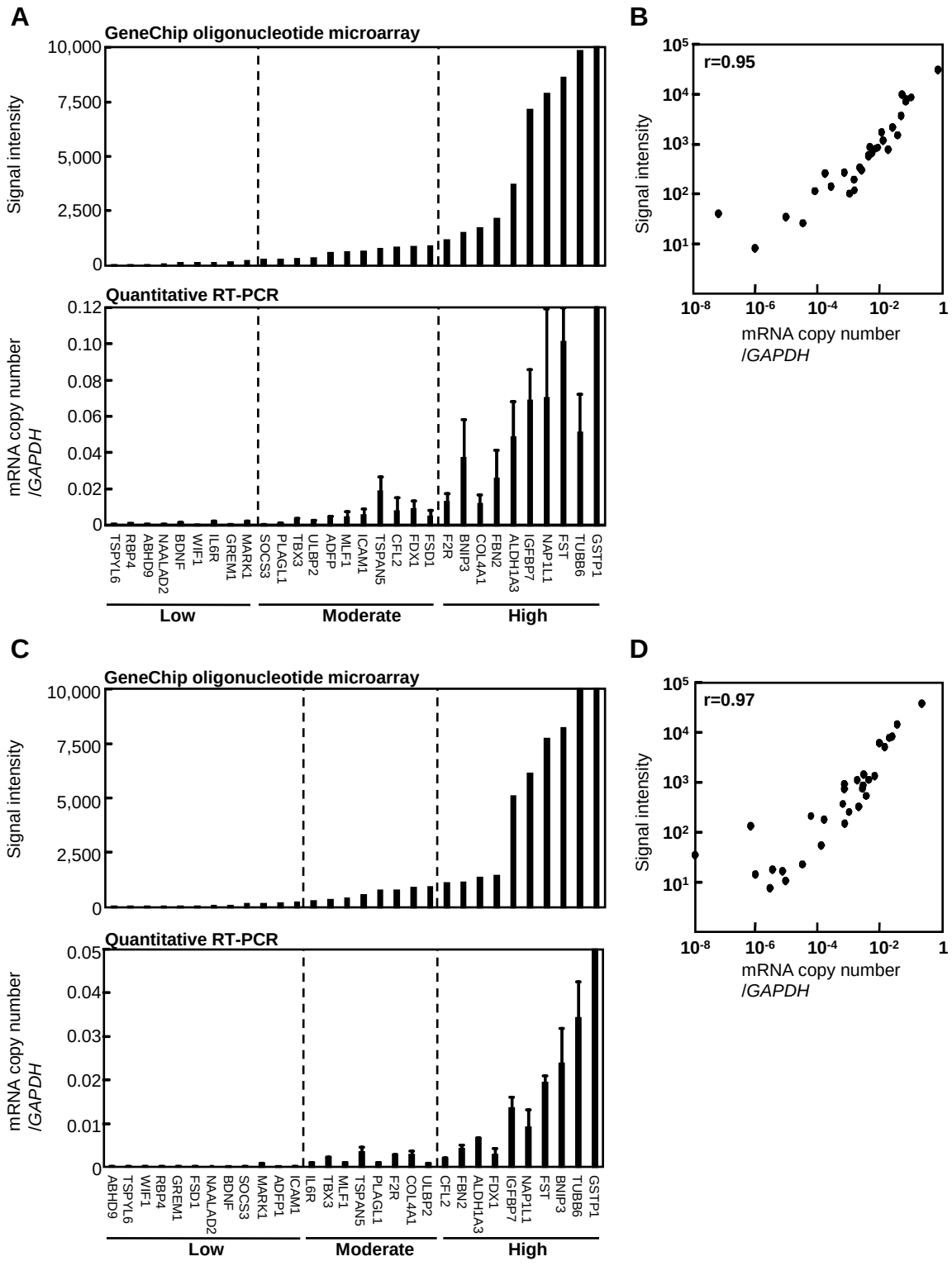
A
GeneChip oligonucleotide microarray
Signal intensity
10,000
7,500
5,000
2,500
0
Quantitative RT-PCR
mRNA copy number /GAPDH
0.12
0.10
0.08
0.06
0.04
0.02
0
TSPYL6 RBP4 ABHD9 NAALAD2 BDNF WIF1 IL6R GREM1 MARK1 SOCS3 PLAGL1 TBX3 ULBP2 ADFP MLF1 ICAM1 TSPAN5 CFL2 FDX1 FSD1 F2R BNIP3 COL4A1 FBN2 ALDH1A3 IGFBP7 NAP1L1 FST TUBB6 GSTP1
Low
Moderate
High
B
r=0.95
Signal intensity
10^5
10^4
10^3
10^2
10^1
10^-8 10^-6 10^-4 10^-2 1
mRNA copy number /GAPDH
C
GeneChip oligonucleotide microarray
Signal intensity
10,000
7,500
5,000
2,500
0
Quantitative RT-PCR
mRNA copy number /GAPDH
0.05
0.04
0.03
0.02
0.01
0
ABHD9 TSPYL6 WIF1 RBP4 GREM1 FSD1 NAALAD2 BDNF SOCS3 MARK1 ADFP1 ICAM1 IL6R TBX3 MLF1 TSPAN5 PLAGL1 F2R COL4A1 ULBP2 CFL2 FBN2 ALDH1A3 FDX1 IGFBP7 NAP1L1 FST BNIP3 TUBB6 GSTP1
Low
Moderate
High
D
r=0.97
Signal intensity
10^5
10^4
10^3
10^2
10^1
10^-8 10^-6 10^-4 10^-2 1
mRNA copy number /GAPDH

Fig. S3

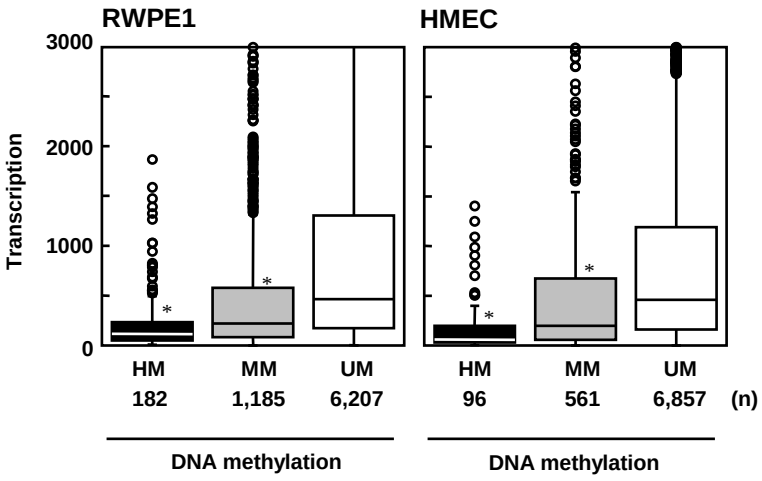
RWPE1
HMEC
3000
2000
1000
0
Transcription
HM
MM
UM
182
1,185
6,207
96
561
6,857
(n)
DNA methylation
DNA methylation

