Regional activation of the cancer genome by long-range epigenetic remodeling

Bert, Saul A ; Robinson, Mark D ; Strbenac, Dario ; Statham, Aaron L ; Song, Jenny Z ; Hulf, Toby ; Sutherland, Robert L ; Coolen, Marcel W ; Stirzaker, Clare ; Clark, Susan J

Abstract: Epigenetic gene deregulation in cancer commonly occurs through chromatin repression and promoter hypermethylation of tumor-associated genes. However, the mechanism underpinning epigenetic-based gene activation in carcinogenesis is still poorly understood. Here, we identify a mechanism of domain gene deregulation through coordinated long-range epigenetic activation (LREA) of regions that typically span 1 Mb and harbor key oncogenes, microRNAs, and cancer biomarker genes. Gene promoters within LREA domains are characterized by a gain of active chromatin marks and a loss of repressive marks. Notably, although promoter hypomethylation is uncommon, we show that extensive DNA hypermethylation of CpG islands or "CpG-island borders" is strongly related to cancer-specific gene activation or differential promoter usage. These findings have wide ramifications for cancer diagnosis, progression, and epigenetic-based gene therapies.

DOI: https://doi.org/10.1016/j.ccr.2012.11.006

Cancer Cell
Article

Regional Activation of the Cancer Genome by Long-Range Epigenetic Remodeling

Saul A. Bert,[1] Mark D. Robinson,[1,2,3] Dario Strbenac,[1] Aaron L. Statham,[1] Jenny Z. Song,[1] Toby Hulf,[1] Robert L. Sutherland,[1,5,6] Marcel W. Coolen,[1,4] Clare Stirzaker,[1,5] and Susan J. Clark[1,5,*]
[1]Cancer Research Program, Garvan Institute of Medical Research, Darlinghurst, New South Wales 2010, Australia
[2]Bioinformatics Division, Walter and Eliza Hall Institute of Medical Research, Parkville, Melbourne, Victoria 3050, Australia
[3]Institute of Molecular Life Sciences, University of Zurich, 8057 Zurich, Switzerland
[4]Department of Human Genetics, Nijmegen Centre for Molecular Life Sciences (NCMLS), Radboud University Nijmegen Medical Centre, P.O. Box 9101, 6500 HB, Nijmegen, The Netherlands
[5]St. Vincent's Clinical School, Faculty of Medicine, University of New South Wales, Darlinghurst, New South Wales 2010, Australia
[6]Robert L. Sutherland deceased.
*Correspondence: s.clark@garvan.org.au
http://dx.doi.org/10.1016/j.ccr.2012.11.006

SUMMARY

Epigenetic gene deregulation in cancer commonly occurs through chromatin repression and promoter hypermethylation of tumor-associated genes. However, the mechanism underpinning epigenetic-based gene activation in carcinogenesis is still poorly understood. Here, we identify a mechanism of domain gene deregulation through coordinated long-range epigenetic activation (LREA) of regions that typically span 1 Mb and harbor key oncogenes, microRNAs, and cancer biomarker genes. Gene promoters within LREA domains are characterized by a gain of active chromatin marks and a loss of repressive marks. Notably, although promoter hypomethylation is uncommon, we show that extensive DNA hypermethylation of CpG islands or "CpG-island borders" is strongly related to cancer-specific gene activation or differential promoter usage. These findings have wide ramifications for cancer diagnosis, progression, and epigenetic-based gene therapies.

INTRODUCTION

The individual epigenome of each cell type is formed during early development and combines CpG DNA methylation and histone modifications to orchestrate or mark tissue-specific gene expression patterns. In normal cells, the bulk of the genome is DNA methylated, but CpG-island-associated promoters of active or bivalent genes commonly remain unmethylated (Cedar and Bergman, 2009). The lack of methylation at CpG island promoters is still an enigma, because although CpG dinucleotides are the primary target for DNA methyltransferase enzymes (DNAMTases), they remain essentially resilient to DNA de novo methylation during normal development and differentiation. Whether this is due to active or poised transcription and/or binding of the transcriptional machinery that obscures CpG sites from DNAMTases, or to active demethylation remains to be resolved (Clark and Melki, 2002). In cancer, however, CpG island promoters are commonly hypermethylated, and this methylation is associated with gene repression and gain of histone repressive marks (Jones and Baylin, 2007). Our group and other investigators (Coolen et al., 2010; Frigola et al., 2006; Hsu et al., 2010) have reported that epigenetic inactivation is not limited to single genes but can also encompass large domains across the genome during tumorigenesis, through long-range epigenetic silencing (LRES). The characteristics of LRES are generally typified by concordant increases in CpG island hypermethylation and gain or reinforcement of the repressive histone modifications H3K9 dimethylation (H3K9me2) and H3K27 trimethylation

Significance

Epigenetic changes, including alterations in histone modifications and DNA methylation, commonly occur in cancer and are associated with aberrant gene expression. However, most studies have focused on epigenetic gene-silencing events, and thus, the mechanism that promotes gene activation in carcinogenesis is still poorly appreciated. Here, using an integrated genomics approach in prostate cancer, we identify a mechanism of regulation that involves coordinated epigenetic activation of multiple adjacent genes by remodeling of chromatin and DNA methylation patterns across domains. The activated regions commonly contain key tumor genes, most notably the prostate cancer biomarker genes *PSA* and *PCA3*, which have not previously been reported to be epigenetically regulated. Importantly, our study reveals a paradigm in epigenetic cancer gene deregulation that promotes widespread oncogenic gene activation in tumorigenesis.

(H3K27me3), in conjunction with the loss of active H3K9 acetylation (H3K9ac) and H3K4 trimethylation (H3K4me3; Coolen et al., 2010).

Studies examining the underlying mechanism of epigenetic deregulation in cancer have primarily concentrated on DNA hypermethylation and gene silencing, rather than DNA demethylation and epigenetic gene activation. However, genome-wide hypomethylation, initially reported by Feinberg et al. (1988), is one of the primary epigenetic aberrations found in tumors. Recently, it was reported that long-range hypomethylation domains cover nearly half the cancer genome (Berman et al., 2012) and commonly occur at partially methylated domains (PMDs) in somatic cells (Hansen et al., 2011; Hon et al., 2012). Traditionally, cancer-associated hypomethylation was attributed to demethylation of the pervasive LINE-1 elements, as well as other repeat sequences (Chalitchagorn et al., 2004; Ehrlich, 2002). More recently, demethylation of repeats was causally implicated in the activation of alternative transcripts (Wolff et al., 2010) and overexpression of oncogenes (Lamprecht et al., 2010). CpG demethylation of gene promoters has also been shown for several individual genes in cancer, including *R-RAS* (Nishigaki et al., 2005) and cancer-testis antigens such as the *MAGE*, *GAGE*, and *XAGE* families (Grunau et al., 2005; Lim et al., 2005).

Even though epigenetic activation of specific genes has been documented in cancer (reviewed in Ross et al., 2010), as yet no genome-wide studies have specifically addressed the extent and genomic context of epigenetic activation in cancer. Here, we investigate the prevalence of regional activation in prostate cancer and determine whether there are predominant chromatin and DNA methylation changes associated with cancer-specific gene activation covering large genomic domains.

RESULTS

Activated Domains Are Common in Prostate Cancer Cells

To identify potential LREA regions in prostate cancer, we used prostate cancer and normal prostate cell lines as well as publicly available clinical expression data sets to categorize regions that commonly display concordant cancer-associated gene activation. First, a list of transcriptionally upregulated regions was created using Affymetrix Human Gene 1.0ST expression array data, carried out in two normal primary prostate cells (PrECs) and three prostate cancer cell lines (LNCaP, DU145, and PC3). We preprocessed the data using robust multichip analysis (RMA; Irizarry et al., 2003) and calculated the moderated t-statistics (using limma; Smyth, 2004), representing the change in expression in LNCaP over PrEC cells, for each represented gene. The median t-statistic over a sliding window of five genes was plotted for each chromosome as a representation of local up- or downregulation (Figure S1A available online). Forty-two activated domains were identified that had a median t-statistic above 4. We found that 43% and 57% of the activated domains identified in LNCaP cells were also consistently activated in PC3 and DU145 cells, respectively (Table 1). Figure 1A displays chromosome 19, which shows two activated domains (regions 27 and 28), each of which harbors cancer-specific overexpression of neighboring genes (Table 1). Notably, region 28 contains a subset of the Kallikrein gene family (*KLK15-KLK4*), which also includes *KLK3*, commonly known as prostate-specific antigen (*PSA*; Figure 1B). Interestingly, region 28 is adjacent to a region of transcriptional repression that also contains a subset of the Kallikrein gene family (*KLK5-KLK12*). We validated the gene expression array data using quantitative RT-PCR (qRT-PCR) on genes from region 28 (Figures 1D and 1E) and regions 14, 23, and 25 (Figure S1B), and in all cases confirmed concordant gene activation in cancer relative to normal cells.

Second, to exclude domains that showed a local increase in expression due to potential copy number amplification, we used genomic DNA inputs hybridized to Affymetrix Promoter 1.0R arrays to estimate promoter-level copy number changes between the cancer cell lines and PrEC (see Supplemental Experimental Procedures). Seven domains were excluded from further analysis because they showed increased copy number ($p < 0.05$) in LNCaP cells (Figure S1C), leaving 35 activated regions. Unexpectedly, two domains showed a significant loss of DNA copy number (4q31.1 and 14q11.1-q11.2) despite concordant gene activation.

Third, to confirm the veracity and relevance of the activated domains in clinical prostate cancer, we analyzed gene expression across clinical samples from nine large Oncomine prostate cancer studies. This allowed us to compare expression in 215 normal prostate and 380 local prostate cancer samples, as exemplified in Figure 1C for region 28 (19q13.33). The Oncomine data from all activated domains are summarized in Figure S1D. We found that 74% (26/35) of the activated domains identified in LNCaP cells were also consistently activated in clinical prostate cancer (Table 1).

Activated Domains Harbor Cancer-Related Genes

Using this rigorous approach, we identified 35 candidate LREA regions harboring 251 genes that showed concordant transcriptional activation (Table 1). The LREA regions span 32.5 Mb (~1% of the genome) and range in size from 85.5 kb to 5.2 Mb. Activated domains were identified on all chromosomes except for 2, 17, 18, 21, and the Y chromosome (Figure S1A), with chromosomes 7, 11, and 12 having the highest coverage (3.3%, 3.9%, and 3.1%, respectively; Table 1). Each region contained on average seven genes, with a mean 5.96-fold change in expression in LNCaP compared with PrEC cells; indeed, 65.7% of contained genes showed at least a 1.5-fold increase in gene expression. Compared with normal genomic distribution, we found no general significant increase in the density of genes or CpG islands, or of SINE, LINE, Long Terminal Repeat, or simple repeat density in activated regions (Figure S2).

