

# Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay

Tibor Rauch\*, Zunde Wang\*, Xinmin Zhang<sup>†</sup>, Xueyan Zhong\*, Xiwei Wu<sup>‡</sup>, Sean K. Lau<sup>§</sup>, Kemp H. Kernstine<sup>¶</sup>, Arthur D. Riggs\*<sup>||</sup>, and Gerd P. Pfeifer\*<sup>||</sup>

\*Division of Biology, <sup>‡</sup>Division of Information Sciences, <sup>§</sup>Division of Pathology, and <sup>¶</sup>Division of Surgery, Beckman Research Institute of the City of Hope, Duarte, CA 91010; and <sup>†</sup>NimbleGen Systems, Inc., Madison, WI 53711

Contributed by Arthur D. Riggs, February 5, 2007 (sent for review December 11, 2006)

**De novo** methylation of CpG islands is a common phenomenon in human cancer, but the mechanisms of cancer-associated DNA methylation are not known. We have used tiling arrays in combination with the methylated CpG island recovery assay to investigate methylation of CpG islands genome-wide and at high resolution. We find that all four *HOX* gene clusters on chromosomes 2, 7, 12, and 17 are preferential targets for DNA methylation in cancer cell lines and in early-stage lung cancer. CpG islands associated with many other homeobox genes, such as *SIX*, *LHX*, *PAX*, *DLX*, and *Engrailed*, were highly methylated as well. Altogether, more than half (104 of 192) of all CpG island-associated homeobox genes in the lung cancer cell line A549 were methylated. Analysis of paralogous *HOX* genes showed that not all paralogues undergo cancer-associated methylation simultaneously. The *HOXA* cluster was analyzed in greater detail. Comparison with ENCODE-derived data shows that lack of methylation at CpG-rich sequences correlates with presence of the active chromatin mark, histone H3 lysine-4 methylation in the *HOXA* region. Methylation analysis of *HOXA* genes in primary squamous cell carcinomas of the lung led to the identification of the *HOXA7*- and *HOXA9*-associated CpG islands as frequent methylation targets in stage 1 tumors. Homeobox genes are potentially useful as DNA methylation markers for early diagnosis of the disease. The finding of widespread methylation of homeobox genes lends support to the hypothesis that a substantial fraction of genes methylated in human cancer are targets of the Polycomb complex.

DNA methylation | HOX genes | chromatin | Polycomb

DNA methylation at CpG dinucleotides is an important epigenetic modification carried out by DNA methyltransferases (1, 2). *De novo* methylation of CpG islands that overlap with promoter regions is commonly associated with gene silencing and is a frequent event that accompanies tumorigenesis (3–9). Cancer-specific hypermethylation of genes that suppress uncontrolled cell proliferation or promote genome stability is a key step in tumor development.

Homeobox genes encode a transcription factor family that plays decisive roles in embryogenesis and differentiation of adult cells (10). Homeobox proteins are classified into one family on the basis of their evolutionary conserved helix–loop–helix DNA-binding motif, called the homeodomain. Apart from the homeodomain, family members share only limited conservation outside of the DNA-binding motif. Most of the family members are scattered throughout the genome, but a subgroup of the homeobox genes, *HOX* genes, are organized into clusters. *HOX* genes were originally identified in *Drosophila* as factors involved in homeotic transformations (11). During mammalian evolution, the ancient *HOX* cluster underwent duplications that were followed by gene losses that finally led to the emergence of the 39 present *HOX* genes organized into four clusters (10). *HOX* proteins are essential switches of developmental stage- and cell-specific gene regulation,

and in this way they are key determinants of cell identity and potential targets during tumorigenesis.

We recently developed a DNA methylation detection technique, the methylated CpG island recovery assay (MIRA) (12). This technique is based on the high affinity of a complex of the MBD2b and MBD3L1 proteins for CpG-methylated DNA (13) and is compatible with microarray-based methodology (14). We previously found that several homeobox genes were methylated in a lung cancer cell line (14). In the present study, we combined the MIRA technique with tiling arrays to obtain genome-wide CpG island coverage as well as high-resolution data for DNA methylation within *HOX* gene clusters in cell lines and primary lung tumors. The data indicate that multiple CpG islands within *HOX* clusters and near other homeobox genes are frequent methylation targets in lung cancer. An in-depth analysis of the *HOXA* cluster revealed DNA methylation markers for stage 1 lung cancer.