Fig. S4

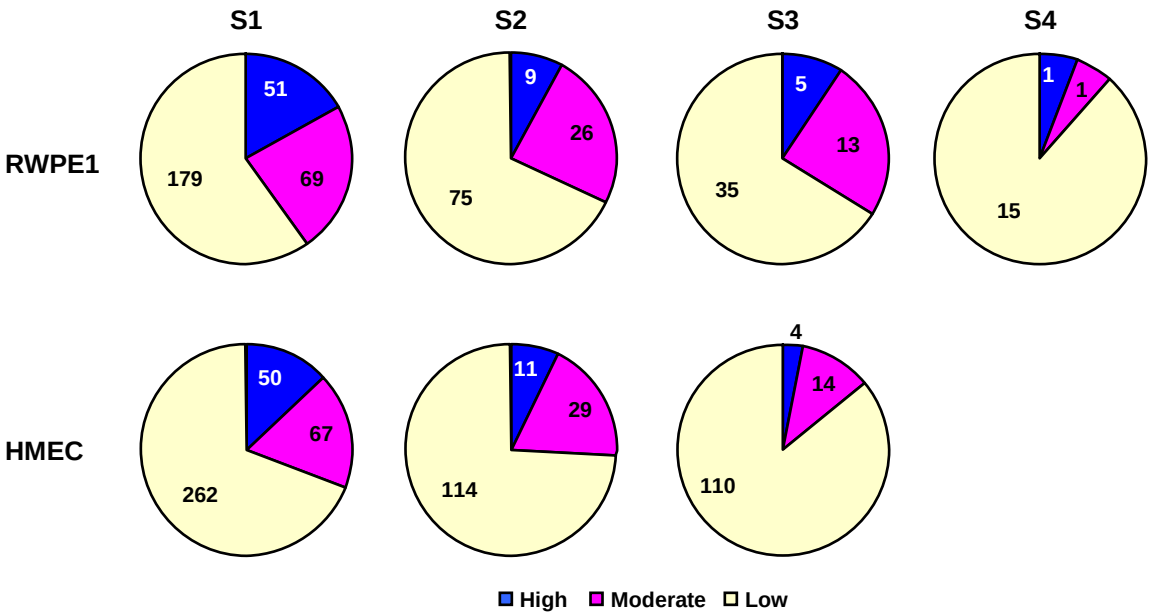
S1
S2
S3
S4
RWPE1
HMEC
51
69
179
9
26
75
5
13
35
1
1
15
50
67
262
11
29
114
4
14
110
High
Moderate
Low

Fig. S5

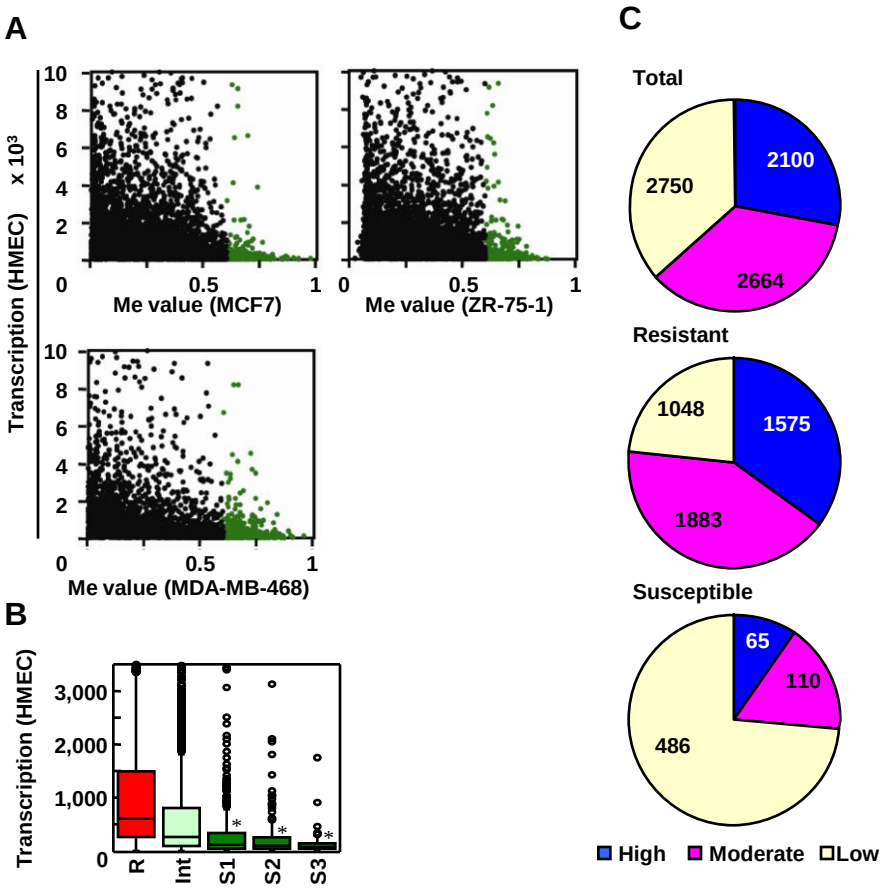
A
Transcription (HMEC) x 10³
Me value (MCF7)
Me value (ZR-75-1)
Me value (MDA-MB-468)
B
Transcription (HMEC)
R
Int
S1
S2
S3
C
Total
2100
2750
2664
Resistant
1048
1575
1883
Susceptible
65
110
486
High
Moderate
Low

Fig. S6

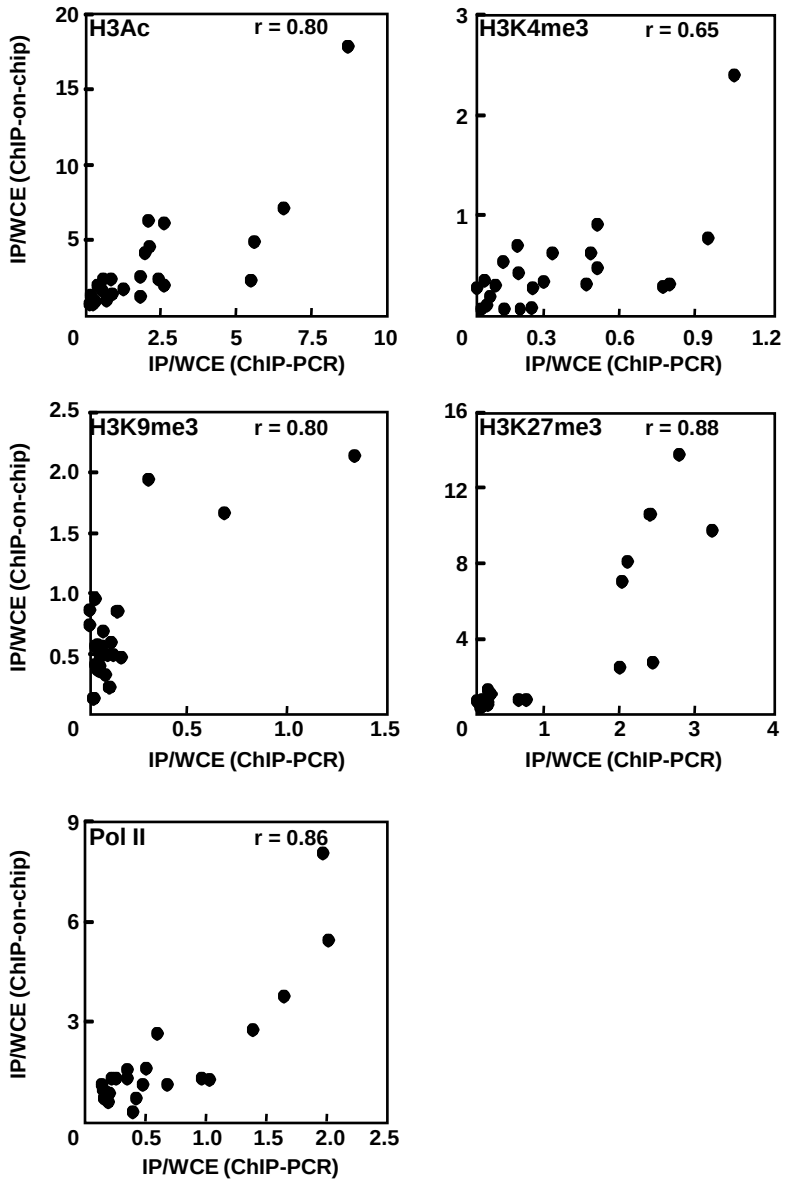
H3Ac
r = 0.80
IP/WCE (ChIP-on-chip)
IP/WCE (ChIP-PCR)
H3K4me3
r = 0.65
H3K9me3
r = 0.80
H3K27me3
r = 0.88
Pol II
r = 0.86