Notably, 15% of the 35 activated domains we identified contained gene clusters (Table 1). These included the *MAGE* (Xq28: region 35) and *GAGE* cancer-testis antigens (Xp11.23: region 32), UDP-glucuronosyltransferase type 2 family genes (*UGT2*; 4q13.2: region 5), as well as genes from the Kallikrein gene family (*KLK*; 19q13.33: region 28). We found that several prostate-cancer-associated genes were also located within the activated regions. In particular, two of the most sensitive prostate cancer biomarkers, *KLK3* (*PSA*; Lilja et al., 2008) and prostate cancer antigen 3 (*PCA3*; Deras et al., 2008) were located within regions 28 and 14, respectively. In addition, the LREA regions harbored several genes, including *C15orf21*, *KLK2*,

Table 1. Summary of LREA Domains in Prostate Cancer

Identifier	Band	Chromosome Coordinates	Size (kb)	No. of Genes[a]	Average Fold Change	% Up in Oncomine	miRNA/ Oncogene	Genes
1 * ^	1q21.2	chr1: 149,035, 308 - 149,234,738	199.4	5	2.07	70	*ARNT*	CTSK, **ARNT**, **SETDB1**, **LASS2**, ANXA9
2 * ^	1q23.3	chr1: 159,357, 388 - 159,457,113	99.7	8	1.69	68.8		**DEDD,** UFC1, **USP21, PPOX, B4GALT3,** ADAMTS4, NDUFS2, FCER1G
3 * ^	3q13.2-q13.31	chr3: 114,948, 598 - 115,164,901	216.3	3	2.552	66.7		ATP6V1A, **GRAMD1C, ZDHHC23**
4 * ^	3q13.33	chr3: 121,797, 818 - 122,748,178	950.4	6	16.96	70		NDUFB4, HGD, RABL3, GTF2E1, **STXBP5L, POLQ**
5	4q13.2	chr4: 68,995, 776 - 70,396,212	1,400.4	11	22.18	40		TMPRSS11E, UGT2B17, UGT2B15, TMPRSS11E, UGT2B15, UGT2B10, UGT2A3, UGT2B11, UGT2B7, UGT2B28, UGT2B4
6 *	4q22.1	chr4: 89,398, 960 - 89,848,709	449.7	5	2.03	60		**PPM1K, HERC6, HERC5, PIGY, HERC3**
7 *	4q31.1	chr4: 140,594, 411 - 141,294,683	700.3	5	2.53	62.5		**RAB33B, SETD7,** MGST2, H3F3A, **MAML3**
8 ^	5p13.2	chr5: 36,139, 171 - 36,724,193	585	5	4.71	50	mir-580	**LMBRD2,** SKP2, **C5orf33,** RANBP3L, SLC1A3
9 ^	6q21	chr6: 110,608, 037 - 111,453,996	846	8	2.63	50		**CDC40, C6orf186,** DDO, **SLC22A16, CDC2L6, AMD1, GTF3C6, BXDC1**
10 * ^	7p12.2-p12.1	chr7: 49,947, 565 - 53,224,113	3,276.5	10	2.30	66.7	*IKZF1*	**ZPBP,** LOC100130988, LOC100130988, IKZF1, **FIGNL1,** DDC, LOC100129427, GRB10, **COBL, DKFZp564N2472**
11 * ^	7q21.13-q21.2	chr7: 89,621, 625 - 91,577,925	1,956.3	9	4.89	55.6	*AKAP9*	**STEAP1, STEAP2, C7orf63, GTPBP10,** CLDN12, PFTK1, **FZD1, MTERF, AKAP9**
12 * ^	8p21.2-p21.1	chr8: 27,224, 916 - 27,528,288	303.4	4	3.96	16.7		**PTK2B,** CHRNA2, **EPHX2,** CLU
13 * ^	8q21.13	chr8: 80,685, 869 - 82,186,858	1501	5	2.41	80		**STMN2, HEY1,** MRPS28, TPD52, **ZBTB10**
14	9q21.13-q21.2	chr9: 78,190, 253 - 79,453,043	1,262.8	9	16.54	58.3		**RFK,** GCNT1, PCA3, **PRUNE2, FOXB2,** LOC645225, **VPS13A,** LOC642947, **GNA14**
15 ^	10q11.21	chr10: 43,201, 071 - 43,464,332	563.3	4	3.11	50		**HNRNPF, ZNF239, ZNF485, ZNF32**
16 * ^	11p15.1-p14.3	chr11: 20,365, 679 - 25,566,989	5,201.3	10	2.6	55.6	*FANCF*	**PRMT3, SLC6A5, NELL1, ANO5,** SLC17A6, **FANCF, GAS2, SVIP, LUZP2,** LOC554234
17 * ^	12p11.21	chr12: 31,118, 046 - 31,773,319	655.3	8	1.7	40		DDX11, OVOS2, FAM60A, **FLJ13224, DENND5B,** AK3L1, C12orf72, **AMN1**
18 * ^	12q14.2-q14.3	chr12: 63,084, 500 - 63,801,383	716.9	6	6.68	60	mir-548c	**XPOT, TBK1, RASSF3, GNS, TBC1D30, WIF1**
19 * ^	12q21.31	chr12: 80,177, 487 - 82,052,212	1,874.7	4	8.83	75		**PPFIA2,** CCDC59, **C12orf26, TMTC2**
20 * ^	12q23.2	chr12: 100,073, 402 - 100,979,977	906.6	10	3.33	44.4		**SLC5A8, UTP20, ARL1,** SPIC, MYBPC1, **CHPT1, SYCP3, GNPTAB, DRAM, CCDC53**

(*Continued on next page*)

Table 1. *Continued*

Identifier	Band	Chromosome Coordinates	Size (kb)	No. of Genes[a]	Average Fold Change	% Up in Oncomine	miRNA/ Oncogene	Genes
21 *	13q12.12	chr13: 22,653,091 - 23,779,210	11,262.1	7	2.44	50		SGCG, **SACS,** TNFRSF19, **MIPEP,** PCOTH, FLJ46358, **SPATA13**
22 * ^	14q11.1-q11.2	chr14: 18,447,594 - 19,251,915	804.3	6	3	0		OR11H1, A26C2, LOC440157, OR11H1, A26C2, OR11H1
23 * ^	14q13.3-q21.1	chr14: 36,218,829 - 37,752,019	1,533.2	6	9.14	70		**SLC25A21, MIPOL1, FOXA1, C14orf25,** TTC6, **SSTR1**
24	15q11.2	chr15: 22,705,169 - 22,888,117	182.9	3	2.76	100		**SNRPN,** SNRPN, SNRPN
25 * ^	15q21.1	chr15: 43,561,968 - 43,689,685	127.7	3	8.85	66.7	*C15orf21*	**SLC30A4,** C15orf21, **PLDN**
26	16p12.2	chr16: 20,542,060 - 20,768,491	226.4	5	5.07	90		ACSM1, **THUMPD1,** ACSM3, **EXOD1, LOC81691**
27 * ^	19p13.2	chr19: 8,036,287 - 8,293,278	257	5	2.29	50		**FBN3, LASS4, CD320, NDUFA7,** RPS28
28 *	19q13.33	chr19: 56,020,357 - 56,105,806	85.5	5	10.21	100	*KLK2*	KLK15, KLK3, KLK2, KLKP1, **KLK4**
29 * ^	20q11.21-q11.22	chr20: 31,334,602 - 31,737,871	406.3	8	1.75	64.3		C20orf114, **CDK5RAP1, SNTA1,** CBFA2T2, **NECAB3, C20orf144, C20orf134, E2F1**
30 * ^	22q11.21	chr22: 16,973,242 - 17,159,474	186.2	5	2.93	33.3		**TUBA8, USP18,** DKFZP434B061, LOC728212, GGT3P
31 * ^	22q11.21	chr22: 18,681,799 - 19,092,752	411	10	2.10	50		DGCR6L, LOC728212, **DGCR6L,** TMEM191B, PI4KAP2, RIMBP3B, LOC728212, LOC728212, USP18, **ZNF74**
32 *	Xp11.23	chrX: 48,978,871 - 49,347,307	368.4	13	30.57	0		**CCDC22,** FOXP3, **PPP1R3F,** GAGE4, GAGE13, GAGE12C, GAGE13, GAGE12C, GAGE12B, GAGE12B, GAGE12C, GAGE12C, PAGE1
33 * ^	Xp22.22	chrX: 52,689,836 - 54,488,645	1,798.8	20	1.73	46.4	mir-98/let-7f/ *SSX2*	SSX7, SSX2, SSX2, SPANXN5, XAGE5, XAGE3, FAM156A, FAM156A, GPR173, **TSPYL2, JARID1C, IQSEC2, SMC1A,** RIBC1, **HSD17B10, HUWE1, PHF8, FAM120C, WNK3,** TSR2
34 * ^	Xq11.1	chrX: 62,435,851 - 63,342,349	906.5	4	2.01	62.5		LOC645251, **SPIN4, ARHGEF9, FAM123B**
35 * ^	Xq28	chrX: 151,033,182 - 151,688,896	655.7	12	10.37	28.6	mir-767/ mir-105	MAGEA5, MAGEA10, GABRA3, **GABRQ,** MAGEA6, CSAG2, MAGEA2, **MAGEA12,** CSAG1, MAGEA2, CSAG2, MAGEA3
		Average	926.81	7.06	5.96	55.77		

See also Figure S2 and Table S1.

[a]Number of genes in LREA regions in LNCaP cells.

*and ^symbols indicate regions that also contain genes upregulated in PC3 or DU145 cells, respectively, compared with PrEC cells. Underlined genes indicate those within gene families, and bold indicates genes that contain CpG-island-associated promoters. Tumor genes denoted by the Wellcome Trust Sanger Institute Cancer Genome Project (http://www.sanger.ac.uk/genetics/CGP/Census).

