Agglomerative Epigenetic Aberrations Are a Common Event in Human Breast Cancer

Petr Novak,[1,4] Taylor Jensen,[1,2] Marc M. Oshiro,[1] George S. Watts,[1] Christina J. Kim,[1,3] and Bernard W. Futscher[1,2]

[1]Arizona Cancer Center, Departments of [2]Pharmacology and Toxicology and [3]Surgery, University of Arizona, Tucson, Arizona; and [4]Institute of Plant Molecular Biology AS CR, Ceske Budejovice, Czech Republic

Abstract

Changes in DNA methylation patterns are a common characteristic of cancer cells. Recent studies suggest that DNA methylation affects not only discrete genes, but it can also affect large chromosomal regions, potentially leading to LRES. It is unclear whether such long-range epigenetic events are relatively rare or frequent occurrences in cancer. Here, we use a high-resolution promoter tiling array approach to analyze DNA methylation in breast cancer specimens and normal breast tissue to address this question. We identified 3,506 cancer-specific differentially methylated regions (DMR) in human breast cancer with 2,033 being hypermethylation events and 1,473 hypomethylation events. Most of these DMRs are recurrent in breast cancer; 90% of the identified DMRs occurred in at least 33% of the samples. Interestingly, we found a nonrandom spatial distribution of aberrantly methylated regions across the genome that showed a tendency to concentrate in relatively small genomic regions. Such agglomerates of hypermethylated and hypomethylated DMRs spanned up to several hundred kilobases and were frequently found at gene family clusters. The hypermethylation events usually occurred in the proximity of the transcription start site in CpG island promoters, whereas hypomethylation events were frequently found in regions of segmental duplication. One example of a newly discovered agglomerate of hypermethylated DMRs associated with gene silencing in breast cancer that we examined in greater detail involved the protocadherin gene family clusters on chromosome 5 (*PCDHA, PCDHB,* and *PCDHG*). Taken together, our results suggest that agglomerative epigenetic aberrations are frequent events in human breast cancer. [Cancer Res 2008;68(20):8616–25]

Introduction

Aberrant levels and patterns of cytosine methylation are ubiquitous events in human cancer (1). DNA hypomethylation generally occurs often in repetitive elements, pericentromeric regions, and within the body of genes (2–4). In contrast, DNA hypermethylation is largely limited to CpG island regions (5–7). Some gene targets of DNA hypermethylation are found in a broad spectrum of different tumor types, whereas others occur in a tumor type–specific fashion (8). Studies into the biological consequences of hypomethylation and hypermethylation suggest that each plays a participative role in carcinogenesis. Earlier studies suggested that hypermethylation events were focal in nature, usually associated with the transcriptional silencing of one gene. Recent studies show that DNA hypermethylation can extend over regions of the genome from 100 to >1,000 kb in size and is linked to long-range epigenetic silencing (LRES; refs. 9–11).

These LRES events have practical parallels to genetic lesions, such as deletions and microdeletions, in that in both cases long contiguous stretches of DNA have become functionally inactivated. With respect to the epigenetic events, the long-range silencing is associated with an epigenetic terrain consisting of landmarks linked to a repressed chromatin state, such as aberrant DNA methylation and/or histone modification state (9–12). Considering the gross genetic and epigenetic changes that occur in a cancer cell, it is likely that diverse inappropriate combinations of epigenetic landmarks will partner together to produce aberrant gene silencing. This LRES is metastable and reversible with drugs that target epigenetic control mechanisms (9, 10).

We set out to better determine the scope and the spatial organization of aberrant DNA methylation and its relationship to LRES in human breast cancer. Using epigenomic scanning approaches, we analyzed a set of normal and cancerous breast specimens, a limited life span human mammary epithelial cell (HMEC) strain, and breast cancer cell lines to identify differentially methylated regions (DMR) between normal and malignant breast tissue. Locations of DMRs revealed that aberrant DNA methylation is not randomly distributed across the cancer genome, and agglomerates of DMRs can be identified, consisting of agglomerates of both hypermethylated and hypomethylated DMRs that ranged from ~30 kbp to almost 600 kbp in size. These aberrant agglomerative methylation events can occur multiple times within an individual tumor and are associated with LRES events. The types of promoters and the regions affected by DNA hypomethylation were distinct from those targeted by hypermethylation; however, a feature common to both types was their apparent concentration in gene family clusters.

Materials and Methods

Cell culture. All cell lines were maintained as previously described (9). The MDA-MB-231 and BT549 breast cancer cells were obtained from the American Type Culture Collection. Cell line identity was assured by genotypic analysis by DNA fingerprinting using single tandem repeats (STRs). Normal HMECs were obtained from Clonetics and were grown according to supplier's instructions.

Breast tumor specimens. Flash-frozen specimens derived from normal or cancerous breast tissue were obtained from patients who underwent surgery for breast cancer, either lumpectomy or mastectomy, at the University Medical Center in Tucson, AZ, from 2003 to 2005. All patients signed surgical and clinical research consents for tissue collection in

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Bernard W. Futscher, Arizona Cancer Center, 1515 North Campbell Avenue, Tucson, AZ 85724. Phone: 520-626-4646; Fax: 520-626-4979; E-mail: bfutscher@azcc.arizona.edu.