Fig. S7

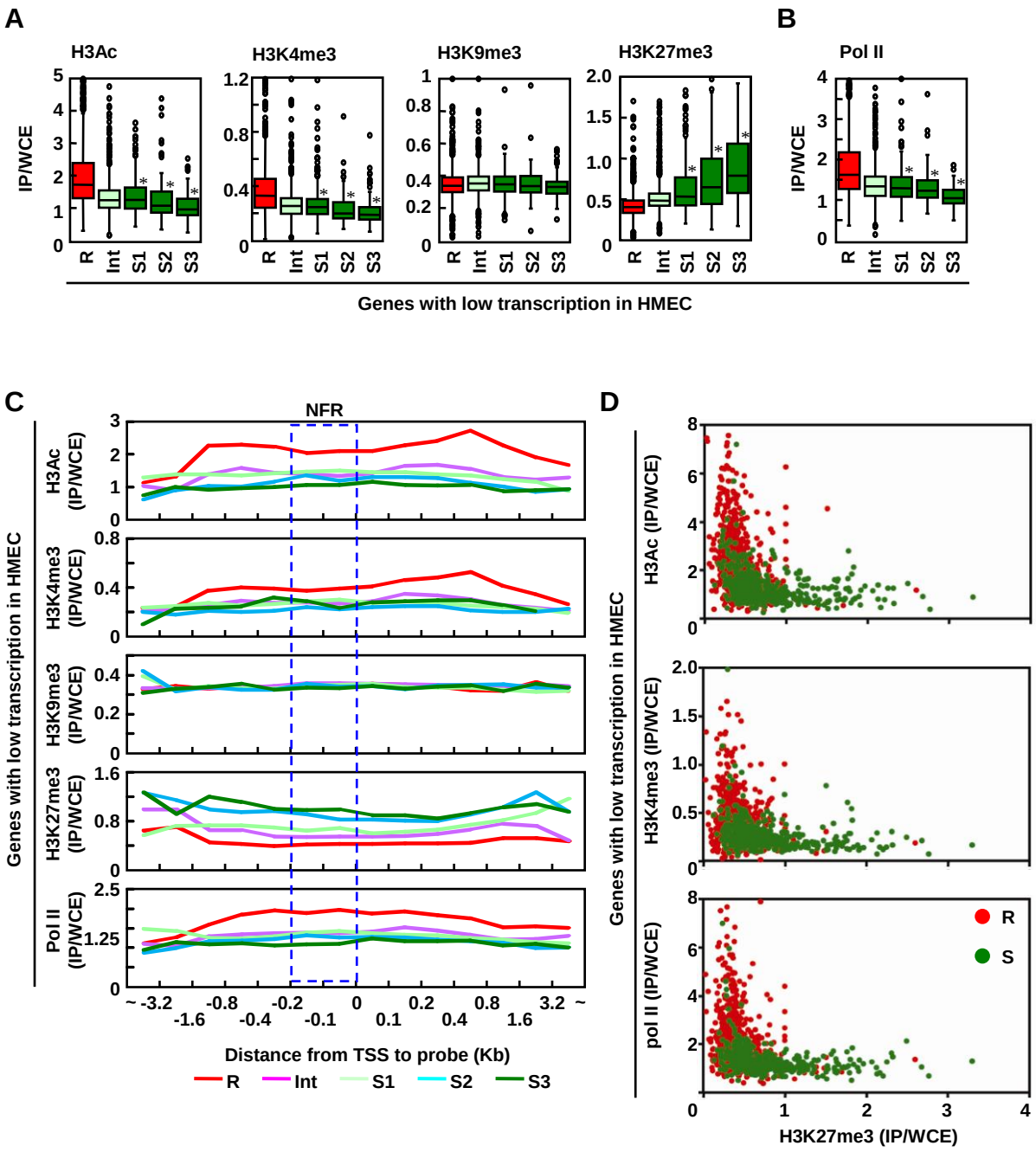
A
H3Ac
H3K4me3
H3K9me3
H3K27me3
IP/WCE
R
Int
S1
S2
S3
Genes with low transcription in HMEC
B
Pol II
C
NFR
H3Ac (IP/WCE)
H3K4me3 (IP/WCE)
H3K9me3 (IP/WCE)
H3K27me3 (IP/WCE)
Pol II (IP/WCE)
Genes with low transcription in HMEC
Distance from TSS to probe (Kb)
D
pol II (IP/WCE)
H3K27me3 (IP/WCE)
R
S