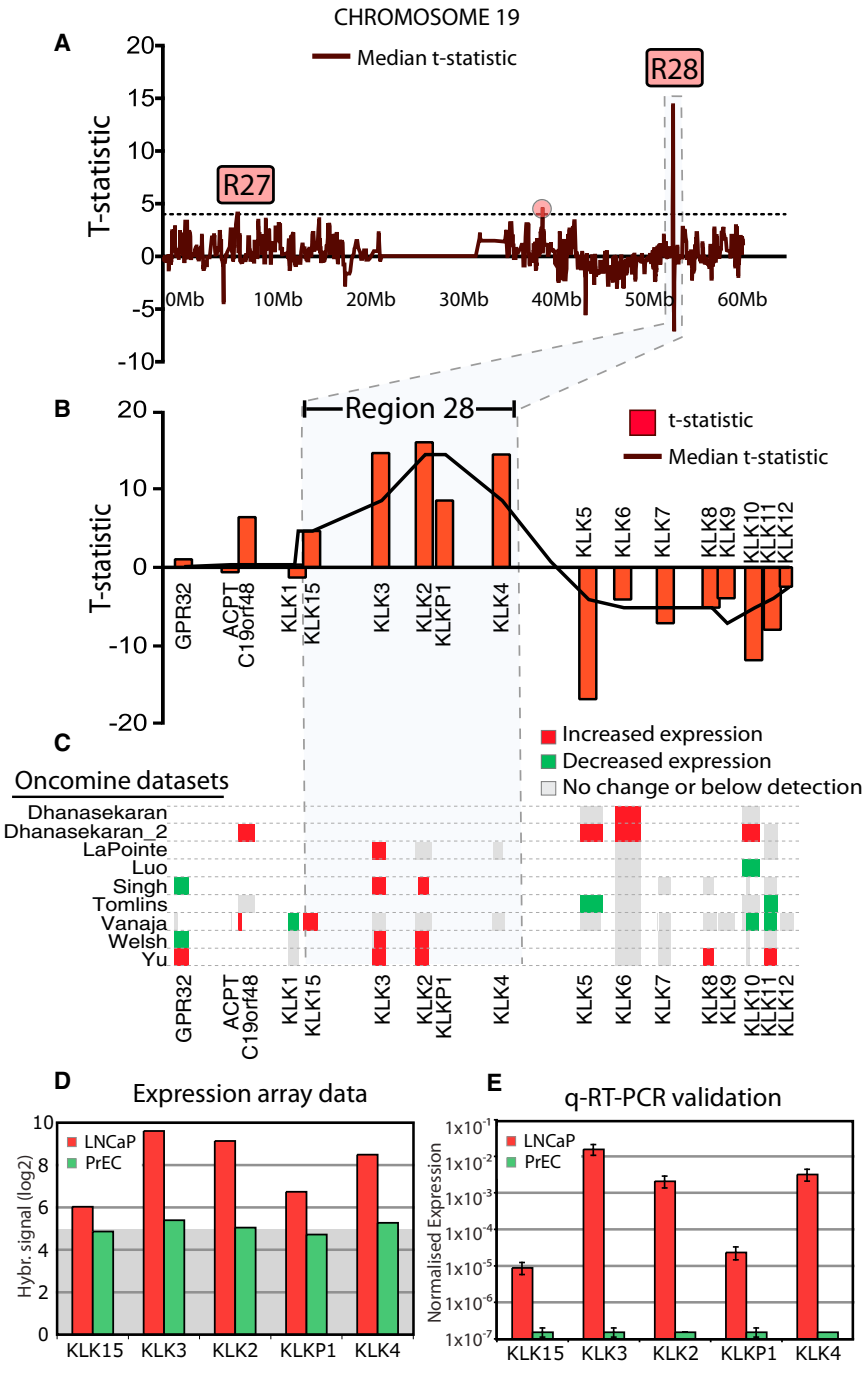

Figure 1. Example of Gene-Activated Domains in Prostate Cancer Cells

(A) The median t-statistic, representing change in expression from LNCaP to PrEC over five genes, is plotted for every locus on chromosome 19. Designated regions 27 (R27) and 28 (R28) are highlighted. The dotted line marks a median t-statistic of 4, above which activated regions were identified.

(B) Region 28 is magnified; red bars represent the calculated t-statistic for each gene labeled on the x axis. The line shows the median t-statistic over five genes. The length of the x axis represents chromosomal coordinates chr19:55,965,000–56, 231,000 (GPR32-KLK12).

(C) Summarized data extracted from nine Oncomine prostate cancer studies are plotted, aligned to genomic coordinates spanning region 28. Red, green, and gray boxes represent probe sets with increased, decreased, and unchanged (or below detection) expression, respectively.

(D) Expression data derived from Affymetrix expression arrays for genes in region 28. The gray box indicates values for which expression is considered background.

(E) Validation of gene expression in region 28 by qRT-PCR. Data are normalized to expression of 18S; error bars indicate ±SD.

See also Figure S1.

and *MIPOL1*, that are implicated in translocations and gene fusions with Ets-transcription factors in prostate cancer (Hermans et al., 2008; Tomlins et al., 2007). We confirmed that translocation of the Ets-transcription factor *ETV1* gene into an intron of *MIPOL1* (14q13.3-q21.1: region 23) occurred in LNCaP cells (Figure S2), as previously reported (Tomlins et al., 2007). Other tumor-associated genes were also located in the activated domains, including *IKZF1* (7p12.2-p12.1: region 10; Mullighan et al., 2009), *FANCF* (11p15.1-p14.3: region 16; Lim et al., 2008), *ARNT* (1q21.2: region 1; Salomon-Nguyen et al., 2000), *AKAP9* (7q21.13-q21.2: region 11; Ciampi et al., 2005), and *SSX2* (Xp22.22: region 33; de Leeuw et al., 1995; Table 1). Functional annotation clustering using DAVID analysis showed significant enrichment in several gene families, including the *MAGE* and *UGT2* families (Table S1).

Epigenome Analysis of LREA Regions

Next, we assessed whether these 35 activated regions also exhibited significant epigenomic changes. We investigated the relative enrichment of the active chromatin marks (H3K9ac and H3K4me3) and repressive marks (H3K27me3 and H3K9me2) in PrEC and LNCaP cells across the domains using Affymetrix Human Promoter 1.0R arrays (chromatin immunoprecipitation [ChIP] on chip). We analyzed DNA methylation using a method that incorporates methyl-CpG binding domain capture with deep sequencing (MBDCap-seq; Robinson et al., 2010). Summaries of tiling array signals for the 85 kb Kallikrein region (19q13.33: region 28), where we observed a broad reorganization of the epigenetic landscape, are shown in Figure 2. All of the genes gained the active H3K9ac mark at their promoters while they simultaneously lost the H3K27me3 repressive Polycomb mark. Changes in H3K9me2 and H3K4me3 were more discrete, with losses of H3K9me2 at the promoter of *KLK2*, and gains of H3K4me3 at the promoter of *KLK4*. There was minimal change of DNA methylation in regions of low CpG

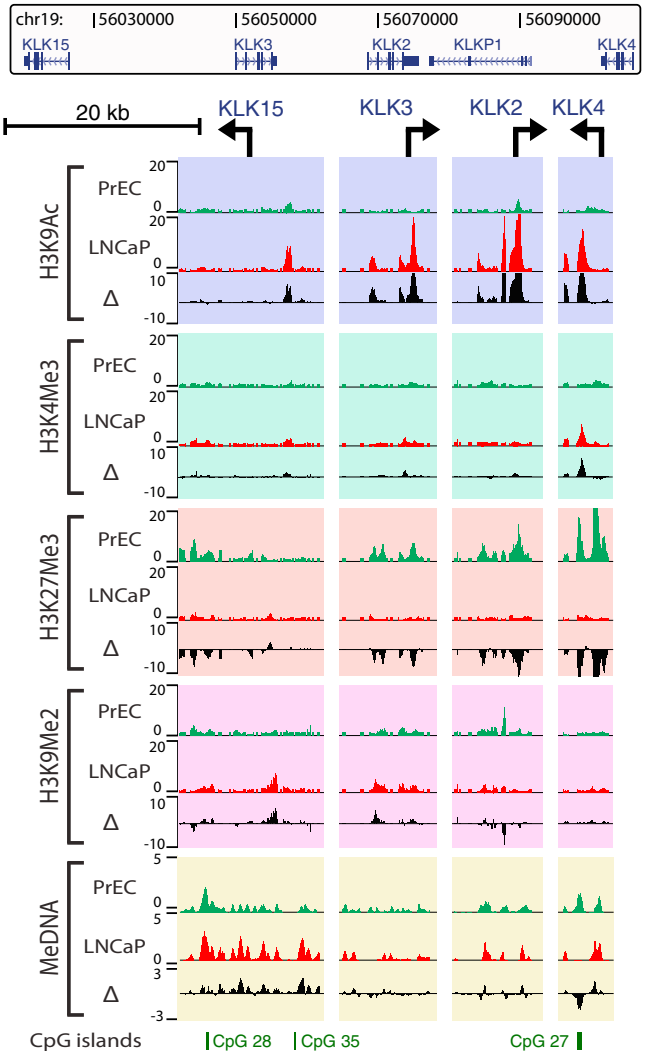

Figure 2. Epigenome Plots across an Example Activated LREA Domain

Histone modifications (H3K9ac, H3K4me3, H3K27me3, and H3K9me2) and DNA methylation profiles (MBDCap-seq) are shown for each TSS from the Kallikrein gene subfamily in region 28 (chromosome 19). For each gene and each modification, the enrichment over input status is shown (green, PrEC; red, LNCaP; black, differential LNCaP-PrEC). Locations of CpG islands in region 28 are indicated. Black arrows mark the TSS for each gene. Note that *KLKP1* has no probes on the Affymetrix promoter array.
See also Figure S3.

density (non-CpG islands) or in regions of high CpG density, including CpG_28 (in the body of *KLK15*), which remained methylated and CpG_35 (4.6 kb upstream of *KLK15*), which remained unmethylated. As discussed in more detail later, CpG_27 (in the body of *KLK4*) was one of the few CpG islands found in LREA regions to be hypomethylated in LNCaP cells (Figure 2).

Interestingly, we found that LREA region 28 is juxtaposed to a neighboring LRES region containing Kallikrein genes (*KLK5*–*KLK12*; Figure 1B). This region also undergoes epigenetic remodeling, but, in contrast to LREA region 28, gains repressive epigenetic marks (LRES; Figure S3). The entire 92 kb region gains the H3K27me3 repressive modification with localized enrichment of H3K9me2. In addition, promoter-associated CpG island hypermethylation (CpG_31, CpG_65, and CpG_69) is associated with repression of *KLK8* and *KLK10* in LNCaP cells. We hypothesized that the boundary region spanning the LREA and LRES might show differential CTCF binding in PrEC and LNCaP cells. However, using CTCF-seq, we found that even though there are two clear CTCF binding sites within the LREA/LRES Kallikrein boundary region, there is little difference in CTCF binding affinities between the normal and prostate cancer cells (Figure S3). In addition, we found no clear differences in CTCF binding flanking the domain boundaries of the other LREA regions (data not shown).