doi:10.1158/0008-5472.CAN-08-1419

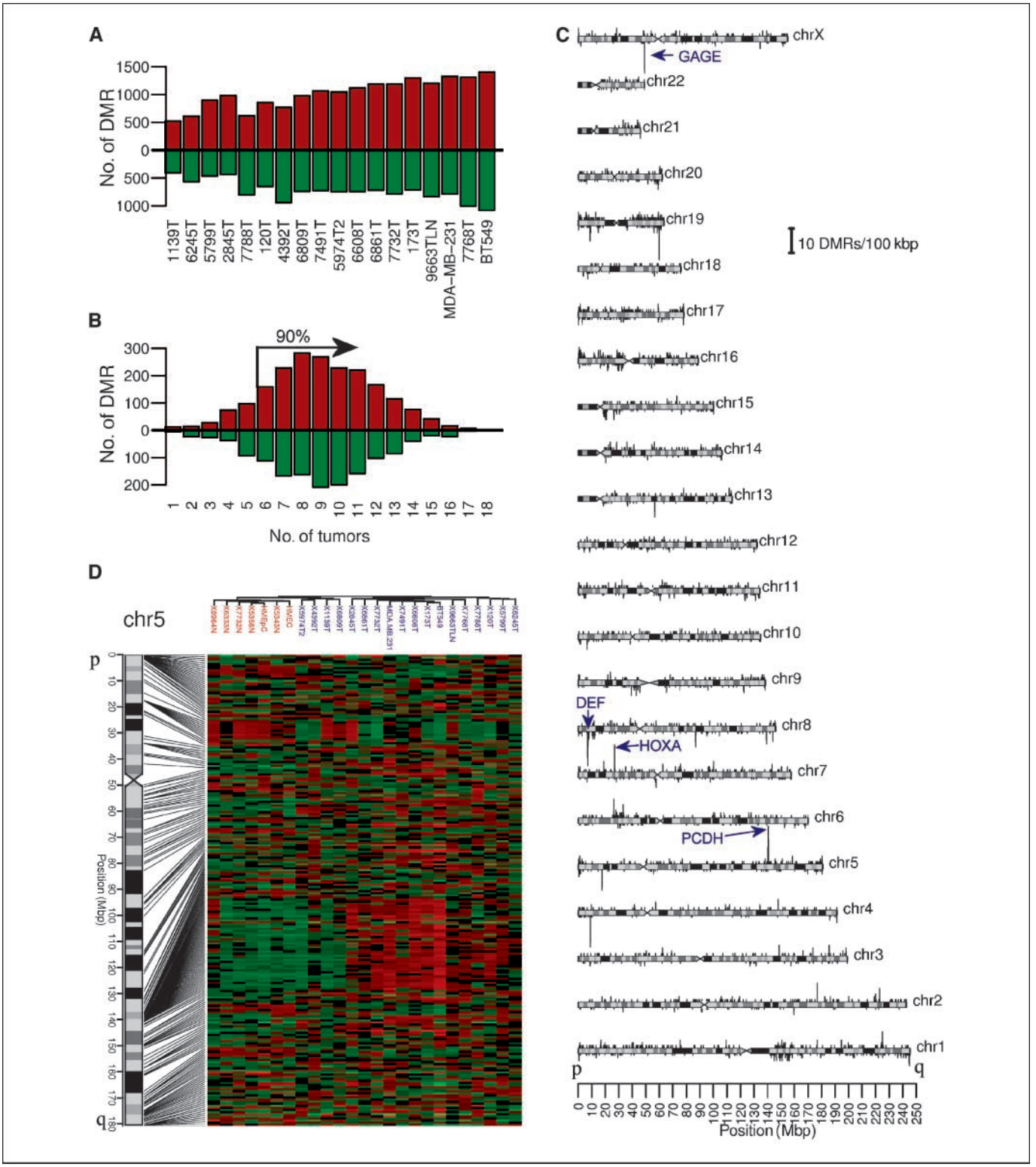
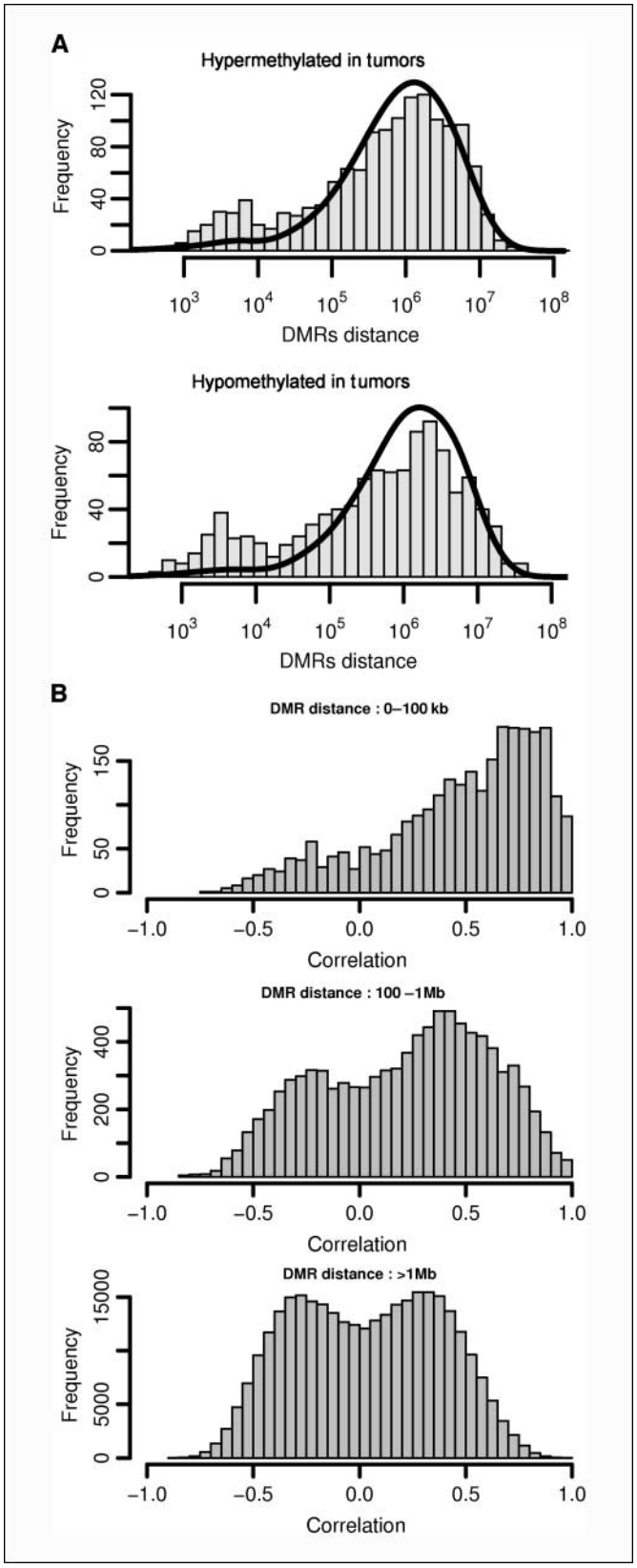


Figure 1. Analysis of the spatial distribution of breast tumor–specific DMRs across the genome. *A,* the number of aberrantly methylated DMRs found in individual tumors. *Red and green bars,* number of sites that were hypermethylated or hypomethylated, respectively. *B,* most of individual DMRs are recurrent and can be found in multiple tumors. For each DMR, the number of tumors in which a specific DMR is aberrantly methylated was counted (*X* axis), and total amount of DMRs in each frequency group is shown on *Y* axis. *Arrow,* delimits 90% of all DMRs. *Red and green bars,* the number of sites that were hypermethylated or hypomethylated, respectively. *C,* distribution of 2,033 regions hypermethylated in breast tumor specimens and 1,473 hypomethylated regions; along chromosomes is shown. The DMR frequency is plotted along the chromosome. Positive peaks correspond to hypermethylated regions, negative peaks to hypomethylated regions. The locations of the HOXA, GAGE, DEF, and PCDH DMRs are labeled. Window size, 100 kb. *D,* a heat map that shows chromosome 5 DMRs ordered by their physical position. *Red and green,* regions of hypermethylation and hypomethylation, respectively. Samples were sorted using hierarchical clustering; *blue,* cancer specimens; *red,* normal specimens. Note the high density of DMR around position 140 Mbp corresponding to PCDH cluster. Heatmaps for other chromosomes can be found as Supplementary Fig. S1.

accordance with the University of Arizona Institutional Review Board and Health Insurance Portability and Accountability Act of 1996 regulations. At the time of surgery, a 1- to 3-cm section of the tumor was immediately snap frozen in liquid nitrogen and stored in our prospective breast tissue bank at −80°C. From each tissue block, a series of 5-μm sections were cut and stained with H&E for pathologic evaluation. All H&E slides were reviewed by two independent pathologists to determine the integrity of the tumor specimen. A partial molecular characterization of these samples have been reported on previously (9). Supplementary Table S1 provides the pathologic assessment of each specimen.



Nucleic acid isolation. RNA and DNA were isolated as previously described (9).

Bisulfite sequencing and DNA sequence analysis. Genomic DNA was analyzed by bisulfite sequencing as described (13). The protocadherin CpG islands were amplified from the bisulfite-modified DNA by two rounds of PCR using nested primers. Primer sequences are available upon request.