Fig. S8

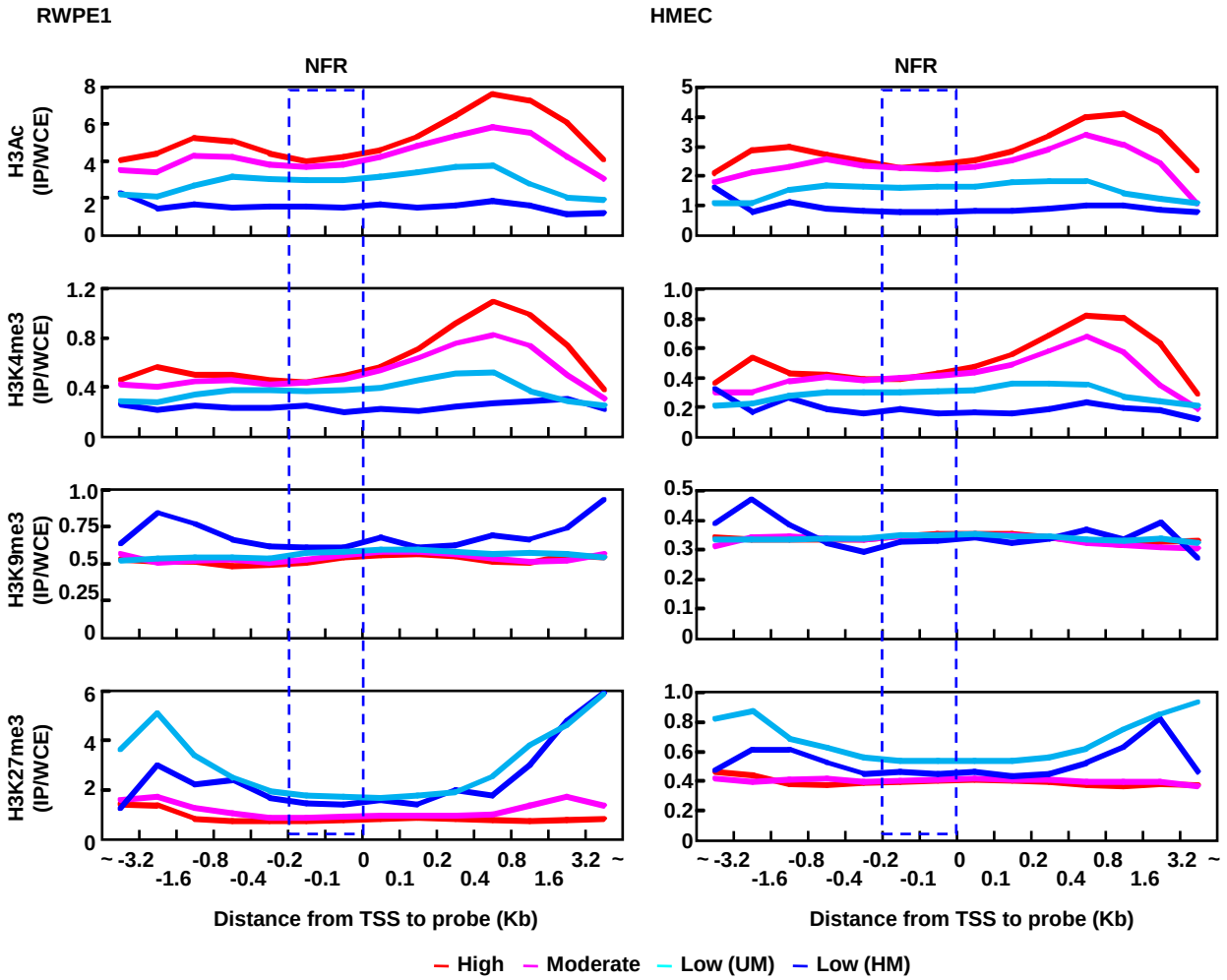
RWPE1
HMEC
NFR
NFR
H3Ac (IP/WCE)
H3K4me3 (IP/WCE)
H3K9me3 (IP/WCE)
H3K27me3 (IP/WCE)
~ -3.2 -1.6 -0.8 -0.4 -0.2 -0.1 0 0.1 0.2 0.4 0.8 1.6 3.2 ~
Distance from TSS to probe (Kb)
High
Moderate
Low (UM)
Low (HM)

Fig. S9

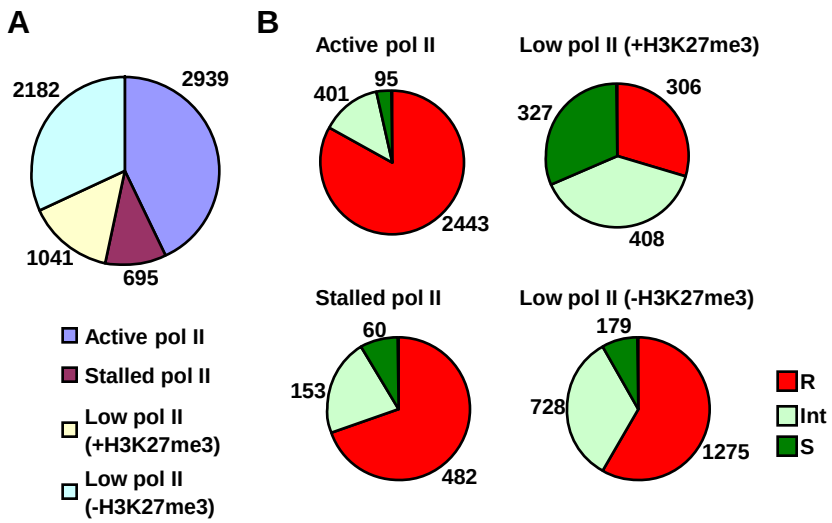
A
2939
2182
1041
695
Active pol II
Stalled pol II
Low pol II (+H3K27me3)
Low pol II (-H3K27me3)
B
Active pol II
95
401
2443
Low pol II (+H3K27me3)
306
327
408
Stalled pol II
60
153
482
Low pol II (-H3K27me3)
179
728
1275
R
Int
S

Fig. S10

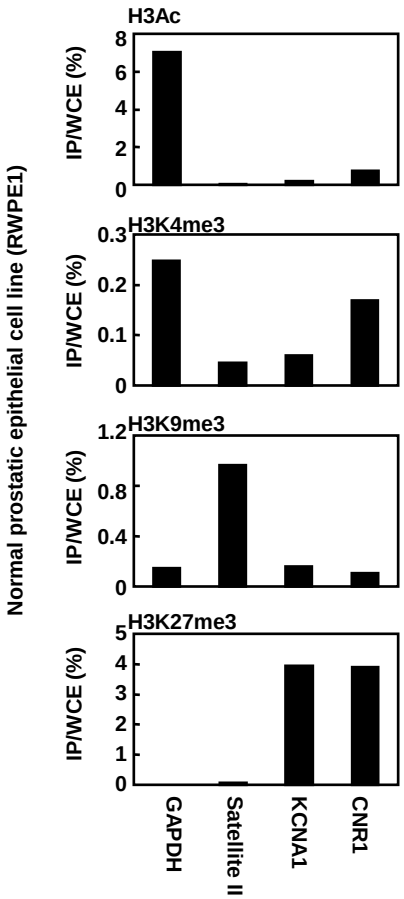
Normal prostatic epithelial cell line (RWPE1)
H3Ac
IP/WCE (%)
8
6
4
2
0
H3K4me3
IP/WCE (%)
0.3
0.2
0.1
0
H3K9me3
IP/WCE (%)
1.2
0.8
0.4
0
H3K27me3
IP/WCE (%)
5
4
3
2
1
0
GAPDH
Satellite II
KCNA1
CNR1

Fig. S11

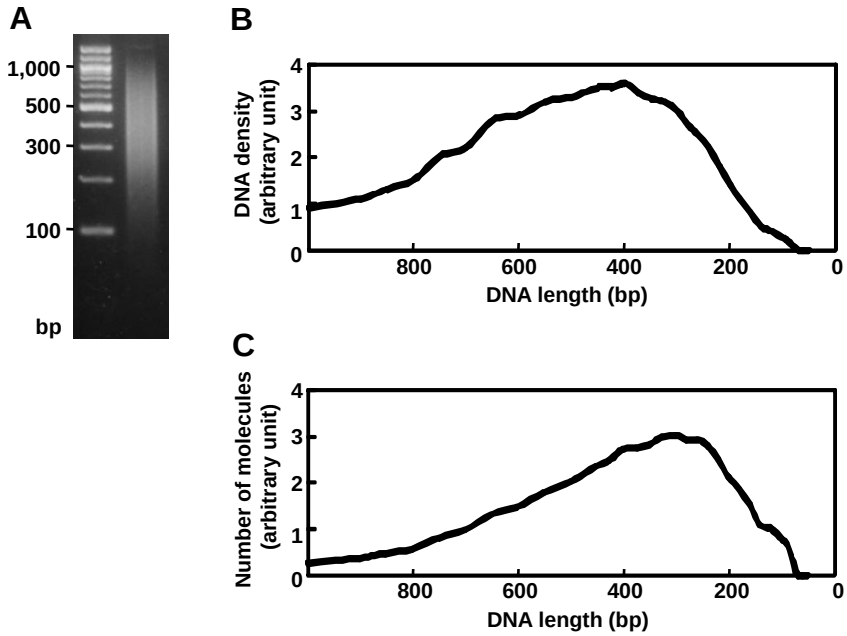
A
1,000
500
300
100
bp
B
DNA density
(arbitrary unit)
4
3
2
1
0
800
600
400
200
0
DNA length (bp)
C
Number of molecules
(arbitrary unit)
4
3
2
1
0
800
600
400
200
0
DNA length (bp)