Activated Domains Are Epigenetically Deregulated in Cancer Cells

Gene expression and histone modifications were analyzed collectively for the 251 gene promoters in the 35 common LREA regions in LNCaP cells. Overall, we found a significant enrichment of the active H3K9ac modification at gene promoters (±1,000 bp) from the transcription start site (TSS) and a general depletion of H3K27me3 across the promoter region within the activated domains compared with randomized gene sets (Figure S4). Although no such overall enrichment or depletion was found for H3K4me3 or H3K9me2 modifications, changes in these marks did occur in localized regions.

From a detailed epigenetic analysis, it is evident that LREA-associated changes occur mainly in blocks of multiple consecutive genes; however, various combinations of changes are observed (Figures 3A–3E). For example, region 5 (4q13.2; *UGT* gene family), region 23 (14q13.3; includes *SLC25A21*, *MIPOL1*, *FOXA1*, *C14orf24*, *TT6*, and *SSTR1*), and region 28 (19q13.33; Kallikrein gene family) all display a regional increase in H3K9ac together with a regional loss of H3K27me3 (Figure 3A), whereas region 4 (3q13.33), region 9 (6q21), and region 13 (8q21.13) all show global increases in H3K9ac and H3K4me3 (Figure 3B). In addition to those regions that exhibit combinations of epigenetic alterations, many LREA regions show global changes predominantly in only one epigenetic mark. For example, regions 14 (9q21.13-q21.2) and 26 (16p12.2) are specifically depleted of H3K27me3 (Figure 3C), regions 3 (3q13.2-q13.31) and 35 (Xq28) are depleted in H3K9me2 (Figure 3D), and regions 11 (7q21.13-q21.2) and 25 (15q21.1) show a predominant gain in H3K9ac (Figure 3E).

Because tumor cell populations are thought to derive from progenitor populations of stem-like cells (Lawson and Witte, 2007), we next asked whether LREA regions also have a different expression and chromatin profile in hES cells. We found significant gene activation in LNCaP cells relative to gene expression in human embryonic stem (hES) cells in the LREA genes, similar to the differential expression observed in PrEC cells (Figure S6), suggesting that this domain activation is a cancer-specific phenomenon. Notably, when we compared the histone modification profiles, we found that over half of the LREA genes were bivalently marked in hES cells, that is, they harbored both H3K4me3 and H3K27me3 marks (Figure S6). Interestingly, gene activation in LNCaP cells for these genes was associated with loss of the H3K27me3 mark but retention of the H3K4me3 modification at promoters. The significant enrichment of the active H3K9ac modification observed in PrEC cells was less

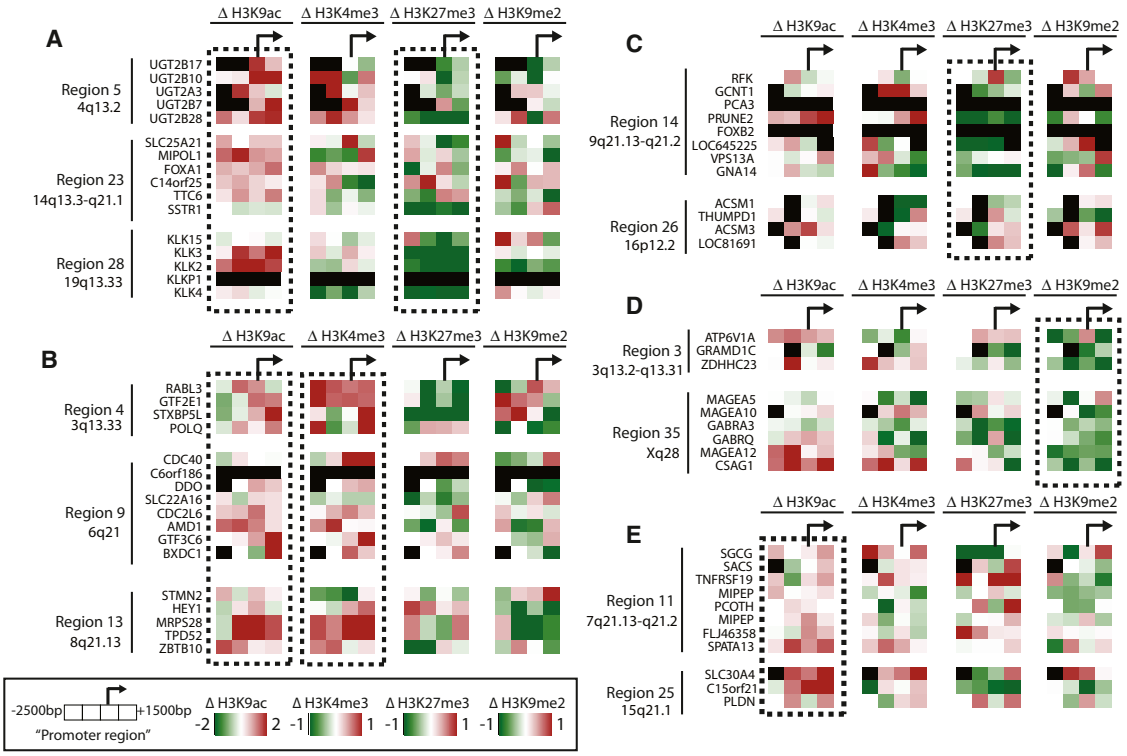

Figure 3. Epigenome Heat Maps of LREA Domains

Histone modification changes between LNCaP and PrEC cells are plotted across the TSS for each gene in an LREA domain, and various chromatin modes are identified.

(A) Gain of H3K9ac and loss H3K27me3.

(B) Gain in both H3K9ac and H3K4me3.

(C–E) Concordant change in only one histone mark (H3K27me3, H3K9me2, or H3K9ac). Each row represents a single gene (named). Heat maps are divided into four blocks showing changes in model-based analysis of tiling-array (MAT) scores at fixed intervals (−2,500 to −1,500, −1,500 to −500, −500 to +500, and +500 to +1,500 bp) relative to the TSS (arrow), for each modification. Relative scale is shown in the bottom panel (green, loss; red, gain; black, not represented on the array). Dotted boxes indicate regions with significant change between LNCaP and PrEC ($p < 0.1$).

See also Figure S4.

pronounced in comparison with hES cells, because many of these genes were already bivalently marked.

Cancer-Associated DNA Methylation Changes

To determine whether DNA methylation changes were common in promoter-associated CpG islands within LREA regions and activated genes across the genome, we collectively compared MBDCap-seq with changes in gene expression (Figure 4A). The majority of CpG islands within 2.5 kb of a TSS were lowly methylated in both PrEC and LNCaP cells, and therefore did not show any significant change in methylation. Of the CpG islands that did show a change in methylation, 2.5% were demethylated, whereas 21% of CpG islands were hypermethylated in LNCaP cells compared with PrECs (false discovery rate [FDR] < 0.05). A similar trend was also observed for CpG islands within LREA regions, with 1% and 23% of promoter-associated CpG islands losing or gaining methylation, respectively (FDR < 0.05; Figure 4A). Moreover, 5% of all hypermethylated CpG islands within 2.5 kb of a TSS were associated with transcripts that gained expression (t-statistic > 4; Figure 4A, boxed area), and 15% were associated with gene repression (t-statistic < −4).

As noted above, a CpG island (CpG_27) spanning the third and fourth exons of the *KLK4* gene was one of the few gene-associated CpG islands that were demethylated in LNCaP cells in association with gene activation (Figures 4A and 4B). However, the adjacent CpG island (CpG_22) 2 kb upstream of the *KLK4* TSS was conversely hypermethylated in LNCaP cells (Figures 4A and 4B). When we examined this relationship further, we noted that the majority of the *KLK4* transcripts in LNCaP cells originated from the second exon (Dong et al., 2005; Figure S5), located 666 bp upstream of the hypomethylated CpG island (CpG_27). The alternate "switching" of CpG island methylation (CpG_27 and CpG_22) in the cancer cells is of particular interest because these CpG islands flank the border between the two alternately transcribed Kallikrein LREA and LRES regions (Figure S3). Because the epigenetic status of these CpG islands could potentially regulate these domains, we investigated CpG methylation and chromatin (H3K27me3 and H3K4me3) data from H1 hES cells (Lister et al., 2009) to determine whether they are differently marked in early development. Figure 4B shows that for PrEC, CpG_22 is also unmethylated in hES cells and is marked bivalently with H3K27me3 and H3K4me3. In