DNA methylation analysis by MassARRAY. Sodium bisulfite (NaBS)-treated genomic DNA was prepared according to manufacturer's instructions (Zymo Research). NaBS-treated DNA (5ng) was seeded into a region specific PCR reaction incorporating a T7 RNA polymerase sequence as described by the manufacturer (Sequenom). Resultant PCR product was then subjected to *in vitro* transcription and RNase A cleavage using the MassCLEAVE T-only kit, spotted onto a Spectro CHIP array, and analyzed using the MassARRAY Compact System MALDI-TOF mass spectrometer (Sequenom). Each NaBS-treated DNA sample was processed in two independent experiments. Data were analyzed using EpiTyper software (Sequenom) and the R-script "Analyze Sequenom Function" using methods described previously (14). Primer sequences were designed using EpiDesigner.[5] Primer sequences are available upon request.

Real-time reverse transcription-PCR. Total RNA was converted to cDNA and then amplified in a real-time PCR format using gene-specific primers and fluorescent probes using Roche UniversalProbe technology and the ABI 7500 Real-time Detection System (Applied Biosystems). Results were calculated using the Delta Ct method normalizing to glyceraldehyde-3-phosphate dehydrogenase expression. Primer sequences are available upon request.

Methyl DNA immunoprecipitation microarrays and data analysis. Methylated fraction of DNA was obtained by immunoprecipitation as described (15). Two hundred nanograms of enriched fraction and corresponding input DNA were hybridized on Affymetrix GeneChip Human Promoter 1.0R Array according to manufacturer's protocol.

Tiling Analysis Software (Affymetrix) was used to extract signal intensities and map them to corresponding genomic position based on National Center for Biotechnology Information human genome assembly (Build 34). Subsequent data processing and data analysis was done using R programming environment (16). Signal intensities were loess normalized, then input intensities were subtracted from corresponding methyl DNA immunoprecipitation (MeDIP) intensities. Resulting MeDIP/Input ratios were again loess normalized. For each probe, a quality weight was calculated as previously described (17). To find DMRs in normal and cancerous samples, t statistics was calculated for each probe. To combine information from neighboring probes, a weighted moving average (window, 1,000 bp) was computed as final summary statistics for each probe. The same procedure was done with probe positions randomly

[5] http://www.epidesigner.com

Figure 2. Cancer-specific DMRs are nonrandomly distributed across the genome. *A,* nonrandom distribution of hypermethylation and hypomethylation DMRs in breast tumors. The histogram shows the distribution of distances between neighboring DMRs. *Solid line,* the theoretical distances between neighboring DMRs based on simulated data sets of random spatial distributions. The increased frequency of DMRs closer than 10^5 bp suggests that tumor-specific DMRs are nonrandomly distributed across the genome. *B,* correlation between pairs of DMRs is dependent on their physical distance. Pearson correlation coefficients were calculated for each pair of DMRs based on microarray data. Correlation coefficients were then divided into three groups according to the distance between DMRs in each DMR pair. Note the strong positive correlation coefficients among proximate DMRs that decreases to no correlation as more distant DMRs are considered. The positive correlation coefficients among nearby DMRs that are closer than 100 kb suggests that aberrant DNA methylation events in these proximate regions are not independent processes.

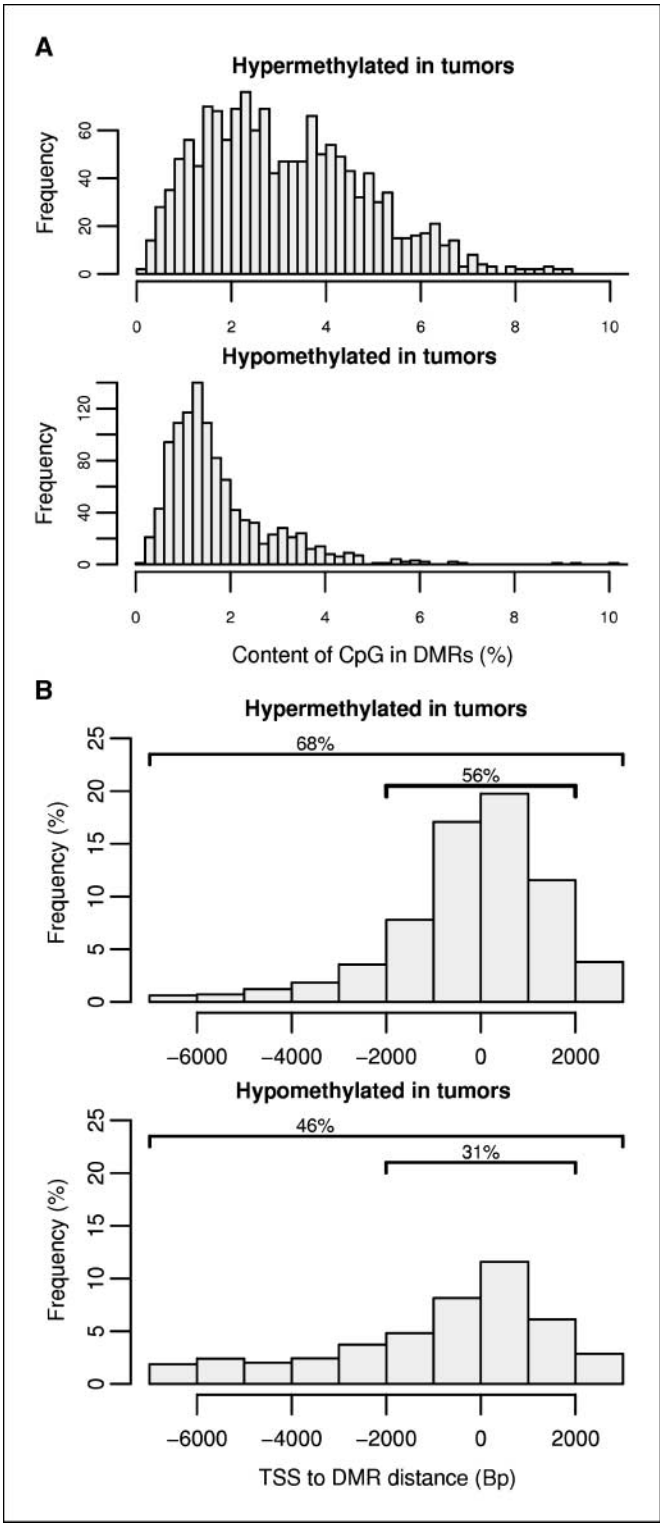


Figure 3. Characteristics of DMRs. *A,* hypermethylated and hypomethylated tumor–specific DMRs differ in CpG content. Regions targeted for hypermethylation in tumor specimens have a higher CpG content than regions of hypomethylation, with 38% of the hypermethylated regions overlapping with known CpG islands, whereas only 6% of hypomethylated sequences overlapping with CpG islands. *B,* localization of tumor-specific DMRs relative to transcription start sites (TSS). Positions of the detected DMRs were compared with the positions of TSS taken from the UCSC genome browser (SwitchGear TSS Track). Overall frequency does not reach 100% because not all DMRs detected on the promoter microarray mapped to the known TSS; 32% of hypermethylated and 54% hypomethylated sites were out of the range that was analyzed (−7 to +3 kb).