Fig. S12

Table S1. DNA methylation status of 29 CGIs determined by MeDIP-CGI microarray and those by MSP

		MeDIP-CGI microarray							MSP				
Gene symbol	Region	Number of probes	Position of probes	RWPE1	PC3	LNCaP	22Rv1	Du145	RWPE1	PC3	LNCaP	22Rv1	Du145
ADFP	Promoter	3	From -301 to -86	UM	UM	UM	UM	UM	U	U	U	U	U
C6orf148	NFR	2	From -142 to -32	UM	UM	UM	UM	UM	U	U	U	U	U
CTSL1	NFR	2	From -155 to -77	UM	UM	UM	UM	UM	U	U	U	U	U
CYP26C1	Promoter	13	From -1,060 to -92	HM	HM	HM	HM	HM	M	M	M	M	M
DAZL	NFR	2	From -168 to -90	HM	HM	HM	HM	HM	M	M	x	M	M
DPEP3	NFR	3	From -161 to -9	HM	HM	HM	HM	HM	M	M	M	M	M
F2R	NFR	2	From -195 to -78	UM	UM	UM	UM	UM	U	U	U	U	U
FADS1	NFR	2	From -136 to -64	UM	UM	UM	UM	UM	U	U	U	U	U
FLJ21963	NFR	2	From -162 to -47	UM	HM	HM	HM	HM	U	M/U	M	M	M
FOXJ1	NFR	2	From -189 to -86	MM	MM	MM	UM	UM	M/U	M/U	M/U	U	U
FOXO1	Promoter	10	From -1,062 to -156	UM	UM	UM	UM	UM	U	U	U	U	U
GREM1	Promoter	7	From -750 to -270	UM	HM	HM	HM	HM	U	M/U	M/U	M	M
GSTP1	NFR	2	From -194 to -78	UM	MM	HM	HM	MM	U	M/U	M	M	U
IGFBP7	NFR	1	-21	UM	MM	MM	UM	UM	U	M	M	U	U
IL6R	NFR	1	-31	UM	UM	MM	UM	UM	U	U	U	U	U
LAMA1	NFR	2	From -148 to -27	MM	UM	MM	MM	MM	U	U	U	U	U
LAYN	NFR	2	From -150 to -57	MM	HM	HM	HM	HM	M/U	M	M	M	M
PAX6	NFR	3	From -143 to -18	MM	HM	HM	UM	UM	M/U	U	M	U	U
PPIC	NFR	2	From -174 to -79	UM	UM	UM	UM	UM	U	U	U	U	U
PYCARD	Inside	7	From 88 to 718	UM	MM	HM	MM	MM	U	M	M	U	M
RORA	Promoter	3	From -298 to -86	-	-	-	UM	UM	U	U	U	U	M/U
SFRP1	NFR	3	From -197 to -20	UM	MM	UM	UM	MM	U	M/U	U	U	M/U
SNAI1	NFR	2	From -173 to -100	MM	MM	MM	MM	MM	U	U	U	U	U
SPAG6	NFR	3	From -199 to -117	MM	HM	HM	HM	HM	U	M	M	M/U	M
TBX3	NFR	2	From -89 to -12	UM	UM	UM	UM	UM	U	U	U	U	U
TFAP2C	NFR	2	From -189 to -22	UM	UM	MM	UM	UM	U	U	U	U	U
TGFBR2	NFR	1	-31	UM	UM	UM	UM	UM	U	U	U	U	U
WDR21C	NFR	3	From -194 to -62	HM	HM	HM	HM	HM	M	M	M	M	M
ZNF177	NFR	2	From -172 to -77	UM	MM	UM	UM	MM	U	M	x	M/U	M

Results of MeDIP-CGI microarray were obtained using cutoff values of 0.6 and 0.4 for methylated and unmethylated CGIs, respectively. Number of probes, number of probes located in the analyzed region. Position of probes, relative location of probes from the TSS. HM, highly methylated. MM, moderately methylated. UM, unmethylated. -, DNA methylation status could not be determined because all the probes in a CGI were considered not functional in the analysis of the cell line. M, only methylated molecules were detected by MSP. M/U, partially methylated (both methylated and unmethylated DNA molecules were detected by MSP). U, only unmethylated molecules were detected by MSP. x, not amplified by MSP.

Table S2. The number of methylated and unmethylated CGIs obtained by MeDIP-CGI microarray analysis

	Prostate					Breast			
	Normal	Cancer				Normal	Cancer		
	RWPE1	PC3	LNCaP	22Rv1	Du145	HMEC	MCF7	ZR-75-1	MDA-MB-468
Promoter regions									
Highly methylated	481	867	1,294	929	849	292	697	824	950
Moderately methylated	1,390	2,024	2,059	1,459	1,873	762	2,027	2,562	2,117
Unmethylated	8,708	7,691	7,229	8,204	7,872	9,541	7,858	7,194	7,542
Divergent promoter regions									
Highly methylated	27	46	93	50	36	16	26	38	41
Moderately methylated	120	177	165	134	164	76	162	223	127
Unmethylated	1,315	1,241	1,207	1,285	1,268	1,376	1,276	1,201	1,305
Gene body regions									
Highly methylated	3,274	3,263	4,341	4,323	3,601	2,899	2,386	2,771	2,385
Moderately methylated	3,016	4,127	3,515	2,581	3,887	2,100	5,101	5,449	5,175
Unmethylated	10,793	9,691	9,220	10,187	9,595	12,084	9,591	8,863	9,536
Downstream of gene regions									
Highly methylated	336	386	506	469	390	264	290	323	288
Moderately methylated	362	444	352	246	397	244	525	570	531
Unmethylated	480	349	318	463	390	670	362	284	360
Unknown									
Highly methylated	1,073	1,031	1,404	1,219	1,240	876	719	799	724
Moderately methylated	1,305	1,739	1,424	916	1,549	899	2,030	2,099	2,056
Unmethylated	1,774	1,383	1,326	2,020	1,365	2,380	1,405	1,256	1,376

The numbers of CGIs with individual DNA methylation statuses are shown for various genomic regions. Promoter region, region within 10 kb upstream of the TSS. Divergent promoter region, region within 10 kb upstream of the TSSs of two genes that are transcribed in opposite directions.