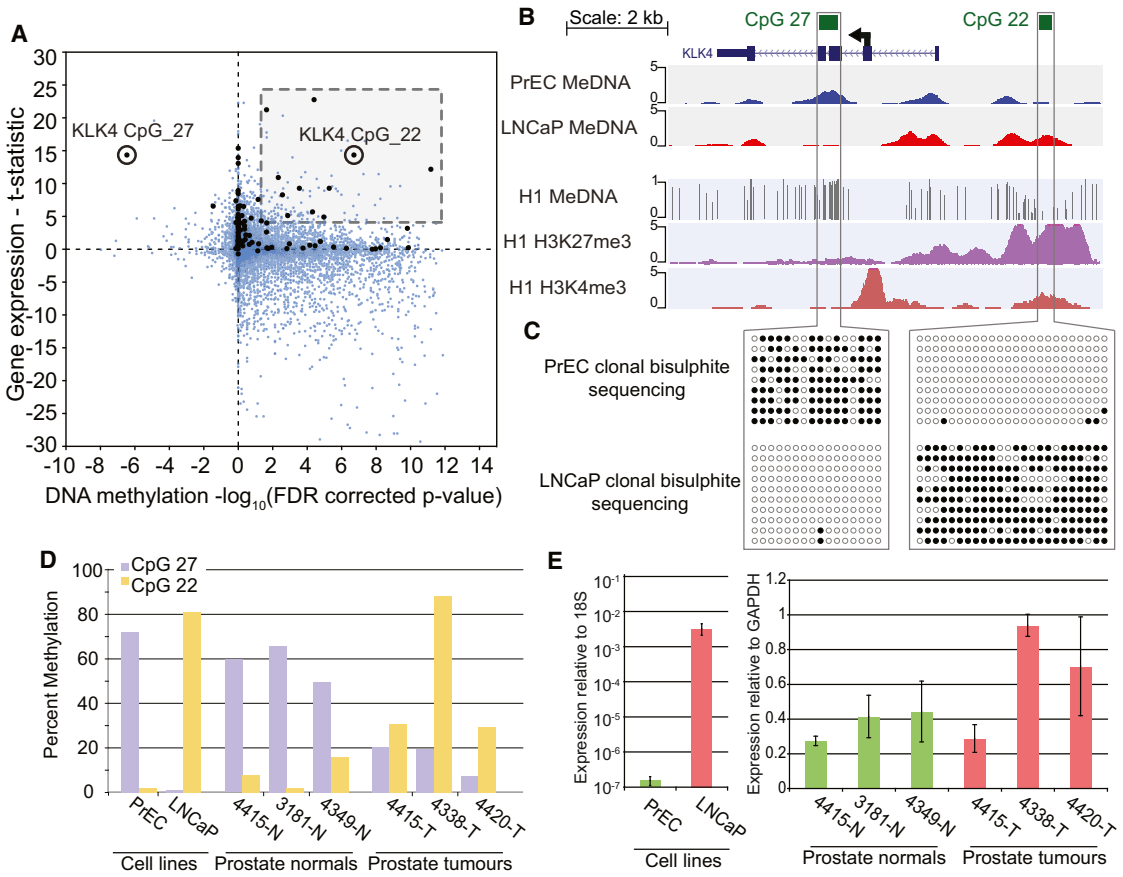

Figure 4. Relationship between DNA Methylation and Gene Expression

(A) Expression change from PrEC to LNCaP is plotted against changes in associated CpG island methylation. All dots represent CpG islands found within 2.5 kb of a TSS; black dots indicate CpG islands found in LREA regions, and blue dots denote all genomic CpG islands. The x axis is the $\pm\log_{10}$ of the FDR-adjusted p value representing differences in MBDCap-seq counts obtained ±500 bp from the center of each CpG island ($\pm\log_{10}$ for islands that lose or gain methylation, respectively). The y axis represents the t-statistic of change in expression between LNCaP and PrEC for genes associated with a given CpG island. The dashed box indicates genes that significantly change in expression (t-statistic > 4) with a significant increase in promoter CpG island methylation (FDR-adjusted p value < 0.05). The two *KLK4*-associated CpG islands are highlighted by black circles.

(B) *KLK4* locus, indicating the relative location of the two associated CpG islands. The arrow above exon 2 indicates the TSS in LNCaP. Data from MBDCap-seq are plotted in blue for PrEC and red for LNCaP. DNA methylation, H3K27me3, and H3K4me3 data for the *KLK4* locus from H1 ES cells (Lister et al., 2009) are represented below.

(C) Bisulphite clonal sequencing, confirming differential methylation in two CpG islands in PrEC and LNCaP.

(D) Percentage of methylation from clonal sequencing (CpG sites methylated / total CpG sites × 100) of the two *KLK4* CpG islands in PrEC, LNCaP, and six clinical samples (three normal [N] and three tumor [T]). See Figure S5C for full clonal sequence data.

(E) qRT-PCR of *KLK4* in PrEC and LNCaP (relative to the 18S); clinical samples (relative to the *GAPDH*). Error bars indicate ±SD.

See also Figure S5.

contrast, CpG_27 has neither of the bivalent marks and is hypermethylated in hES cells.

We confirmed the switch in CpG methylation in cancer and normal cells using bisulphite clonal sequencing and found that CpG_27 was extensively methylated in PrEC, whereas only 1% CpG sites were methylated in LNCaP (Figure 4C). Conversely, CpG_22 was essentially unmethylated (2% of sites were methylated) in PrEC and extensively methylated in LNCaP cells. Prostate tumors (n = 3) and normal prostate tissue (n = 3) isolated from cancer patients were also investigated for changes in methylation and expression associated with *KLK4* (Figures 4D and 4E). Clonal sequencing showed that all normal samples were extensively methylated at CpG_27 and were essentially unmethylated at CpG_22, whereas all tumor samples were hypomethylated at CpG_27 and hypermethylated at CpG_22 (Figure S5, summarized in Figure 4D). This methylation switch corresponded to *KLK4* overexpression in these samples (Figure 4E).

Hypermethylation of Promoter-Associated CpG Islands Associated with Gene Activation

The majority of the promoter-associated CpG islands that were hypermethylated and showed transcriptional activation (as depicted in Figure 4A, boxed area) could be divided into two main groups based on the cancer-specific methylation signature and the TSS, as determined by RNA-seq and cap analysis gene expression (CAGE)-seq (Figure 5). Group I (hypermethylation of

promoter-associated CpG island borders, but TSS remains unmethylated and TSS is unaltered), includes *PRUNE2* (CpG_75), *MMP16* (CpG_37), and *IQGAP2* (CpG_126; Figure 5A). Group II (extensive hypermethylation of CpG islands, including TSS, but there is change in the TSS), includes *TRIM36* (CpG_129), *ALOX15* (CpG_146), and *MPP2* (CpG_63; Figure 5B). For each activated transcript, we also found an increase in the H3K4me3 signal in the cancer cells at either the existing TSS for group I or at the new TSS for group II genes (Figure 5). Interestingly, for group I genes, we found that the cancer-specific hypermethylation across the CpG island borders was mutually exclusive to the H3K4me3 signal. Genome-wide, we found that the group I profile showing CpG island border methylation and gene activation was more common (88%) than the group II profile of hypermethylation and promoter switching (12%; Figure 5C). To determine whether hypermethylated borders potentially promote gene activation through binding inhibition of repressive factors, we interrogated the hypermethylated loci of group I regions for enrichment of transcription factor binding sites using the TRANSFAC database, and found significant enrichment of many transcriptional repressor-associated DNA elements (Table S2).

Recently, focal hypermethylation was also reported to occur in regions of long-range hypomethylation in colorectal cancer, and these regions showed some enrichment of silenced genes within LRES regions (Berman et al., 2012) and were associated with cancer-gene-silencing programs (Berman et al., 2012; Hon et al., 2012). However, we found no statistical enrichment (two-tailed chi-square test, $p > 0.05$) of LREA regions in these hypomethylated PMDs, supporting the conclusion that the type I and II hypermethylated CpG islands are uniquely associated with gene activation events independently of the long-range hypomethylation observed in colorectal and breast cancers (Berman et al., 2012; Hon et al., 2012).

We next asked whether these methylation changes were cancer specific. By comparing Illumina 450K data, we found that the changes in methylation observed in LNCaP cells at promoter CpG islands are comparable if measured against the PrEC or ES cell methylome ($R^2 = 0.804$; Figure 5D). Notably, both the group I and group II genes also showed a similar change in methylation in LNCaP cells when compared with PrEC and ES cells ($R^2 = 0.690$), indicating that the gain of aberrant methylation associated with activated transcription is cancer specific.

Epigenetic Changes Are Influenced by Changes in Adjacent Genes

Finally, to ascertain whether epigenetic changes are influenced by fluctuations in the local epigenomic environment, we quantified the incidence of epigenetic remodeling in adjacent genes in LNCaP as compared with PrEC cells over the whole genome. Figure S7 demonstrates how the frequency with which neighboring genes exhibit the same epigenetic change over the TSS compares with the "expected" random distribution. We found that at each position relative to the TSS, all of the epigenetic alterations (PrEC versus LNCaP) exhibited a significantly higher frequency of neighboring modification than would be expected by chance ($p < 0.01$). Increased H3K9ac and decreased H3K27me3 at neighboring loci were strongly significant over the entire promoter region (±1,000 bp from the TSS; $p < 1 \times 10^{-10}$), while decreases in DNA methylation were found to have the highest significance for enrichment of changes of any epigenetic mark ($p < 1 \times 10^{-43}$ over the whole promoter).

DISCUSSION

Transcriptional deregulation is common in cancer, and changes involve both gene repression and gene activation. Given that one of the main effectors of transcriptional deregulation is an alteration in the epigenetic landscape, it is surprising that most cancer genome-wide studies have focused on epigenetic repression in preference to epigenetic-driven gene activation. Primarily, this is due to the key role of DNA hypermethylation in contributing to gene silencing, including silencing of tumor suppressor genes, which is commonly promoted through deregulation of the Polycomb complex (Ohm et al., 2007; Schlesinger et al., 2007). In contrast, gene activation studies in cancer have mainly focused on promoter demethylation of individual genes and global demethylation of repeat regions (Kalari and Pfeifer, 2010; Ross et al., 2010). However, global demethylation is generally thought to contribute to carcinogenesis by promoting genome instability rather than oncogenic activation (Frigola et al., 2006). We previously reported that regional epigenetic repression, or LRES, commonly occurs in cancer and encompasses multiple genes that gain or exchange repressive histone marks and are typified by DNA hypermethylation of neighboring CpG islands (Coolen et al., 2010; Frigola et al., 2006). We now show that similar epigenetic processes that shaped the cancer epigenome into repressive domains can conversely create large, multigene domains of epigenetic accessibility and consequential transcriptional activation in cancer. The key findings of this study can be summarized as follows:

First, by integrating gene expression, chromatin, and DNA methylation genome-wide profiles in prostate cancer, we were able to identify 35 LREA domains that harbor 251 genes, including multiple gene families, and tumor-related genes. We also found loci encoding microRNAs (miRNAs) within the LREA regions, including let-7f, a member of the well-studied miRNA let-7 family, which exhibits tumor suppressor and antigrowth activity in prostate cancer (Liang et al., 2011), and miR-98, which has been reported to potentially target *EZH2* (Alajez et al., 2010). In addition, two prominent prostate cancer biomarkers, *KLK3* (*PSA*) and *PCA3*, were found embedded in LREA regions. Although *KLK3* can be expressed in normal prostate tissues, it is highly overexpressed in cancer (Shaw and Diamandis, 2007). *PCA3* is a noncoding messenger RNA that is extraordinarily prostate cancer specific (de Kok et al., 2002). Little is known about the transcriptional regulation of *PCA3*, but reports suggest that its regulation is independent of the overlapping *PRUNE2* gene (Salagierski et al., 2009). Here, we show that the surrounding *PRUNE2* locus becomes depleted of the repressive H3K27me3 modification in cancer cells, indicating that epigenetic remodeling may contribute to the biomarker's prostate-cancer-specific activation. Moreover, we noted that several genes and loci in these remodeled LREA regions are commonly involved in Ets-transcription-factor translocations (*KLK2*, *C15orf21*, and *MIPOL1*) in prostate cancer (Kumar-Sinha et al., 2008), suggesting that a more accessible chromatin structure may potentially favor genetic instability and consequently prime these genes for genomic rearrangement in prostate tumorigenesis.