Table 1. Hypermethylated and hypomethylated DMRs occurrences in single copy and duplicate copy genomic sequences

	Hypermethylation	Hypomethylation
DMRs total	2,033	1,473
Single copy DMRs	1,835 (90.3%)	1,016 (69%)
Duplicated DMRs	198 (9.7%)	457 (31%)
Average copy number of duplicated DMRs	4.5	8.8

permutated to find the appropriate cutoff for false discovery rate of 5%. The positions of all positive regions can be found in Supplementary Data as an annotation track that can be visualized in University of California at Santa Cruz (UCSC) Genome Browser. Multiple occurrences of DMR sequences were analyzed using BLAT alignment tool (18). The raw data from these microarray studies have been deposited in the Array Express database, accession E-MEXP-1481.[6]

The expected (theoretical) distances of randomly distributed DMRs were calculated by taking the number of DMRs detected in the sample set and randomly distributing this number through the regions covered by the promoter array. Distances between DMRs from these simulations were then calculated; the average results of 50 simulations are shown. The random distribution simulation was performed separately for hypermethylated and hypomethylated DMRs.

Chromatin immunoprecipitations. Chromatin immunoprecipitations were performed as previously described (19). The antibodies directed toward histone H3 lysine-9 dimethylation (DiMeK9) and H3 lysine-27 trimethylation (TriMeK27) were purchased from Upstate/Millipore. Primer sequences are available upon request.

Results

In this study, we sought to examine the nature and spatial organization of aberrant DNA methylation in invasive human breast cancers and corresponding cell line models. The DNA methylation-state of 16 invasive breast tumors, 5 histologically normal breast specimens, and 4 pure cell line/strain populations derived from cancerous and normal mammary tissue was analyzed using the Affymetrix Human gene promoter array 1.0R, a more comprehensive approach that provided greater and more focused genomic coverage than our previous studies (9, 20). Labeled targets from each of the samples for microarray hybridizations were generated from the 5-methylcytosine–specific antibody enriched fraction of the DNA, as well as the total the genomic DNA (Input). Input DNA was used as a reference control ensuring that each sample analyzed for DNA methylation was compared with its own unique genome. This reference approach served to minimize false discovery due to other genomic aberrations, such as gene amplification.

We identified 3,506 cancer-specific DMRs, which were either aberrantly hypermethylated (2,033) or hypomethylated (1,473) in cancerous tissue. A DMR was defined as any region covered by a group of consecutive probes on the tiling array that exhibited statistically significant differences in normalized signal (i.e., MeDIP/Input ratio) between the group of normal and the group of cancerous specimens. A vast majority (99%) of these DMRs covered 8 to 115 tiling array probes and therefore ranged in size from 280 to 4,000 bp; 1,000 bp was the most frequent DMR size. The methylation status of each of these DMRs was analyzed in each individual cancer

[6] http://www.ebi.ac.uk/arrayexpress

specimen relative to the normal specimens, which enabled us to determine the number of aberrantly methylated DMRs in each individual tumor as well as the frequency of aberrant methylation for each DMR within the sample population as whole. This analysis revealed that each cancerous specimen exhibits hypomethylation and hypermethylation in ~25% to 75% of the 3,506 DMRs identified (Fig. 1*A*). Finally, most of the DMRs that we detected are recurrent in the breast cancer. Approximately 90% of these DMRs are aberrantly methylated in at least 33% of our sample set (Fig. 1*B*). Overall, these results indicate that there are thousands of recurrent DMRs in a substantial fraction of human breast cancers.

To identify potential long-range events, we analyzed the spatial distribution of the DMRs across all 23 chromosomes (Fig. 1*C*). Positive peaks indicate hypermethylated regions, whereas negative peaks indicate hypomethylated regions. The amplitude of each peak is proportional to the number of DMRs found within a 100-kb

Table 2. Genomic regions of agglomerative epigenetic aberrations in breast cancer

Chromosome	From (bp)	To (bp)	Size (kbp)	No of DMRs	Frequency* (% of patients)	Genes within region
Hypermethylated in tumors						
chr1	16567504	16838375	271	8	33	*BC006312, AK123337*
chr2	176771543	176880741	109	13	44	*HOXD13, HOXD11, HOXD10, HOXD9, AK092134, HOXD8, HOXD4, HOXD3*
chr5	140145947	140287745	142	11	50	*AK091409, PCDHA1, PCDHA10, PCDHA11, PCDHA12, PCDHA13, PCDHA2, PCDHA3, PCDHA4, PCDHA5, PCDHA6, PCDHA7, PCDHA8, PCDHA9, PCDHAC1*
chr5	140455538	140838940	383	32	67	*BC031837, BC051788, PCDHB10, PCDHB11, PCDHB12, PCDHB13, PCDHB14, PCDHB15, PCDHB16, PCDHB17, PCDHB18, PCDHB2, PCDHB3, PCDHB4, PCDHB5, PCDHB6, PCDHB7, PCDHB8, PCDHGA1, PCDHGA10, PCDHGA11, PCDHGA12, PCDHGA2, PCDHGA3, PCDHGA4, PCDHGA5, PCDHGA6, PCDHGA7, PCDHGA8, PCDHGA9, PCDHGB1, PCDHGB2, PCDHGB3, PCDHGB4, PCDHGB5, PCDHGB6, PCDHGB7, PCDHGC3, SLC25A2, TAF7*
chr6	26290824	26381754	91	8	28	*HIST1H2BE, HIST1H4D, HIST1H3D, HIST1H2AD, BC056264, HIST1H2BF, HIST1H4E, BC082232, HIST1H2BG, HIST1H2AE, HIST1H3E, HIST1H1D, HIST1H4F, HIST1H4G, HIST1H3F, HIST1H2BH, HIST1H3G, HIST1H2BI*
chr6	33273119	33532803	260	9	22	*RXRB, X66424, SLC39A7, HSD17B8, RING1, VPS52, BC040114, AK093057, AK001725, RPS18, AY536376, B3GALT4, C6orf11, HKE2, RGL2, AB012295, BC064966, BC080574, TAPBP, ZNF297, DAXX, AF090423, KIFC1, PHF1, C6orf82, SYNGAP1, AB067525, ZBTB9*
chr7	26909767	27055875	146	13	44	*HOXA3, AK056230, HOXA4, HOXA5, HOXA6, HOXA7, HOXA9, HOXA10, BC007600, HOXA11, HOXA13*
Hypomethylated in tumors						
chr1	244610973	245171804	561	23	39	*OR2M2, OR2M3, OR2M4, OR2T33, OR2T12, OR2M7, OR5BF1, OR2T4, OR2T1, OR2T2, OR2T3, OR2T5, OR2G6, OR2T29, OR2T34, OR2T10, OR2T11, OR2T35*
chr4	9019357	9047147	28	12	89	*DUB3, AY533200*
chr5	17551068	17642039	91	9	6	no genes
chr8	6826023	6934557	109	12	22	*DEFA1, DEFA3, DEFA5*
chr8	7061201	7315981	255	18	56	*DEFB4, DEFB103A, SPAG11, DEFB104A*
chr8	7729174	7935902	207	8	39	*DEFB4, DEFB103A, SPAG11, DEFB104A*
chr19	59862651	60076000	213	26	39	*KIR2DL1, KIR2DL3, KIR2DL4, KIR2DS2, KIR2DS4, KIR3DL1, KIR3DL2, KIR3DL3, KIR3DP1, L76664, L76668, LILRB4, U33328, X93596, X99481, AF002256, AF276292, AF283988, AY102623, AY366253, AY601812, BC028206, BC069344, CR609786*
chrX	48930476	49163928	233	18	28	*GAGE1, GAGE2, GAGE4, GAGE5, GAGE6, BC036094, BC081536*

NOTE: DMRs were grouped together when the distance between them was smaller than 100 kb. Regions with clusters of ≥8 consecutive DMRs are shown.
*Frequency represents the percentage of tumors in which aberrant methylation occurs in at least 75% of all DMRs within the analyzed region.

window, and therefore, the largest peaks in Fig. 1*C* reveal agglomerates of DMRs—regions with the highest density of aberrant DNA methylation. The agglomerates containing hypermethylated DMRs then represent regions with the potential for DNA methylation-associated LRES.