Table S3. The association between the levels of candidate instructive factors and susceptibility to DNA methylation (S only)

	Lowest quintile	2nd quintile	3rd quintile	4th quintile	Highest quintile
Prostate					
H3Ac	1	0.73 (0.48-1.10)	0.79 (0.47-1.31)	0.68 (0.38-1.22)	0.56 (0.26-1.17)
H3K4m3	1	1.14 (0.75-1.73)	1.30 (0.81-2.08)	1.19 (0.71-2.00)	1.04 (0.60-1.81)
Pol II	1	0.91 (0.61-1.35)	0.64 (0.39-1.06)	0.44 (0.24-0.82)	0.21 (0.09-0.51)
H3K9me3	1	1.54 (0.97-2.45)	1.14 (0.70-1.86)	0.99 (0.60-1.62)	1.03 (0.63-1.70)
H3K27me3	1	0.64 (0.33-1.26)	1.06 (0.56-1.98)	1.71 (0.94-3.13)	5.63 (3.16-10.05)
Mammary gland					
H3Ac	1	0.86 (0.61-1.21)	0.95 (0.64-1.41)	1.07 (0.68-1.70)	1.51 (0.85-2.69)
H3K4m3	1	1.03 (0.74-1.45)	0.68 (0.46-1.00)	0.39 (0.25-0.62)	0.53 (0.32-0.89)
Pol II	1	1.14 (0.82-1.59)	1.01 (0.69-1.49)	0.86 (0.55-1.36)	0.40 (0.23-0.72)
H3K9me3	1	1.28 (0.90-1.82)	1.01 (0.70-1.46)	1.15 (0.78-1.68)	1.03 (0.70-1.52)
H3K27me3	1	0.95 (0.62-1.46)	1.13 (0.73-1.74)	1.73 (1.14-2.62)	6.56 (4.43-9.71)

Multivariate-adjusted odds ratio (OR) (95 % confidence interval; 95% CI) to become highly methylated (S1-S4 for the prostates; and S1-S3 for the mammary glands) is shown for each group. The multivariate-adjusted OR (95% CI) was derived from analyses in which all other listed variables were included into the model.

Table S4. The association between pol II binding and resistance to DNA methylation among genes involved in a specific biological process

	Enriched in	Term	Pol II-high				Pol II-low				Over-representation of S in pol II-low	Pol II high-to-low ratio
			S	R	Total	Fraction of S (%)	S	R	Total	Fraction of S (%)		
Prostate	R	Macromolecule metabolic process	57	1402	1459	3.9	123	561	684	18.0	4.60	2.13
	R	Biopolymer metabolic process	43	1112	1155	3.7	103	444	547	18.8	5.06	2.11
	R	RNA processing	13	301	314	4.1	65	116	181	35.9	8.67	1.73
	R	Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	39	868	907	4.3	83	328	411	20.2	4.70	2.21
	R	Protein transport	4	197	201	2.0	12	97	109	11.0	5.53	1.84
	S	Multicellular organismal process	36	408	444	8.1	127	230	357	35.6	4.39	1.24
	S	Multicellular organismal development	28	331	359	7.8	101	179	280	36.1	4.62	1.28
	S	System development	18	238	256	7.0	86	132	218	39.4	5.61	1.17
	S	Anatomical structure development	23	301	324	7.1	91	170	261	34.9	4.91	1.24
	S	System process	11	97	108	10.2	49	78	127	38.6	3.79	0.85
	S	Nervous system development	7	104	111	6.3	55	73	128	43.0	6.81	0.87
	S	Developmental process	35	516	551	6.4	108	250	358	30.2	4.75	1.54
	S	Organ development	17	187	204	8.3	60	92	152	39.5	4.74	1.34
	All genes		132	3167	3299	4.0	347	1400	1747	19.9	4.96	1.89
Mammary gland	R	Macromolecule metabolic process	59	1279	1338	4.4	176	595	771	22.8	5.18	1.74
	R	Nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	34	784	818	4.2	115	357	472	24.4	5.86	1.73
	R	Macromolecule localization	2	207	209	1.0	6	101	107	5.6	5.86	1.95
	R	Biopolymer metabolic process	43	1015	1058	4.1	146	466	612	23.9	5.87	1.73
	S	Multicellular organismal process	48	370	418	11.5	177	213	390	45.4	3.95	1.07
	S	Multicellular organismal development	37	293	330	11.2	139	158	297	46.8	4.17	1.11
	S	System development	25	218	243	10.3	109	114	223	48.9	4.75	1.09
	S	Anatomical structure development	30	272	302	9.9	122	153	275	44.4	4.47	1.10
	S	Developmental process	49	460	509	9.6	155	238	393	39.4	4.10	1.30
	S	Nervous system development	9	92	101	8.9	66	55	121	54.5	6.12	0.83
	S	Organ development	21	168	189	11.1	75	74	149	50.3	4.53	1.27
	All genes		155	2925	3080	5.0	506	1581	2087	24.2	4.82	1.48

Among genes involved in a specific biological process, genes were classified according to DNA methylation susceptibility and pol II binding levels, and numbers of genes in individual categories are shown. In the prostate and mammary glands, S1-S4 and S1-S3 genes, respectively, were classified as "S" genes, and resistant genes are listed as "R" genes. Pol II binding levels were classified as in Fig. 4. Biological processes that had more than 100 genes in pol II-high and -low categories are shown.

Table S5. Primers for quantitative ChIP-PCR

Gene symbol	Forward primer sequence	Reverse primer sequence	Reference
ARX	TATCTAACCCCACCCCCAAC	GCCCTCCTCCTGGTACTGATTG	this study
C11orf46	CTGTGTGGACTCGCATCTTG	TGTCTCCACTCTCCCACTCC	this study
CDKL2	GTTCCCAGGCTCCAGGTACTG	GAAGGCCAAGGGCTAGAGAG	this study
CNR1	GCAGAGCTCTCCGTAGTCAG	AACAGGCTGGGGCCATACAG	1)
DCC1	AGGGATCCGTGTCCCAAC	GCGACTGCAGGAAGGAAAC	this study
DLX2	TTGCTTCTGATTGGCTGTTG	AGCGGCTTTACGATTGTCTG	this study
EDIL3	GGAGGGAGGAAGGAGAGAGAG	GTTATTCACTCCCCGGGTCTC	this study
EID3	CTCTAACTGCCGCCACTTTC	ACGAGCTGCTCCTCTCCTTTC	this study
FZD1	CTCCCTCCTGGTGAAAGACAG	TGTGTCAATCCCTCAACTCG	this study
GAPDH	CGGCTACTAGCGGTTTTACG	AAGAAGATGCGGCTGACTGTC	this study
GDA	GCCTCAGCTGTGGTTGATCTTG	TGGGGTGTGGTCTTTCTCTC	this study
HCRTR2	TTCCTCAGCTGCCTATCTTCC	GCGCGGAGAGGCTAGAAG	this study
HPCA	CTGCGCAGTCGGTGTCTC	GTGTTGGGAAGCCCTGGAC	this study
HSPA4L	ATGTGCCTAGCCTCCTTTCC	AAATGGAGGCTGCTGAGCTATC	this study
JAK2	AGACAACTGTGACGGGCTTC	CCCTTCTGCTCCTCTTCCTC	this study
KCNA1	TGACGGTGATGTCTGGGGAG	GGTTGCGGTCGAAGAAGTAC	1)
MORC3	CCCTGCTTTAACCAGTCAGG	AACGACTGTGGAGCCCTATG	this study
NEUROG1	CGGGTACTTAAGGGGTCCTG	CCGGTCTCCTGAGTGATGTC	this study
Satellite II	GGAATTGCATGGAATCATCATA	GACCATTGGATGATTGCAGT	this study
SRD5A2	AAGCGGGAGGTGAATGTAAAG	GCAGCAATACCCCTTTCTCAAG	this study
TAC1	GTTATGGGCATCGACGAGTTAC	AGACCCACGTGACATTCTCC	this study
TBX2	ACCCACAATTGGTCCAAAAAG	GCGCGACTGGTTAGATCTTG	this study
TRIM58	TGGAAGGAAGGAGGGAATTTAG	CACACACACCCAGGAAACTG	this study
WSB2	TGCCCCATTTTTATCTGGTTC	CAGCTGAGGGAAAAGATGGA	this study
ZNF454	GTTCTAAAGCGAACGGAACG	CCAGACTCCACTTCCCAGAG	this study
ZNF571	TTTCTGGGGCGTTACTTGAG	GCTCCAGAACAGAACGATCC	this study
ZSWIM2	CCTCTGAAAGGCGAAGACAC	ATGTATTCCGGCTCCTCCTC	this study