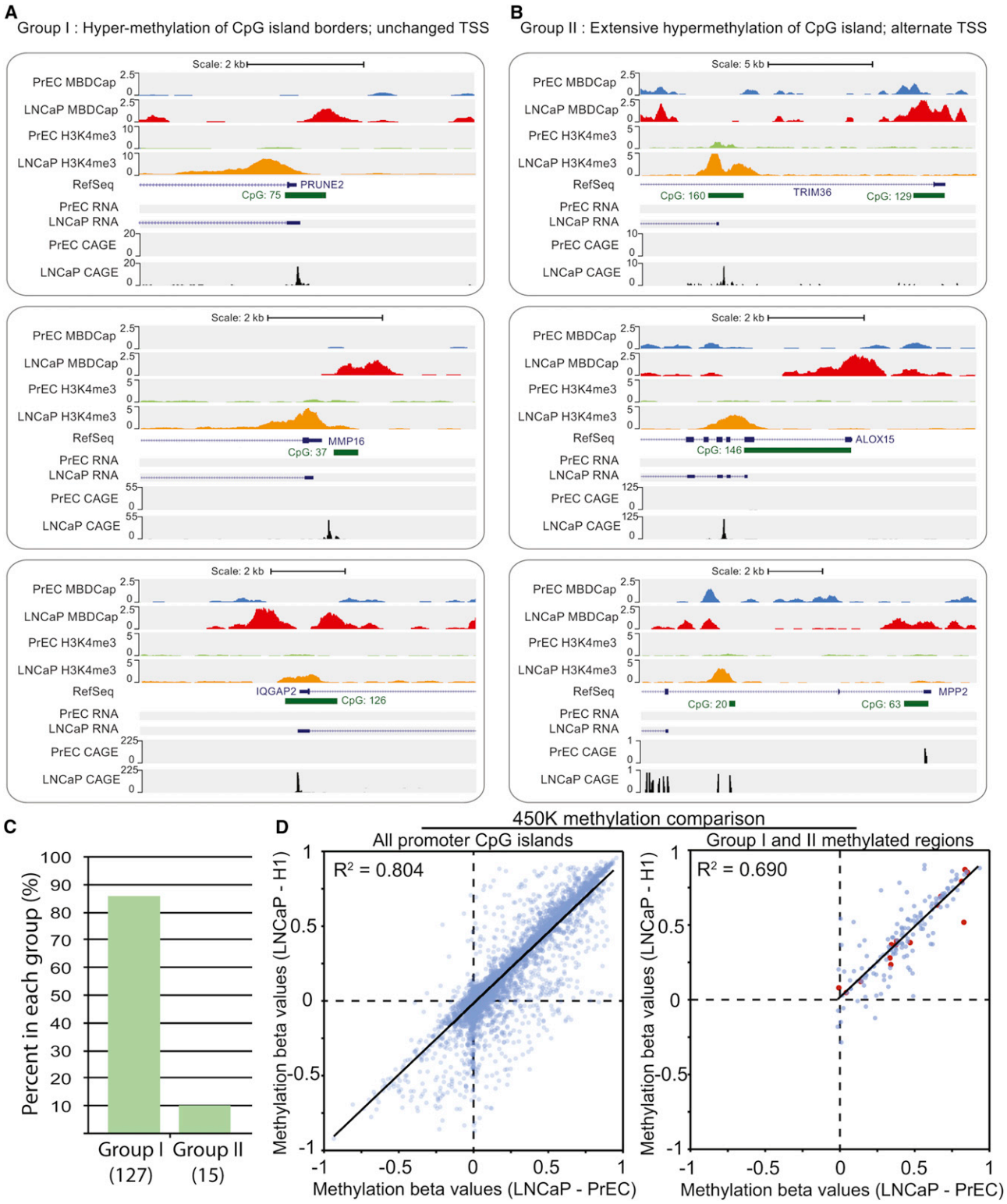

Figure 5. CpG Island Hypermethylation and Gene Activation

Two groups of CpG island methylation were found to be associated with activation of gene expression in LNCaP cells.

(A) Group I includes genes that showed a gain of DNA methylation at one or both of the flanking borders of a CpG island in association with gene activation.

(B) Group II genes showed extensive DNA methylation spanning the entire CpG island, associated with a change in the TSS. Data from MBDCap-seq are plotted in blue (PrEC) and red (LNCaP). Data from H3K4me3 ChIP-seq are plotted in green (PrEC) and orange (LNCaP). RNA-seq and CAGE-seq profiles are plotted below the RefSeq transcripts for each gene.

(legend continued on next page)

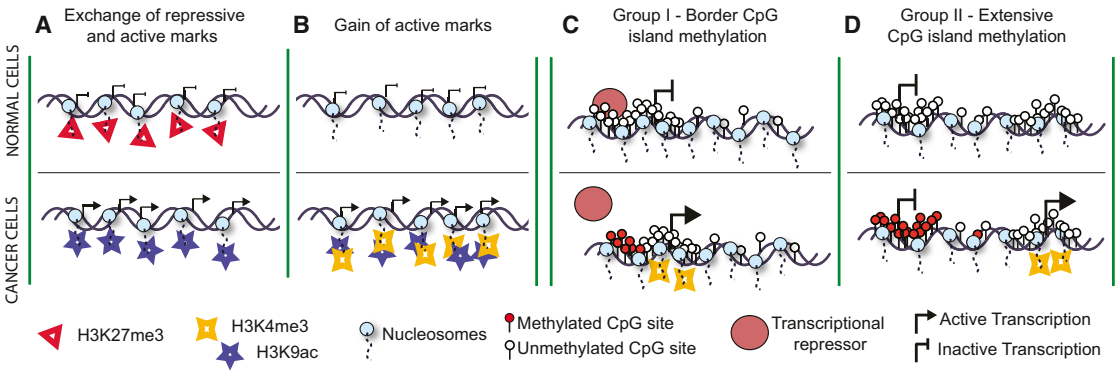

Figure 6. Summary of Epigenetic Changes that Occurred in LREA Domains

In relation to regional gene activation in cancer cells, two main modes of epigenetic change are found across promoter regions, in both histone modifications and DNA methylation patterns.

(A) Exchange of histone marks: active H3K9ac is gained and repressive H3K27me3 mark is depleted.

(B) Gain of active marks: both H3K9ac and H3K4me3 levels are elevated.

(C) Gain of border CpG island methylation and H3K4me3, which we propose may prevent the binding of a repressor element, thereby activating ectopic transcription.

(D) Extensive methylation at CpG islands results in a change of promoter usage and transcription from an alternative TSS.

See also Figure S7 and Table S2.

The second key finding of the study is that the LREA domains showed extensive changes in chromatin remodeling and could be divided into two prominent modes of histone modification alterations. Mode 1 (exchange of histone marks) is generally characterized by an enrichment of the active histone modification H3K9ac and depletion of the repressive histone modification H3K27me3 (Figure 6A). Mode 2 (gain of active histone marks) is characterized by an enrichment of both H3K9ac and H3K4me3 (Figure 6B). Notably, over half of the LREA genes were bivalently marked in hES cells, and gene activation in LNCaP cells was associated with loss of the H3K27me3 mark but retention of the H3K4me3 modification at promoters. These features are in contrast to LRES domains, which are notably marked by a depletion of the active epigenetic mark H3K9ac, an enrichment of the repressive H3K27me3 modifications. In fact, the entire Kallikrein locus presents a remarkable example of the prominent long-range switching that can occur in prostate cancer, where the LREA domain that gains H3K9ac and loses H3K27me3 modification is adjacent to an LRES domain that gains H3K27me3 and loses H3K9ac. It is interesting that the opposing transcriptional domains are bordered by two CTCF sites. CTCF is reported to be a boundary element/insulating protein that occurs at the bases of known chromatin looping interactions (Tsai et al., 2008), and is implicated in nuclear periphery positioning (Ottaviani et al., 2009). Even though we found that CTCF binding at the LREA/LRES boundary appears to be unaltered in the cancer cells, other deregulated CTCF cofactors may be involved in promoting switching between the two opposing epigenetic domains.

The third key finding of this study is that CpG island DNA hypermethylation is more commonly associated with gene activation in prostate cancer than DNA promoter hypomethylation, primarily because the majority of CpG islands are already unmethylated in normal prostate cells. However, we also found isolated examples of CpG island DNA demethylation, specifically at the *KLK4* locus, where concordant hypermethylation of an upstream CpG island and demethylation of a downstream CpG island within *KLK4*, resulted in transcriptional activation from a new TSS at the second exon. Interestingly, we found that the CpG island promoters associated with activated transcripts were more commonly hypermethylated than hypomethylated, which is an intriguing contradiction. Evidence suggests that methylated CpG sites can activate gene transcription at low-CpG-density promoters (Rishi et al., 2010). In this study, we show that DNA hypermethylation of CpG-rich promoter regions is associated with cancer gene activation, with two distinct profiles: group I methylation, where DNA methylation encroaches the flanking regions of the CpG islands, but specifically not the H3K4me3-marked TSS (Figure 6C); and group II methylation, which is characterized by extensive methylation across the CpG island, including the TSS (Figure 6D).

Group I hypermethylation is different from previously reported cancer-specific methylation at CpG island shores, which

(C) Summary of the percentages of group I and group II genes identified in LNCaP cells. Numbers in brackets indicate how many islands fall under a given characterization.

(D) Differences in methylation between LNCaP and H1 hES, and LNCaP and PrEC, are plotted for all promoter-associated CpG islands (left) and group I and II methylated regions (right). Group I and group II CpG islands are colored blue and red, respectively. Each point represents the average beta value derived from Infinium HumanMethylation450 BeadChip arrays (Statham et al., 2012), calculated ±500 bp from the center of each promoter-associated CpG island. Group I islands are represented as the average methylation over the specifically methylated region.

See also Figure S6.

primarily occurs up to 2 kb distant from the CpG islands in more CpG-depleted regions (Irizarry et al., 2009) and is associated with gene repression in cancer and tissue-specific expression (Doi et al., 2009) or a change in promoter usage (Irizarry et al., 2009). Hypermethylation has also been documented at gene body CpG islands, resulting in the silencing of intragenic transcription initiation sites (Maunakea et al., 2010). Gene repression is also associated with focal CpG island hypermethylation located in regions of long-range hypomethylation (Berman et al., 2012) or PMDs (Hon et al., 2012) and cancer-specific DNA-methylated regions (Hansen et al., 2011). In contrast, we found that group I hypermethylation at the borders of CpG islands results in an augmentation of gene transcription rather than gene repression. We propose that these flanking regions could harbor repressive factor binding elements, and that DNA hypermethylation relieves this repression through binding inhibition, subsequently promoting gene activation (Figure 6C). In fact, a survey of the methylated CpG island flanking sequences, using the TRANSFAC database, showed a highly significant enrichment in transcription factor binding sites in these regions (Table S2), many of which have repressor functions. One of the most common transcription factor binding sites found in the border regions was ZF5, which is a ubiquitous zinc finger transcriptional repressor that also binds to two sites in the c-*myc* promoter, modulating expression (Numoto et al., 1993). However, future experiments are required to determine whether the transcriptional repressors identified in this study are bona fide targets of the border CpG island group I genes. In contrast to group I genes, group II genes displayed extensive CpG island hypermethylation across the TSS, and this resulted in ectopic gene activation from alternative promoters in the cancer cell. In all cases, this was also associated with a gain of H3K4me3 across the new promoter regions. Our results clearly demonstrate that cancer-specific DNA hypermethylation of CpG islands contributes to deregulation of promoter usage, and can result not only in gene repression but, notably, also in cancer-associated gene activation.