## Results

**Methylation of the *HOXA* Cluster.** To explore the use of MIRA-assisted tiling platforms for genome-wide DNA methylation analysis, we first used NimbleGen's ENCODE tiling arrays. The ENCODE array is designed to identify functional elements in  $\approx 1\%$  of the human genome. On these arrays, the neighboring tiling oligonucleotides (50 bp long) overlap with each other at 12-bp-long sequences, and in this way 30 Mb, including the *HOXA* chromosomal region, is fully covered. First, we analyzed methylation patterns by using DNA obtained from the lymphoblastoid cell line GM06990, which is one of the cell lines used in the ENCODE project. We compared the MIRA-enriched fraction with the input fraction. Samples were prepared as described previously (14) by using MseI digestion, linker ligation, amplification, and labeling of input and MIRA-enriched DNA. Upon examination of the ENCODE regions, we observed that the highest level of methylation was found on chromosome 7 in a region including the *HOXA* gene cluster. Three independent MIRA reactions with GM06990 DNA were conducted. Fig. 1 *Upper* shows the high reproducibility of this approach, which is aided by the overlap features of the tiling array. We found that several CpG islands within the *HOXA* cluster were strongly methylated in GM06990 cells. Confirmation of the methylation status was obtained by combined bisulfite restriction analysis

Author contributions: K.H.K., A.D.R., and G.P.P. designed research; T.R., Z.W., X. Zhang, X. Zhong, and S.K.L. performed research; X. Zhang contributed new reagents/analytic tools; T.R., Z.W., X.W., A.D.R., and G.P.P. analyzed data; and T.R., A.D.R., and G.P.P. wrote the paper.

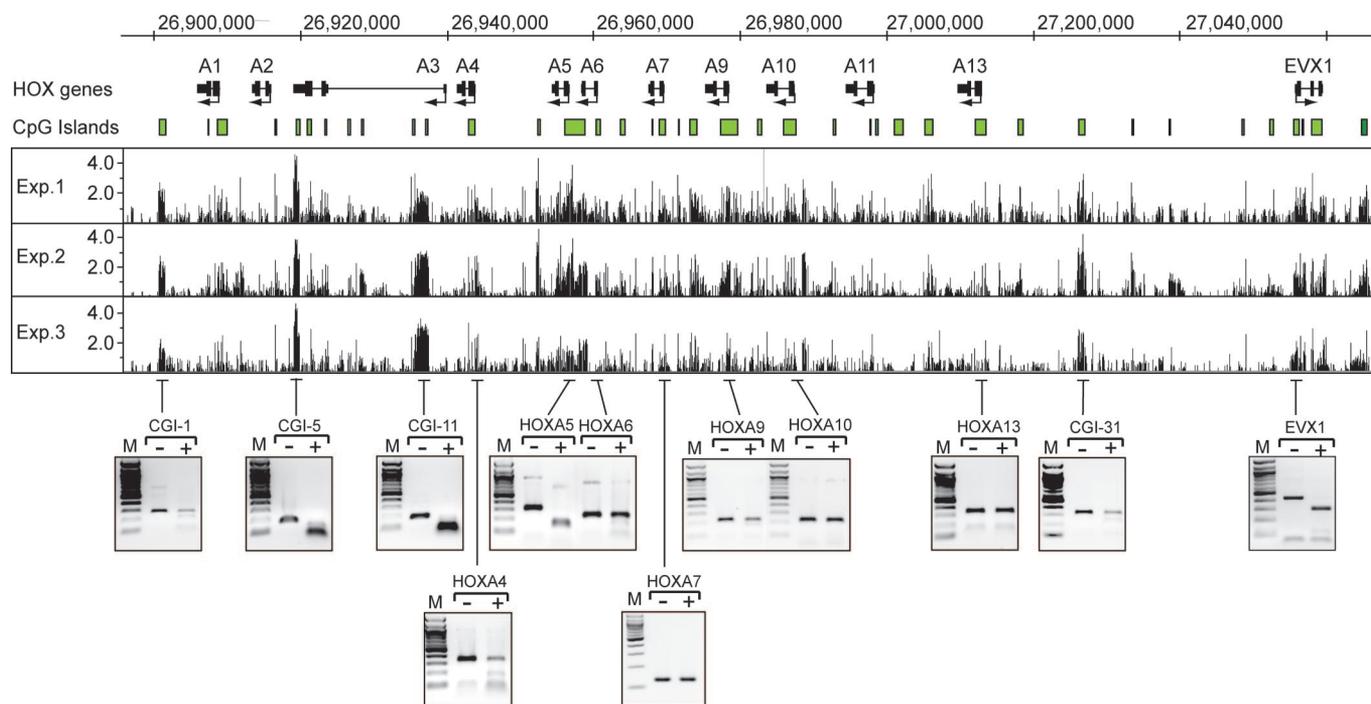
The authors declare no conflict of interest.

Abbreviations: COBRA, combined bisulfite restriction analysis; MIRA, methylated CpG island recovery assay; NHBE cells, normal human bronchial epithelial cells.

<sup>||</sup>To whom correspondence may be addressed. E-mail: ariggs@coh.org or gpfeifer@coh.org.

This article contains supporting information online at [www.pnas.org/cgi/content/full/0701059104/DC1](http://www.pnas.org/cgi/content/full/0701059104/DC1).