Examination of the DNA methylation chromosomal plot in Fig. 1*C* shows that we could detect multiple regions of agglomerative DNA methylation. One such region is the *HOXA* gene family cluster, which we previously identified in breast cancer (9), thereby providing empirical support for the analytic approach used to detect aberrant DNA methylation related to LRES. One novel region of agglomerative DNA hypermethylation discovered is located on chromosome 5 and is associated with the three protocadherin gene clusters (*PCDHA, PCDHB, and PCDHG*). A heat map of the DNA methylation state of all DMRs ordered by physical position along chromosome 5 is shown in Fig. 1*D*. The patient samples were ordered by hierarchical clustering and the heat map allows visualization of the hypermethylated DMRs associated with the *PCDH* gene clusters. We also detected agglomerative hypomethylation events specific to cancer; one of which was localized to the GAGE gene family cluster on the X chromosome. The GAGE genes are tumor-specific antigens not expressed in adult somatic tissues but undergo a transcriptional activation during carcinogenesis that has been associated with DNA hypomethylation (21, 22).

We evaluated whether the agglomeration of DMRs in apparent hotspots was simply due to chance or if DMRs clustered in a cancer-driven fashion. One way in which this was assessed was through an analysis of the distributions of genomic distances between neighboring hypermethylated and hypomethylated DMRs. Figure 2*A* shows the frequency of distances between DMRs in the data obtained (histogram) compared with the frequency of theoretical distances based on a simulated random spatial distribution of DMRs (*overlaid solid line*). Overall, the results were quite similar for both hypermethylated and hypomethylated DMRs. In both cases, the observed and expected (theoretical) distribution shows the majority of DMRs occur $\sim 10^6$ bp apart; however, when the observed data are further compared with the expected data, there is an increased frequency of hypermethylated as well as hypomethylated DMRs that have a spacing of $<10^4$ bp. Thus, the divergence of the observed DMR distributions from the expected DMR distributions suggests that aberrant DNA methylation is nonrandomly distributed across the breast cancer genome.

We explored the distribution of DMRs further and found that the location of individual DMRs are positively correlated with methylation changes of proximate DMR sites (<100 kb apart). This effect is practically null for distant sites (>1 Mb; Fig. 2*B*). In other words, there is a higher probability that proximate DMR sites will exhibit the same aberrant methylation status in a particular sample. Examples of such correlated regions can be seen on Fig. 1*D* around positions 20 and 140 Mbp. This phenomenon suggests that there is a common factor affecting methylation status of such aberrant methylation prone regions, and provides further support for the idea that cancer-specific aberrant methylation is distributed across the genome in a nonrandom fashion.

Further examination of the genomic characteristics of the promoter-associated DMRs revealed other features that distinguished hypermethylation from hypomethylation events in the invasive breast cancers. Results in Fig. 3*A* show that the aberrant hypermethylation events in breast cancer were linked to CpG-rich regions, with 38% of hypermethylated DMRs coincident with CpG islands, whereas only 6% of hypomethylation DMRs were linked to CpG islands. The locations of these DMRs within the 10-kb promoter regions assayed on the gene promoter array were also determined (Fig. 3*B*). DMRs associated with aberrant hypermethylation were concentrated in the region near transcription start, whereas hypomethylated DMRs were more evenly distributed throughout the promoter region.

The types of promoter regions targeted by DNA hypermethylation and hypomethylation differed further. Cancer-specific DNA hypomethylation events occurred at a higher frequency in regions of segmental duplications, whereas DNA hypermethylation events were heavily biased toward single copy sequences (Table 1). Taken together, these results suggest that both long-range DNA hypermethylation and hypomethylation events occur in human breast cancer, and that the occurrence of these events show regional selectivity both within promoter regions and throughout the genome.

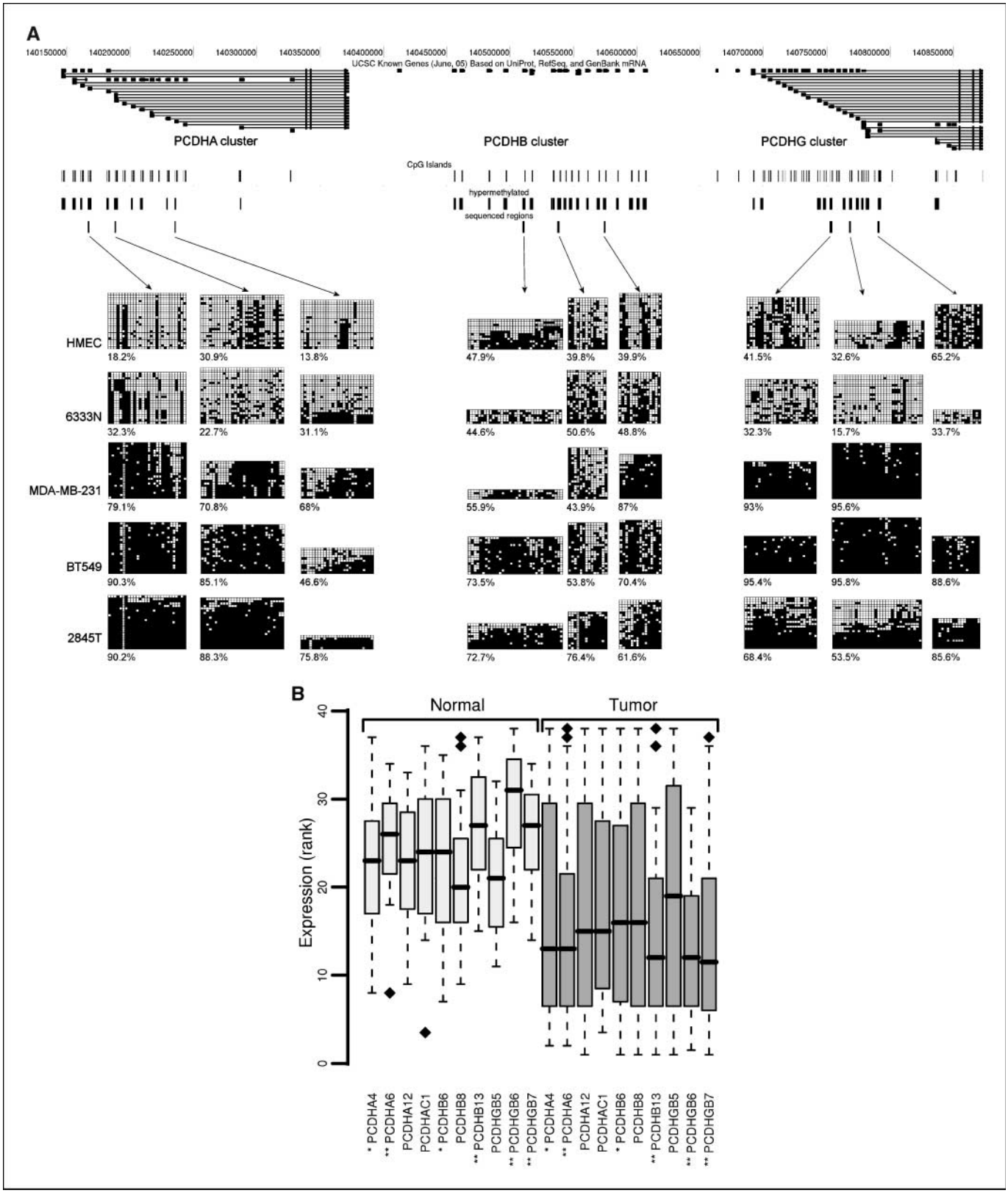
Table 2 shows the chromosomal regions associated with the highest density of DMRs within the breast cancer samples when compared with normal tissue. The *HOXA* genes are present and representative of hypermethylation events, whereas *GAGE* genes are present and representative of hypomethylation events. Table 2 also shows a range of 22% to 89% for the frequency of occurrence of the 15 agglomerative DNA methylation events detected in the breast cancer samples. In general, these results show that DNA agglomerative events occur at frequencies that are comparable with and exceed the frequency of common genetic mutations in breast cancer, such as p53 (23). Interestingly, there was a prevalence of gene family clusters in the list of agglomerates of DMRs including the protocadherin gene family clusters (*PCDHA, PCDHB, and PCDHG*), the *HOXA* gene cluster, the *HOXD* gene cluster, and a histone gene cluster.