Finally, evidence suggesting regional genomic deregulation is further reinforced by our pairwise analysis of chromatin change, which showed that neighboring promoters undergo similar epigenetic remodeling in cancer for all chromatin marks studied, including DNA methylation (Figure S7). Importantly, neighboring remodeling was found to be significant for both enrichment and depletion of all of the studied epigenetic modifications. The causes of this regional bidirectional epigenetic remodeling are still unclear, but it is possible that factors that typically organize the genome into epigenetically and transcriptionally appropriate domains themselves become deregulated. Currently, several organizational processes have been described that could be causally related in the establishment of LREA and LRES in cancer. It will be of considerable interest to investigate the role of the spatial organization of the genome, such as radial positioning within the nucleus (Singh Sandhu et al., 2011; Strasák et al., 2009), association with the nuclear lamina (Berman et al., 2012), and chromatin looping structures (Hsu et al., 2010), in the deregulation of long-range epigenetic control in cancer, as these processes are all instrumental in establishing epigenetic domains in differentiation and disease.

EXPERIMENTAL PROCEDURES

ChIP Assays

ChIP assays were carried out according to the manufacturer's protocol (Upstate Biotechnology) as previously described (Coolen et al., 2010). The antibodies used were specific for acetylated-histone H3 (H3K9ac; catalog no. 06-599, Millipore), dimethyl-histone H3 (H3K9me2; catalog no. ab1220, Abcam), tri-methyl-histone H3 (H3K27me3; catalog no. 07-449, Millipore) and tri-methyl-histone H3 (H3K4me3; catalog no. ab8580, Abcam). More details regarding the protocol can be found in Supplemental Experimental Procedures.

Methylation Profiling by MBDCap-seq

The MethylMiner Methylated DNA Enrichment Kit (Invitrogen) was used to isolate methylated DNA from LNCaP and PrEC cell lines as previously described (Nair et al., 2011). We used 10 ng of captured DNA for library preparation, and Illumina GAII sequencing to generate a 36 bp read length. We mapped the 36 bp reads to the hg18 reference genome using Bowtie (Langmead et al., 2009), with up to three mismatches. More details regarding the protocol can be found in Supplemental Experimental Procedures.

Chromatin Mark Heat Map Analysis

Using the blocksStats procedure in the Repitools software package (Coolen et al., 2010), we summarized the enrichment in 1,000 bp blocks (−2,500 to −1,500, −1,500 to −500, −500 to +500, and +500 to +1,500) for each epigenetic mark surrounding each LREA-associated TSS. We tested changes in epigenetic marks for individual regions using geneSetTest from the R limma package (Smyth, 2004).

Identification of Regions Showing Long-Range Epigenetic Activation

Array data for two replicates each of LNCaP and PrEC RNA were preprocessed using RMA (Irizarry et al., 2003) as implemented in the R package aroma.affymetrix (Bengtsson et al., 2008). Moderated t-statistics representing differential expression between LNCaP and PrEC cells were calculated using limma (Smyth, 2004) for each gene on the array. We identified domains of LREA by first defining the core region as the region in which the median t-statistic over five genes was >4 for two sequential genes. This core was then extended bidirectionally to encompass flanking genes that were also assigned positive t-statistics for change in expression. We used a five-gene sliding window as an arbitrary definition because this size allows the identification of sizable activated domains. A detailed protocol can be found in the findClusters procedure in the Repitools R package (Statham et al., 2010).

RNA-seq, CAGE-seq, and Chip-seq Mapping

For RNA-seq, 75 bp single-end reads were mapped to the human genome (build hg18) using TopHat (Trapnell et al., 2010), with RefSeq used as a reference transcriptome. Putative transcripts were assembled using Cufflinks (Langmead et al., 2009) with the default parameters. CAGE libraries were made by DNAFORM (Japan), and sequencing was performed by GeneWorks (Australia). We mapped 29 bp reads derived from CAGE tags to the human genome (hg18) using bowtie (Langmead et al., 2009), allowing up to three mismatches. Reads that mapped to multiple places with equal numbers of mismatches were discarded. For Chip-seq, we mapped 50 bp reads derived from H3K4me3 and CTCF ChIP to the human genome (hg18) using bowtie (Langmead et al., 2009), allowing up to three mismatches. Nonunique reads and reads that mapped more than once (i.e., identical start sites) to a single genomic location were excluded.

Infinium HumanMethylation450 BeadChip Arrays

HumanMethylation450 BeadChip array data for PrEC and LNCaP cell lines were adapted from Statham et al. (2012). Arrays were hybridized and data were processed as previously described (Statham et al., 2012). For CpG island analysis, beta values representing methylation at individual probes were averaged ±500 bp from the center of all CpG islands located within 2.5 kb of a TSS.

Tumor Samples

Fresh-frozen clinical tissue samples were obtained with informed consent from St. Vincent's Campus Prostate Cancer Group (i.e., from men with localized prostate cancer treated by radical prostatectomy), with appropriate ethical approval from St. Vincent's Campus Research Ethics Committee (Approval No. H00/088). RNA and DNA were extracted using the TRIzol reagent (Invitrogen).

ACCESSION NUMBERS

All data sets have been deposited in NCBI's Gene Expression Omnibus (GEO). Raw and analyzed tiling and expression arrays for LNCaP and PrEC can be found under GSE24546; CAGE-seq, RNA-seq, CTCF-seq, H3K4me3-seq, and DU145 and PC3 expression array profiling data sets can be found under GSE38685.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.ccr.2012.11.006.

ACKNOWLEDGMENTS

We thank Dr. Kate Patterson for help with the figures and the Ramacotti Centre, University of New South Wales (Sydney, Australia), for array hybridization, ChIP-seq, RNA-seq, and MBDCap-seq. H1 hES DNA was kindly supplied by Phillippe Collas. This work was supported by CINSW Fellowships (to M.W.C. and T.H.), CINSW Student Scholarships (to S.A.B. and A.L.S.), and National Health and Medical Research Council (NHMRC) project and fellowship support (to S.J.C. and R.L.S.). S.J.C., C.S., and M.W.C. conceived and designed the experiments; J.Z.S. and S.A.B. performed the experiments; M.D.R., D.S., and A.L.S. analyzed the data; R.L.S. and T.H. provided clinical samples; and S.A.B. and S.J.C. wrote the paper. This work is dedicated to the memory of Robert L. Sutherland (October 10, 2012).

Received: January 8, 2012
Revised: May 24, 2012
Accepted: November 15, 2012
Published: December 13, 2012

REFERENCES

Alajez, N.M., Shi, W., Hui, A.B., Bruce, J., Lenarduzzi, M., Ito, E., Yue, S., O'Sullivan, B., and Liu, F.F. (2010). Enhancer of Zeste homolog 2 (EZH2) is overexpressed in recurrent nasopharyngeal carcinoma and is regulated by miR-26a, miR-101, and miR-98. Cell Death Dis. *1*, e85.

Bengtsson, H., Simpson, K., Bullard, J., and Hansen, K. (2008). aroma.affymetrix: a generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory. Tech Report #745, Department of Statistics (Berkeley: University of California).

Berman, B.P., Weisenberger, D.J., Aman, J.F., Hinoue, T., Ramjan, Z., Liu, Y., Noushmehr, H., Lange, C.P., van Dijk, C.M., Tollenaar, R.A., et al. (2012). Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. Nat. Genet. *44*, 40–46.

Cedar, H., and Bergman, Y. (2009). Linking DNA methylation and histone modification: patterns and paradigms. Nat. Rev. Genet. *10*, 295–304.

Chalitchagorn, K., Shuangshoti, S., Hourpai, N., Kongruttanachok, N., Tangkijvanich, P., Thong-ngam, D., Voravud, N., Sriuranpong, V., and Mutirangura, A. (2004). Distinctive pattern of LINE-1 methylation level in normal tissues and the association with carcinogenesis. Oncogene *23*, 8841–8846.

Ciampi, R., Knauf, J.A., Kerler, R., Gandhi, M., Zhu, Z.W., Nikiforova, M.N., Rabes, H.M., Fagin, J.A., and Nikiforov, Y.E. (2005). Oncogenic AKAP9-BRAF fusion is a novel mechanism of MAPK pathway activation in thyroid cancer. J. Clin. Invest. *115*, 94–101.

Clark, S.J., and Melki, J. (2002). DNA methylation and gene silencing in cancer: which is the guilty party? Oncogene *21*, 5380–5387.

Coolen, M.W., Stirzaker, C., Song, J.Z., Statham, A.L., Kassir, Z., Moreno, C.S., Young, A.N., Varma, V., Speed, T.P., Cowley, M., et al. (2010). Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. Nat. Cell Biol. *12*, 235–246.

de Kok, J.B., Verhaegh, G.W., Roelofs, R.W., Hessels, D., Kiemeney, L.A., Aalders, T.W., Swinkels, D.W., and Schalken, J.A. (2002). DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. Cancer Res. *62*, 2695–2698.

de Leeuw, B., Balemans, M., Olde Weghuis, D., and Geurts van Kessel, A. (1995). Identification of two alternative fusion genes, SYT-SSX1 and SYT-SSX2, in t(X;18)(p11.2;q11.2)-positive synovial sarcomas. Hum. Mol. Genet. *4*, 1097–1099.

Deras, I.L., Aubin, S.M., Blase, A., Day, J.R., Koo, S., Partin, A.W., Ellis, W.J., Marks, L.S., Fradet, Y., Rittenhouse, H., and Groskopf, J. (2008). PCA3: a molecular urine assay for predicting prostate biopsy outcome. J. Urol. *179*, 1587–1592.

Doi, A., Park, I.H., Wen, B., Murakami, P., Aryee, M.J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., et al. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat. Genet. *41*, 1350–1353.