1) Kirmizis, A. et al. (2004) Genes Dev, 18, 1592-1605.

Table S6. Primers for MSP

Gene symbol	Status	Chromosome	TSS (NCBI)	Position	Forward primer sequence	Position	Reverse primer sequence	Reference
ADFP	M	9	19117573	-169	GGTCGGGTTTTCGTTCGGTTTTC	-36	ACCCGAATATCACCCTCGAACACG	1)
	U			-170	AGGTTGGGTTTTTGTTTGGTTTTT	-36	ACCCAAATATCACCCTCAAACACA	1)
C6orf148	M	6	74029640	-122	GTTGGGTTTTTTGCGCGTC	1	ATAACGACGACCCCGCCG	2)
	U			-123	TGTTGGGTTTTTTGTGTGTT	-16	AAAATAACAACAACCCCACCA	2)
CTSL1	M	9	89530800	-182	GATTTTATTTTGCGTCGTTTC	-40	ACGCTACGATTAACTATACCG	1)
	U			-186	GTTTGATTTTATTTTGTGTTGTTTT	-40	ACTACACTACAATTAACTATACCA	1)
CYP26C1	M	10	94811011	-619	TTTTTTCGTGAGCGCGTC	-544	CCGATTCACACAACGTAACG	3)
	U			-621	AGTTTTTTTGTGAGTGTGTTGT	-544	TCCCAATTCACACAACATAACA	3)
DAZL	M	3	16622010	-103	GGTATCGGATTTGCGTATAC	-29	AAAAACTACGAAAAACGACG	2)
	U			-103	GGTATTGGATTTGTGTATAT	-29	AAAAACTACAAAAAACAACA	2)
DPEP3	M	16	66571867	-46	CGTCGGTTTTAGAGTCGC	65	CTACATATTACGCGAAAATCGACCG	this study
	U			-51	TGGTGTGTTGGTTTTAGAGTTGT	65	TAAACTACATATTACACAAAAATCAACCA	this study
F2R	M	5	76047542	-187	TTAGGAGGGTCGAGACGGTCGC	-96	TCCTCTAAACACCGTTAATTCG	1)
	U			-189	TTTTAGGAGGGTTGAGATGGTTGT	-98	TCCTCTAAACACCATTAATTCACA	1)
FADS1	M	11	61340886	-234	GTTCGTTTGACGTTAGGAAGTC	-34	GCCCAAAACCAACCGCCTACG	1)
	U			-234	GTTTGTTTGATGTTAGGAAGTT	-34	CACCCAAAACCAACCACCTACA	1)
FLJ21963	M	12	79995940	-232	CGTTTTGGAGTTTGAGCGC	-113	ACAAAAATATCACAACGCCG	2)
	U			-233	ATGTTTTGGAGTTTGAGTGT	-113	ACAAAAATATCACAACACCA	2)
FOXJ1	M	17	71648161	-162	ATGATGTTAGAGAGGGCGTTTTC	-25	TTACACGACCTCCCGAACG	this study
	U			-161	TGATGTTAGAGAGGGTGTTTTT	-25	CCATTACACAACCTCCCAAACA	this study
FOXO1 (FOXO1A)	M	13	40138734	-123	CGTATGTTTATTGGTCGCGC	-33	GACTTACGAAATCTACCGCCG	2)
	U			-124	TTGTATGTTTATTGGTTGTGT	-33	AACTTACAAAATCTACCACCA	2)
GREM1	M	15	30797497	-37	CGTCGGTATTTAAACGGGAGAC	63	GAAACTCGACGCGAAATCAACG	1)
	U			-134	TGTTGGTATTTAAATGGGAGAT	-35	CAAAACTCAACACAAAATCAACA	1)
GSTP1	M	11	67107862	-100	GCGGGATTTTTTAGAAGAGC	23	GCGCGTACTCACTAATAACG	this study
	U			-100	GTGGGATTTTTTAGAAGAGTGGTT	23	ACCACACATACTCACTAATAACA	this study
IGFBP7	M	4	57671296	-195	GGGTCGGTTACGTCGGGTGTTC	-18	GACAAAAACGCGAATAAACCG	1)
	U			-197	ATGGGTTGGTTATGTTGGGTGTTT	-18	CAACAACAAAAACACAAATAAACCA	1)
IL6R	M	1	152644293	-117	TTTTTATAGCGTAATTTCGTTTAC	78	AACCGAAACGAATAACGCAACA	1)
	U			-124	GGTGTGTTTTTTATAGTGTAATTTT	65	TAACACAACAACCCCACACACCA	1)
LAMA1	M	18	7107813	-333	GGATTGTAGGGTCGCGGC	-139	TAAATCCCGACGCACGCG	2)
	U			-335	GGGGATTGTAGGGTTGTGGT	-135	CCTCTAAATCCCAACACACACA	2)
LAYN	M	11	110916443	-315	TTTTGGATGTTATTCGCGC	-198	CCAACACGAAAAACGACG	2)
	U			-320	AAGTTTTTTGGATGTTATTTGTGTG	-194	AATCCCAACACAAAAAACAACA	2)

Table S6. Primers for MSP (continued)