The *PCDH* gene family clusters showed a strong hypermethylation signal in the DNA methylation chromosomal plot (Fig. 1*C*). Considering their unique genomic organization and the potential for these clusters to be a novel target for epigenetic silencing in human breast cancer, we analyzed this region in greater detail. This family of *PCDH* genes (α,β, and γ) resides in 3 clusters on chromosome 5, and the genes are expressed in normal human breast tissue and down-regulated in breast cancer (Fig. 4*B*). The α and γ gene clusters share similar characteristics. The genes within each cluster share the same last three 3′ exons and differ only in their unique 5′ exons. In contrast, each gene in the β cluster is distinct, ordered along the chromosome, and consists of only one exon (24). Each *PCDH* gene in these three gene family clusters also has a CpG island promoter.

To confirm and extend the *PCDH* differential methylation results detected on the gene promoter microarrays, the methylation states of nine 5′ CpG island promoters (three each from the *PCDHA, PCDHB, and PCDHG* clusters) were analyzed by bisulfite sequencing in a selected set of samples. The genomic organization of the CpG island regions analyzed as well as the results of the bisulfite sequencing are shown in Fig. 4*A*. The bisulfite sequencing analysis of normal mammary tissue and the nontumorigenic, mortal cell strain HMEC revealed surprisingly high and widespread methylation of the *PCDH* CpG island promoters. The overall levels of methylation between the normal tissue and the HMEC cell strain were comparable and, based on the reverse transcription-PCR (RT-PCR) data, not at a level sufficient to repress transcription. The results further show that DNA hypermethylation of *PCDH* CpG island promoters occurs in human breast cancer. Importantly, the breast cancer sample 2845T, which shows LRES of the *PCDH* gene

cluster, also contains an LRES event at the *HOXA* cluster (9). Thus, this tumor is a clear example showing the occurrence of multiple LRES events within a single cancer genome. Overall, the results from this high-resolution analysis confirm that aberrant DNA hypermethylation of multiple *PCDH* CpG islands is a common event in human breast cancer.

In addition to these agglomerative DNA methylation events, analysis of the microarray results also allowed for the detection of



focal events previously determined to be aberrantly methylated in human breast cancer specimens. For example, similar to earlier reports we detected aberrant DNA methylation in the promoters of *p16, WT-1, RASSF1A, DSC3, SIM1, PCDH10,* and *FOXA2* (25–29). Furthermore, it is worthwhile to note that some focal events previously associated with breast tumorigenesis are contained within agglomerative DNA methylation events such as *HOXA5, HOXA9, HOXD11,* and *PCDHGB6* (12, 28, 30). Taken together, these results indicate that the human promoter microarrays are able to detect common focal methylation changes as well as agglomerates of aberrantly methylated DMRs in human breast cancer specimens even using relatively small sample populations.

The prevalence of hypomethylated DMRs in the breast cancer samples detected by differential methylation microarray led us to validate these results. We analyzed 4 hypomethylated DMRs that are part of agglomerative epigenetic aberration located within a segmental duplication on chromosome 8. This region was also chosen because it was recently found to be hypomethylated in lung cancer (31). Quantitative analysis of DNA methylation patterns was determined using the MassArray pipeline of bisulfite PCR, *in vitro* transcription, base-specific cleavage of the RNA products, and mass spectrometry (32). Results from the analysis of three noncancer and three cancer samples are presented in Supplementary Fig. S2 and confirm the microarray data. Overall, these results show that both agglomerative DNA hypomethylation and hypermethylation events occur in breast cancer and represent a facet of long range epigenetic aberrations linked to human carcinogenesis.

Discussion

Using the Affymetrix Human gene promoter array 1.0R, we took a genome-wide approach to profile the DNA methylation state of gene promoter regions in normal and cancerous human breast specimens and cell line models. We extended earlier observations that showed LRES in cancer and now find multiple agglomerative DNA hypermethylation events and, in the case of the PCDH gene family cluster, we confirmed that this event led to LRES. We also found agglomerative events characterized by DNA hypomethylation. The regions of hypermethylation and hypomethylation were distinct from one another. DNA hypermethylation events concentrated in CpG islands near transcription start sites, whereas hypomethylation events were found typically in CpG-poor regions and were not concentrated near transcription start, but rather the hypomethylated DMRs were more evenly spread across the 10-kb promoter regions analyzed by the array. Interestingly, the hypomethylation events also were disproportionately found in regions of segmental duplication, and it seems that a region we identified on chromosome 8 as hypomethylated in breast cancer also maps to a region of hypomethylation recently described in lung cancer (31).

A feature common to both the long-range DNA hypermethylation and hypomethylation events was their frequent occurrence in gene family clusters. The functional significance of the agglomerative DNA hypermethylation events, as opposed to the hypomethylation events, is more evident and is clearly linked to gene repression. One gene family cluster we identified as a new target of aberrant hypermethylation in breast cancer was the protocadherin cluster (*PCDHA, PCDHB,* and *PCDHG*) on chromosome 5. Further analysis showed that an overall decrease in the expression of the gene family cluster in breast cancer was correlated with the increased methylation of the *PCDH* cluster's CpG islands.

One mechanism that may be involved in the aberrant hypermethylation of entire gene family clusters is by disrupting the function of transcription factors and gene regulators that participate in the control of their expression. With respect to the protocadherin cluster, a mechanism involving this possibility exists. Recent studies indicate that binding sites for the insulator protein, CTCF, preferentially concentrate to the *PCDH* cluster (33, 34). Because CTCF DNA binding activity is reported to be attenuated in breast cancer (35–37) and loss of its DNA binding activity has been linked to the spreading of aberrant DNA methylation into regions once protected by CTCF (38–40), the aberrant DNA methylation of the *PCDH* gene cluster may be a result of compromised CTCF function. An alternative possibility is that aberrant DNA methylation appears first in the *PCDH* regions, and the increased levels of DNA methylation in CTCF recognition sites directly block CTCF interaction with the DNA because CTCF binding is inhibited by DNA methylation and therefore its functional role in regulating gene transcription (33, 34, 41, 42).