Dong, Y., Bui, L.T., Odorico, D.M., Tan, O.L., Myers, S.A., Samaratunga, H., Gardiner, R.A., and Clements, J.A. (2005). Compartmentalized expression of kallikrein 4 (KLK4/hK4) isoforms in prostate cancer: nuclear, cytoplasmic and secreted forms. Endocr. Relat. Cancer *12*, 875–889.

Ehrlich, M. (2002). DNA methylation in cancer: too much, but also too little. Oncogene *21*, 5400–5413.

Feinberg, A.P., Gehrke, C.W., Kuo, K.C., and Ehrlich, M. (1988). Reduced genomic 5-methylcytosine content in human colonic neoplasia. Cancer Res. *48*, 1159–1161.

Frigola, J., Song, J., Stirzaker, C., Hinshelwood, R.A., Peinado, M.A., and Clark, S.J. (2006). Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. Nat. Genet. *38*, 540–549.

Grunau, C., Sanchez, C., Ehrlich, M., van der Bruggen, P., Hindermann, W., Rodriguez, C., Krieger, S., Dubeau, L., Fiala, E., and De Sario, A. (2005). Frequent DNA hypomethylation of human juxtacentromeric BAGE loci in cancer. Genes Chromosomes Cancer *43*, 11–24.

Hansen, K.D., Timp, W., Bravo, H.C., Sabunciyan, S., Langmead, B., McDonald, O.G., Wen, B., Wu, H., Liu, Y., Diep, D., et al. (2011). Increased methylation variation in epigenetic domains across cancer types. Nat. Genet. *43*, 768–775.

Hermans, K.G., Bressers, A.A., van der Korput, H.A., Dits, N.F., Jenster, G., and Trapman, J. (2008). Two unique novel prostate-specific and androgen-regulated fusion partners of ETV4 in prostate cancer. Cancer Res. *68*, 3094–3098.

Hon, G.C., Hawkins, R.D., Caballero, O.L., Lo, C., Lister, R., Pelizzola, M., Valsesia, A., Ye, Z., Kuan, S., Edsall, L.E., et al. (2012). Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res. *22*, 246–258.

Hsu, P.Y., Hsu, H.K., Singer, G.A., Yan, P.S., Rodriguez, B.A., Liu, J.C., Weng, Y.I., Deatherage, D.E., Chen, Z., Pereira, J.S., et al. (2010). Estrogen-mediated epigenetic repression of large chromosomal regions through DNA looping. Genome Res. *20*, 733–744.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. *31*, e15.

Irizarry, R.A., Ladd-Acosta, C., Wen, B., Wu, Z., Montano, C., Onyango, P., Cui, H., Gabo, K., Rongione, M., Webster, M., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat. Genet. *41*, 178–186.

Jones, P.A., and Baylin, S.B. (2007). The epigenomics of cancer. Cell *128*, 683–692.

Kalari, S., and Pfeifer, G.P. (2010). Identification of driver and passenger DNA methylation in cancer by epigenomic analysis. Adv. Genet. *70*, 277–308.

Kumar-Sinha, C., Tomlins, S.A., and Chinnaiyan, A.M. (2008). Recurrent gene fusions in prostate cancer. Nat. Rev. Cancer *8*, 497–511.

Lamprecht, B., Walter, K., Kreher, S., Kumar, R., Hummel, M., Lenze, D., Kochert, K., Bouhlel, M.A., Richter, J., Soler, E., et al. (2010). Derepression of an endogenous long terminal repeat activates the CSF1R proto-oncogene in human lymphoma. Nat. Med. *16*, 571–579.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Lawson, D.A., and Witte, O.N. (2007). Stem cells in prostate cancer initiation and progression. J. Clin. Invest. *117*, 2044–2050.

Liang, S., He, L., Zhao, X., Miao, Y., Gu, Y., Guo, C., Xue, Z., Dou, W., Hu, F., Wu, K., et al. (2011). MicroRNA let-7f inhibits tumor invasion and metastasis by targeting MYH9 in human gastric cancer. PLoS ONE *6*, e18409.

Lilja, H., Ulmert, D., and Vickers, A.J. (2008). Prostate-specific antigen and prostate cancer: prediction, detection and monitoring. Nat. Rev. Cancer *8*, 268–278.

Lim, J.H., Kim, S.P., Gabrielson, E., Park, Y.B., Park, J.W., and Kwon, T.K. (2005). Activation of human cancer/testis antigen gene, XAGE-1, in tumor cells is correlated with CpG island hypomethylation. Int. J. Cancer *116*, 200–206.

Lim, S.L., Smith, P., Syed, N., Coens, C., Wong, H., van der Burg, M., Szlosarek, P., Crook, T., and Green, J.A. (2008). Promoter hypermethylation of FANCF and outcome in advanced ovarian cancer. Br. J. Cancer *98*, 1452–1456.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Maunakea, A.K., Nagarajan, R.P., Bilenky, M., Ballinger, T.J., D'Souza, C., Fouse, S.D., Johnson, B.E., Hong, C., Nielsen, C., Zhao, Y., et al. (2010). Conserved role of intragenic DNA methylation in regulating alternative promoters. Nature *466*, 253–257.

Mullighan, C.G., Su, X., Zhang, J., Radtke, I., Phillips, L.A., Miller, C.B., Ma, J., Liu, W., Cheng, C., Schulman, B.A., et al.; Children's Oncology Group. (2009). Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N. Engl. J. Med. *360*, 470–480.

Nair, S.S., Coolen, M.W., Stirzaker, C., Song, J.Z., Statham, A.L., Strbenac, D., Robinson, M.D., and Clark, S.J. (2011). Comparison of methyl-DNA immunoprecipitation (MeDIP) and methyl-CpG binding domain (MBD) protein capture for genome-wide DNA methylation analysis reveal CpG sequence coverage bias. Epigenetics *6*, 34–44.

Nishigaki, M., Aoyagi, K., Danjoh, I., Fukaya, M., Yanagihara, K., Sakamoto, H., Yoshida, T., and Sasaki, H. (2005). Discovery of aberrant expression of R-RAS by cancer-linked DNA hypomethylation in gastric cancer using microarrays. Cancer Res. *65*, 2115–2124.

Numoto, M., Niwa, O., Kaplan, J., Wong, K.K., Merrell, K., Kamiya, K., Yanagihara, K., and Calame, K. (1993). Transcriptional repressor ZF5 identifies a new conserved domain in zinc finger proteins. Nucleic Acids Res. *21*, 3767–3775.

Ohm, J.E., McGarvey, K.M., Yu, X., Cheng, L., Schuebel, K.E., Cope, L., Mohammad, H.P., Chen, W., Daniel, V.C., Yu, W., et al. (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. Nat. Genet. *39*, 237–242.

Ottaviani, A., Schluth-Bolard, C., Rival-Gervier, S., Boussouar, A., Rondier, D., Foerster, A.M., Morere, J., Bauwens, S., Gazzo, S., Callet-Bauchu, E., et al. (2009). Identification of a perinuclear positioning element in human subtelomeres that requires A-type lamins and CTCF. EMBO J. *28*, 2428–2436.

Rishi, V., Bhattacharya, P., Chatterjee, R., Rozenberg, J., Zhao, J., Glass, K., Fitzgerald, P., and Vinson, C. (2010). CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. Proc. Natl. Acad. Sci. USA *107*, 20311–20316.

Robinson, M.D., Stirzaker, C., Statham, A.L., Coolen, M.W., Song, J.Z., Nair, S.S., Strbenac, D., Speed, T.P., and Clark, S.J. (2010). Evaluation of affinity-based genome-wide DNA methylation data: effects of CpG density, amplification bias, and copy number variation. Genome Res. *20*, 1719–1729.

Ross, J.P., Rand, K.N., and Molloy, P.L. (2010). Hypomethylation of repeated DNA sequences in cancer. Epigenomics *2*, 245–269.

Salagierski, M., Verhaegh, G.W., Jannink, S.A., Smit, F.P., Hessels, D., and Schalken, J.A. (2009). Differential expression of PCA3 and its overlapping PRUNE2 transcript in prostate cancer. Prostate *70*, 70–78.

Salomon-Nguyen, F., Della-Valle, V., Mauchauffe, M., Busson-Le Coniat, M., Ghysdael, J., Berger, R., and Bernard, O.A. (2000). The t(1;12)(q21;p13) translocation of human acute myeloblastic leukemia results in a TEL-ARNT fusion. Proc. Natl. Acad. Sci. USA *97*, 6757–6762.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben-Shushan, E., Reubinoff, B.E., et al. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. Nat. Genet. *39*, 232–236.

Shaw, J.L., and Diamandis, E.P. (2007). Distribution of 15 human kallikreins in tissues and biological fluids. Clin. Chem. *53*, 1423–1432.

Singh Sandhu, K., Li, G., Sung, W.K., and Ruan, Y. (2011). Chromatin interaction networks and higher order architectures of eukaryotic genomes. J. Cell. Biochem. *112*, 2218–2221.

Smyth, G.K. (2004). Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. *3*, Article3.

Statham, A.L., Strbenac, D., Coolen, M.W., Stirzaker, C., Clark, S.J., and Robinson, M.D. (2010). Repitools: an R package for the analysis of enrichment-based epigenomic data. Bioinformatics *26*, 1662–1663.

Statham, A.L., Robinson, M.D., Song, J.Z., Coolen, M.W., Stirzaker, C., and Clark, S.J. (2012). Bisulfite sequencing of chromatin immunoprecipitated DNA (BisChIP-seq) directly informs methylation status of histone-modified DNA. Genome Res. *22*, 1120–1127.

Strasák, L., Bártová, E., Harnicarová, A., Galiová, G., Krejcí, J., and Kozubek, S. (2009). H3K9 acetylation and radial chromatin positioning. J. Cell. Physiol. *220*, 91–101.

Tomlins, S.A., Laxman, B., Dhanasekaran, S.M., Helgeson, B.E., Cao, X., Morris, D.S., Menon, A., Jing, X., Cao, Q., Han, B., et al. (2007). Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. Nature *448*, 595–599.

Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. *28*, 511–515.

Tsai, C.L., Rowntree, R.K., Cohen, D.E., and Lee, J.T. (2008). Higher order chromatin structure at the X-inactivation center via looping DNA. Dev. Biol. *319*, 416–425.

Wolff, E.M., Byun, H.M., Han, H.F., Sharma, S., Nichols, P.W., Siegmund, K.D., Yang, A.S., Jones, P.A., and Liang, G. (2010). Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. PLoS Genet. *6*, e1000917.