Gene symbol	Status	Chromosome	TSS (NCBI)	Position	Forward primer sequence	Position	Reverse primer sequence	Reference
PAX6	M	11	31789455	-74	AGGGAGTATTTAATCGGTTGGC	68	CTCCTACGCCTAAACCAAAACG	1)
	U			-138	GTAATATTTTGTGTGAGAGTGAGT	-47	TCCTCCTACACCTAAACCAAAACA	1)
PPIC	M	5	122400324	-162	GTTTTTCGTATTCGTTTAAGGC	-33	AAAATAAAAATCGAACAATCCG	1)
	U			-165	GGTGTTTTTTGTATTTGTTTAAGGT	-57	AAAAACAAAAACCCAAAACACA	1)
PYCARD	M	16	31121752	-186	CGGGGAATCGCGGAGGTTTC	-36	AATAAAACCCGAAAAAAAACCG	1)
	U			-190	GGTTTGGGGAATTGTGGAGGTTTT	-13	ATCACACCCTCCAACTAACCTACA	1)
RORA	M	15	59308794	-213	GGTTGGAGAAGTTTTCGTTAGC	-111	GACGAACGAACAAACAAAAACG	1)
	U			-215	TTGGTTGGAGAAGTTTTTGTTAGT	-123	CAAACAAAAACACAAAAAAACACA	1)
SFRP1	M	8	41286137	-28	CGGTCGTAGGAGTTTCGC	54	GACTCCCGAAAATACGACG	this study
	U			-32	GGTTTGGTTGTAGGAGTTTTGT	54	CCCCAACTCCCAAAAATACAACA	this study
SNAI1	M	20	48032934	-155	ATTTGTTCGGGGAGTGGTTTTC	-91	AAAACGAAACCTTATCTACCACG	1)
	U			-213	GGAGTTTTTGTTTGGGTTTTTATT	-91	AAAAACAAAACCTTATCTACCACA	1)
SPAG6	M	10	22674405	-198	GGTAGTGTAGGGATATTCGAC	-62	ACGTAACGTCACGACCG	3)
	U			-199	GGGTAGTGTAGGGATATTTGAT	-62	CCAAACATAACATCACAACCA	3)
TBX3	M	12	113606352	-98	TTGGTTCGAAAGCGTTAAAGAG	-22	ACCGAACGTCTACTCGACGACT	1)
	U			-110	GTAGTAATATAATTGGTTTGAAAGT	-33	CTACTCAACAACTCTAAAAAATCA	1)
TFAP2C	M	20	54637765	-146	GCGTTGCGTTAGGTTCGGGTGC	40	CGCGAATATCAAAACCGCTCCG	1)
	U			-148	TGGTGTTGTGTTAGGTTTGGGTGT	40	ACCACAAATATCAAAACCACTCCA	1)
TGFBR2	M	3	30622998	-87	GTAGTTGAAAGTCGGTTAAAGTTTTC	-26	CGACGTCCAACCCCTAACTCTCTCG	4)
	U			-88	GGTAGTTGAAAGTTGGTTAAAGTTTTT	-26	TCAACATCCAACCCCTAACTCTCTCA	4)
WDR21C	M	8	88955412	-142	AGTAGCGGAATTTTTGTTTTAGATTAC	-35	CCTACGCCTATACGTATAAAATCG	this study
	U			-142	AGTAGTGGAATTTTTGTTTTAGATTAT	-40	ACACCTATACATATAAAATCAACCCA	this study
ZNF177	M	19	9334696	-119	GTAGGAGTATTTGCGATGTTTC	-12	AAAATAACGAAACGACGAACG	1)
	U			-97	GTTTTTAAGTTTTTAGGGTGAATTT	-22	AAACAACAAACACCCACTTCCA	1)

1) Yamashita, S. et al. (2006) Cancer Sci, 97, 64-71.

2) Moriguchi, K. et al. (2007) Cancer Lett, 249, 178-187.

3) Abe, M. et al. (2008) Oncology, 74, 50-60.

4) Yamashita, S. et al. (2008) Cancer Res, 68, 2112-2121.

Table S7. Primers for quantitative RT-PCR

Gene Symbol	Forward primer sequences	Reverse primer sequences	Anneal (°C)	Reference
ABHD9	ACATCCTGCCAGGCATAGGG	CAGGCCAGCAAGGACCACT	56	2)
ADFP	ACACCCTCCTGTCC ACATC	AAGTGAGGAGGCTGTCAGAC	57	2)
ALDH1A3	ATTCTTCTGGAGGCTTTACA	TCACAGACAACCTATAGGCA	56	2)
BDNF	AACTACCCAGTCGTACGTGC	CCCCTTTTAATGGTCAATGTA	56	2)
BNIP3	CTATATTGGAAGGCGTCTGAC	CACCCCAGGATCTAACAGCTC	56	1)
CFL2	GGAAGCAAAGCAGATCTTGG	TGTTTCGTATGTGGCATCG	62	This study
COL4A1	CGCTGCCAAGTCTGTATGAGA	GGGTTCGTTGCTGTTAACAAA	55	1)
F2R	TCTCAGGAACCCCAATGAT	AGGAGCTGGTCAAATATCCG	54	2)
FBN2	CAAGAAGAAGGAGCTTAAGAA	CCAAGTCTGTGAAGGGTTAAT	56	1)
FDX1	TGGCTTGTTCAACCTGTCAC	TGTTTCAGGCACTCGAACAG	62	This study
FSD1	TCCAACACCAGCCTCACCTA	GCTCCAAGCAGCAAGTGACA	57	2)
FST	AGATGAAGACCAGGACTACA	CACAAAGGCTATGTCAACAC	56	2)
GAPDH	AGGTGAAGGTCGGAGTCAACG	AGGGGTCATTGATGGCAACA	54	This study
GREM1	CTCAACTGCCCTGAACTACAG	TGCACCTGGATTTGGCTTAAT	51	2)
GSTP1	TCGCTGACTACAACCTGCTG	GGGAGGTTCACGTACTCAGG	56	This study
ICAM1	TGTCCCCCTCAAAAGTCATC	TAGGCAACGGGGTCTCTATG	62	This study
IGFBP7	GGGTGCTGGTATCTCCTCTAAGT	TAAGGCATCAACCACTGTAATTT	56	1)
IL6R	AGGAAGGCAAGACAAGCAT	GGGAGATGAGAGGAACAAGC	59	2)
MARK1	GAGGGTAAAGATTCTAAGCC	GCATCTAACACTTTTCGGA	54	2)
MLF1	CTCCACATCAAAGGCTCATC	TGCCAAGGACAGAACTGATT	56	2)
NAALAD2	CCAGGAAAGCTGTTCTATAGGCA	TTCCAGGCCAAACGAGAGT	64	1)
NAP1L1	GAAAATGATCCAGACTATGACCC	CCTCAAGGCCACATACATC	56	1)
PLAGL1	TGTGAGAAGACGTTCAACCG	GCCAGGTGCCTCTTATAGCC	55	2)
RBP4	GGAGTTTAATTTGCCCTTC	AACTTTCAGGAAAGGCAAGC	56	2)
SOCS3	AGGCTCCTTTGTGGACTTCAC	AATCGAAGTCTCCGTCCTTG	62	This study
TBX3	CTGTCTTCTTGGCGTGGT	CCCCAGTAGCTCAATGCAAC	56	1)
TSPAN5	TGCAAGACACTGGACAGACC	TGTCAGTGAGCAGCATTTCC	62	This study
TSPYL6	AGAGTCCGACAGGATTGCTCA	AGGCGACGTCTAGCTCTAT	64	1)
TUBB6	CGCCCCAACTCAGATCCTACAAC	GTGAGGGGCCGACACCAAC	68	1)
ULBP2	CACGGTCTTGATCAAACTCG	GAATGAGCTATTGGGTCCAT	53	2)
WIF1	ATTACATCTGGTGAACTCCG	CAGGCCAGTATTCTTAAGTG	51	2)

1) Moriguchi, K. et al. (2007) Cancer Lett, 249, 178-187.

2) Nakajima, T. et al. (2009) Int J Cnacer, 124, 905-910.