In another scenario, aberrant methylation may occur because of an altered chromatin structure that is secondary to changes in the protein composition of the chromatin interactions. We have previously observed DNA methylation–linked LRES in the *HOXA* gene cluster (9). This situation represents a potential case where changes in DNA-protein interactions may play a role in changes in DNA methylation. The histone methyltransferases EZH2 and MLL1 play important and opposing roles in *HOXA* gene regulation and their dysregulation has been linked to cancer etiology (43–46). Considering their role in the maintenance of epigenetic control, their disruption may allow for or participate in the appearance of regional aberrant methylation. Another factor that may ultimately participate in the LRES seen in the *HOXA* cluster in breast cancer is retinoid receptors because retinoids are important participants in *HOXA* transcriptional control and important regulators of breast carcinogenesis (47, 48).

Additional mechanisms may participate in the establishment of the agglomerative DNA hypermethylation and likely act in concert with one another to produce stable LRES. It has been proposed

Figure 4. Aberrant DNA methylation and transcriptional repression of the PCDH cluster. *A,* map of PCDH cluster taken from UCSC Genome Browser (http://genome.ucsc.edu). Individual tracks, from the top to the bottom, show positions of CpG islands, the regions detected as hypermethylated by microarray, and then the regions analyzed by bisulfite sequencing. Below these tracks are the bisulfite sequencing results showing the methylation status of selected DMRs across the PCDH cluster in breast specimens and cell lines. The methylation data obtained by bisulfite sequencing of the samples analyzed are shown. Sample with a suffix "N" is derived from normal tissue and sample with a suffix "T" is derived from tumor tissue. The cell lines shown are the normal cell strain HMEC and the breast cancer cell lines MDA-MB-231 and BT549. In these diagrams, each row represents an individual cloned and sequenced PCR product, whereas the columns contain the data for each of the CpG sites analyzed (■, methylated sites; □, unmethylated sites; ■, poor sequence data). Clones were sorted from least to most methylated for presentation purposes. Results from the three regions analyzed by bisulfite sequencing for each of the gene clusters, PCDHA, PCDHB, and PCDHG are shown. The numbers below each bisulfite sequence plots are the percentage of methylated CpG sites for that sample in the region analyzed. *B,* gene expression of the PCDH cluster in normal and breast tumor specimens. mRNA levels of PCDH genes that were bisulfite sequenced were measured using real-time RT-PCR. The expression was analyzed in 15 normal and 23 cancerous specimens. For visualization purposes, expression values were converted to ranks. Wilcoxon rank-sum test was used to test whether expression of PCDH genes is higher in normal than in tumor specimens. Significance of change is indicated by each gene name: **, P value <0.05; *, P value <0.1. Box plots display 50% of data points and show median, upper, and lower quartiles.

that the low level of stochastic DNA methylation found in many CpG islands may serve as seeds or nucleation sites that trigger *de novo* methylation of the surrounding area leading to a more stable epigenetic silencing event (49). This model would be consistent with the results seen in the *PCDH* and *HOXA* gene clusters because there is a surprisingly high level of DNA methylation in some of their CpG islands in normal tissue and finite life span mammary epithelial cells. Indeed, it is further possible that these sequences themselves are more prone to DNA methylation (50).

In summary, recent epigenome-wide scanning approaches have allowed for the identification of new types of epigenetic lesions in the cancer genome (9–12). One of these epigenetic lesions, LRES, is characterized by epigenetic mechanisms that work together to functionally inactivate long contiguous stretches of the genome; however, it seems that the combination of epigenetic landmarks associated with the long range silencing may arise from different epigenetic partnering schemes. For instance, Frigola and colleagues (10) and Hitchins and colleagues (11) found aberrant increases in DNA methylation and histone H3 K9 dimethylation that were linked to long range silencing, and pharmacologic reactivation studies implicated histone acetylation state, as well. Stransky and colleagues (12) found long-range silencing associated with H3 K9 dimethylation in the absence of aberrant DNA hypermethylation. In contrast, in our previous studies (9), as well as the ones reported here, we found long range silencing associated with aberrant increases in DNA methylation and loss of the histone acetylation modifications linked to a permissive or active chromatin state, but we found no consistent increases in the repressive histone modification marks, K9 or K27 methylation (Supplementary Fig. S3).

The differences in epigenetic landmarks associated with long-range gene silencing observed between the separate studies could be technical in nature where the ability to detect long-range epigenetic aberrations currently is unrefined and limited in power. It is likely that increases in the sensitivity and precision of epigenomic research tools and analytic schemes will help identify and more precisely define long-range epigenetic aberrations in human disease. Alternatively, these differences may reflect unique tumor-specific mechanisms because the studies by Frigola and colleagues (10) and Hitchins and colleagues (11) analyzed colon tumors, whereas Starnsky and colleagues (12) analyzed bladder tumors, and our studies focused on breast cancer. We speculate that different external stimuli (e.g., carcinogens and environmental agents) may disturb different nodes of epigenetic control that then combine to create aberrant epigenetic terrains that are cancer- or disease-specific and not seen in normal, nondiseased cells.

Disclosure of Potential Conflicts of Interest

The authors declare that there are no conflicts of interests.

Acknowledgments

Received 4/15/2008; revised 8/6/2008; accepted 8/13/2008.

Grant support: R01CA65662 and R33CA0951 (B.W. Futscher); Center Grants P30ES06694 and P30CA023074, and the BIO5 interdisciplinary biotechnology center at the UA supported the Genomics Shared Service; and training grants ES007091 and CA09213 (T. Jensen).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Jose Munoz-Rodrigues for outstanding technical support.

References

1. Jones PA, Baylin SB. The epigenomics of cancer. Cell 2007;128:683–92.
2. Narayan A, Ji W, Zhang XY, et al. Hypomethylation of pericentromeric DNA in breast adenocarcinomas. Int J Cancer 1998;77:833–8.
3. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature 1983;301:89–92.
4. Jackson K, Yu MC, Arakawa K, et al. DNA hypomethylation is prevalent even in low-grade breast cancers. Cancer Biol Ther 2004;3:1225–31.
5. Vertino PM, Spillare EA, Harris CC, Baylin SB. Altered chromosomal methylation patterns accompany oncogene-induced transformation of human bronchial epithelial cells. Cancer Res 1993;53:1684–9.
6. Baylin SB. Abnormal regional hypermethylation in cancer cells. AIDS Res Hum Retroviruses 1992;8:811–20.
7. Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B. Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. Hum Genet 1989;83:155–8.
8. Costello JF, Fruhwald MC, Smiraglia DJ, et al. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. Nat Genet 2000;24:132–8.
9. Novak P, Jensen T, Oshiro MM, et al. Epigenetic inactivation of the HOXA gene cluster in breast cancer. Cancer Res 2006;66:10664–70.
10. Frigola J, Song J, Stirzaker C, Hinshelwood RA, Peinado MA, Clark SJ. Epigenetic remodeling in colorectal cancer results in coordinate gene suppression across an entire chromosome band. Nat Genet 2006;38:540–9.
11. Hitchins MP, Lin VA, Buckle A, et al. Epigenetic inactivation of a cluster of genes flanking MLH1 in microsatellite-unstable colorectal cancer. Cancer Res 2007;67:9107–16.
12. Stransky N, Vallot C, Reyal F, et al. Regional copy number-independent deregulation of transcription in cancer. Nat Genet 2006;38:1386–96.
13. Clark SJ, Harrison J, Paul CL, Frommer M. High sensitivity mapping of methylated cytosines. Nucleic Acids Res 1994;22:2990–7.
14. Coolen MW, Statham AL, Gardiner-Garden M, Clark SJ. Genomic profiling of CpG methylation and allelic specificity using quantitative high-throughput mass spectrometry: critical evaluation and improvements. Nucleic Acids Res 2007;35:e119.
15. Weber M, Davies JJ, Wittig D, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nat Genet 2005;37:853–62.
16. Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol 2004;5:R80.
17. Chen Z, McGee M, Liu Q, Scheuermann RH. A distribution free summarization method for Affymetrix GeneChip(R) arrays. Bioinformatics 2007;23:321–7.
18. Kent WJ. BLAT-the BLAST-like alignment tool. Genome Res 2002;12:656–64.
19. Jensen TJ, Novak P, Eblin KE, Gandolfi AJ, Futscher BW. Epigenetic Remodeling During Arsenical-Induced Malignant Transformation. Carcinogenesis 2008;29:1500–8.
20. Nouzova M, Holtan N, Oshiro MM, et al. Epigenomic changes during leukemia cell differentiation: analysis of histone acetylation and cytosine methylation using CpG island microarrays. J Pharmacol Exp Ther 2004;311:968–81.
21. Li J, Yang Y, Fujie T, et al. Expression of BAGE, GAGE, and MAGE genes in human gastric carcinoma. Clin Cancer Res 1996;2:1619–25.
22. De Backer O, Arden KC, Boretti M, et al. Characterization of the GAGE genes that are expressed in various human cancers and in normal testis. Cancer Res 1999;59:3157–65.
23. Coles C, Condie A, Chetty U, Steel CM, Evans HJ, Prosser J. p53 mutations in breast cancer. Cancer Res 1992;52:5291–8.
24. Wu Q, Maniatis T. A striking organization of a large family of human neural cadherin-like cell adhesion genes. Cell 1999;97:779–90.
25. Dammann R, Yang G, Pfeifer GP. Hypermethylation of the cpG island of Ras association domain family 1A (RASSF1A), a putative tumor suppressor gene from the 3p21.3 locus, occurs in a large percentage of human breast cancers. Cancer Res 2001;61:3105–9.
26. Herman JG, Merlo A, Mao L, et al. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant dna methylation in all common human cancers. Cancer Res 1995;55:4525–30.
27. Huang TH-M, Laux DE, Hamlin BC, Tran P, Tran H, Lubahn DB. Identification of DNA methylation markers for human breast carcinomas using the methylation-sensitive restriction fingerprinting technique. Cancer Res 1997;57:1030–4.
28. Miyamoto K, Fukutomi T, Akashi-Tanaka S, et al. Identification of 20 genes aberrantly methylated in human breast cancers. Int J Cancer 2005;116:407–14.
29. Oshiro MM, Kim CJ, Wozniak RJ, et al. Epigenetic silencing of DSC3 is a common event in human breast cancer. Breast Cancer Res 2005;7:R669–80.

30. Raman V, Martensen SA, Reisman D, et al. Compromised HOXA5 function can limit p53 expression in human breast tumours. Nature 2000;405:974–8.
31. Rauch TA, Zhong X, Wu X, et al. High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. Proc Natl Acad Sci U S A 2008;105:252–7.
32. Ehrich M, Nelson MR, Stanssens P, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A 2005;102:15785–90.
33. Xie X, Mikkelsen TS, Gnirke A, Lindblad-Toh K, Kellis M, Lander ES. Systematic discovery of regulatory motifs in conserved regions of the human genome, including thousands of CTCF insulator sites. Proc Natl Acad Sci U S A 2007;104:7145–50.
34. Kim TH, Abdullaev ZK, Smith AD, et al. Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome. Cell 2007;128:1231–45.
35. Filippova GN, Lindblom A, Meincke LJ, et al. A widely expressed transcription factor with multiple DNA sequence specificity, CTCF, is localized at chromosome segment 16q22.1 within one of the smallest regions of overlap for common deletions in breast and prostate cancers. Genes Chromosomes Cancer 1998;22:26–36.
36. Filippova GN, Qi CF, Ulmer JE, et al. Tumor-associated zinc finger mutations in the CTCF transcription factor selectively alter tts DNA-binding specificity. Cancer Res 2002;62:48–52.
37. Butcher DT, Rodenhiser DI. Epigenetic inactivation of BRCA1 is associated with aberrant expression of CTCF and DNA methyltransferase (DNMT3B) in some sporadic breast tumours. Eur J Cancer 2007;43:210–9.
38. Filippova GN, Cheng MK, Moore JM, et al. Boundaries between chromosomal domains of X inactivation and escape bind CTCF and lack CpG methylation during early development. Dev Cell 2005;8:31–42.
39. Butcher DT, Mancini-DiNardo DN, Archer TK, Rodenhiser DI. DNA binding sites for putative methylation boundaries in the unmethylated region of the BRCA1 promoter. Int J Cancer 2004;111:669–78.
40. De La Rosa-Velazquez IA, Rincon-Arano H, Benitez-Bribiesca L, Recillas-Targa F. Epigenetic regulation of the human retinoblastoma tumor suppressor gene promoter by CTCF. Cancer Res 2007;67:2577–85.
41. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM. CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. Nature 2000;405:486–9.
42. Bell AC, Felsenfeld G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. Nature 2000;405:482–5.
43. Whitman SP, Liu S, Vukosavljevic T, et al. The MLL partial tandem duplication: evidence for recessive gain-of-function in acute myeloid leukemia identifies a novel patient subgroup for molecular-targeted therapy. Blood 2005;106:345–52.
44. Guenther MG, Jenner RG, Chevalier B, et al. Global and Hox-specific roles for the MLL1 methyltransferase. Proc Natl Acad Sci U S A 2005;102:8603–8.
45. Yu BD, Hess JL, Horning SE, Brown GAJ, Korsmeyer SJ. Altered Hox expression and segmental identity in Mll-mutant mice. Nature 1995;378:505–8.
46. Kleer CG, Cao Q, Varambally S, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. Proc Natl Acad Sci U S A 2003;100:11606–11.
47. Decensi A, Serrano D, Bonanni B, Cazzaniga M, Guerrieri-Gonzaga A. Breast cancer prevention trials using retinoids. J Mammary Gland Biol Neoplasia 2003;8:19–30.
48. Langston AW, Gudas LJ. Retinoic acid and homeobox gene regulation. Curr Opin Genet Dev 1994;4:550–5.
49. Song JZ, Stirzaker C, Harrison J, Melki JR, Clark SJ. Hypermethylation trigger of the glutathione-S-transferase gene (GSTP1) in prostate cancer cells. Oncogene 2002;21:1048–61.
50. Feltus FA, Lee EK, Costello JF, Plass C, Vertino PM. DNA motifs associated with aberrant CpG island methylation. Genomics 2006;87:572–9.

Agglomerative Epigenetic Aberrations Are a Common Event in Human Breast Cancer

Petr Novak, Taylor Jensen, Marc M. Oshiro, et al.

Cancer Res 2008;68:8616-8625.