© 2007 by The National Academy of Sciences of the USA



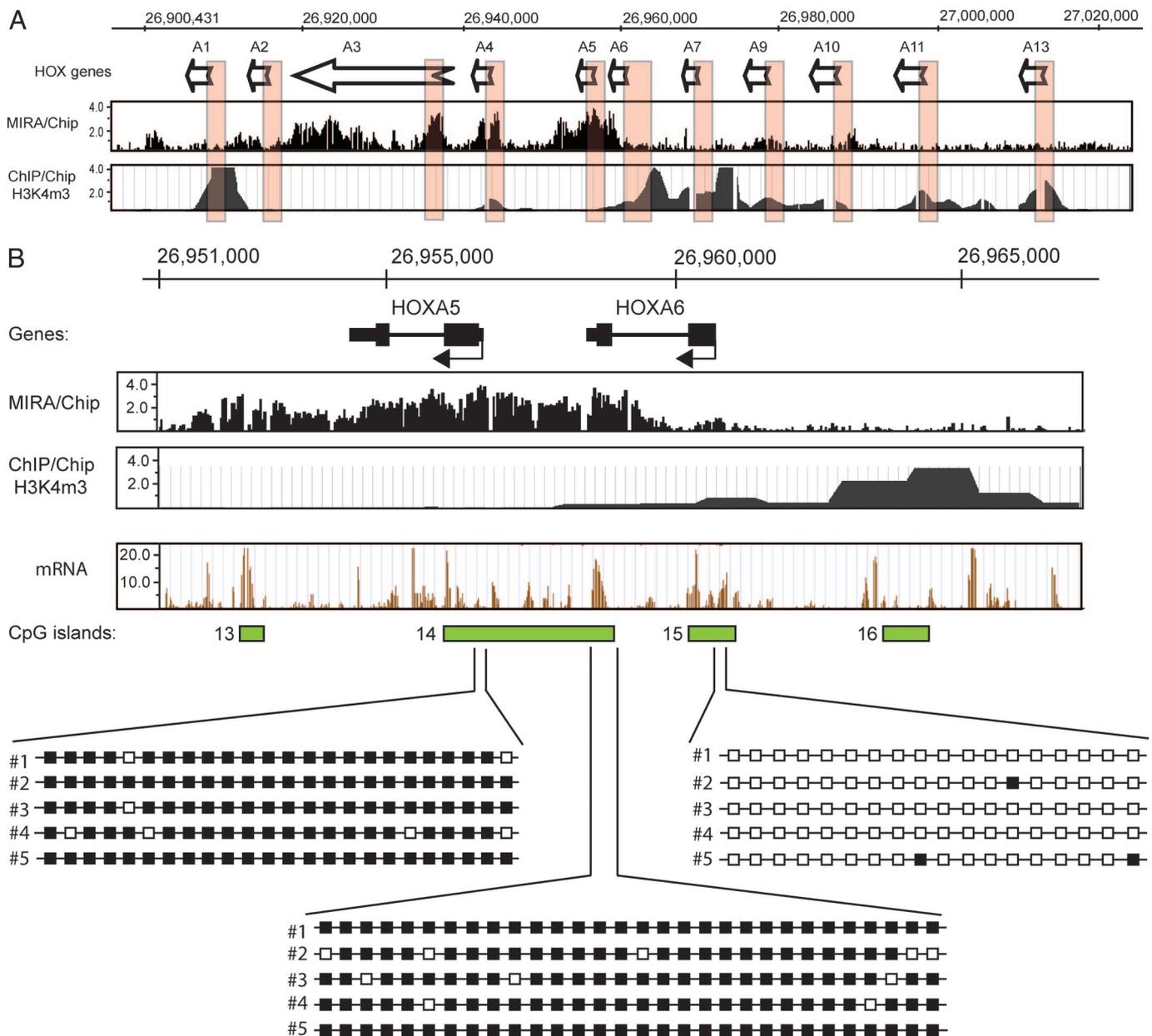
**Fig. 1.** Analysis of the *HOXA* cluster region on chromosome 7 with ENCODE NimbleGen tiling arrays. The analysis covers a 155-kb region. DNA from the GM06990 cell line was used for three independent MIRA reactions by using the amplification and labeling procedure as described in *Materials and Methods*. (Upper) The signal ratio of MIRA-enriched versus input DNA is plotted along the chromosome. The green boxes indicate the 39 CpG islands (see [SI Table 1](#)). The numbering refers to *HOXA* gene number, and the arrows indicate the direction of transcription. A COBRA using *Bst*UI digestion (see *Materials and Methods*) was conducted for the regions indicated to confirm the methylation data obtained by MIRAs. Some of the *Bst*UI cleavage fragments are too small to be visible on the gels. –, to control digestion with no *Bst*UI; +, *Bst*UI-digested samples.

(COBRA) methylation analysis (Fig. 1 Lower) and by bisulfite sequencing [Fig. 2 and [supporting information \(SI\) Fig. 5](#)].

The *HOXA* cluster contains 12 genes (11 *HOX* genes and *EVX1*) and is contained in a 155-kb-long genomic region. Most of the *HOXA* promoters are embedded in one of the 39 CpG islands ([SI Table 1](#)) residing in the locus. The high CpG island density makes the cluster an ideal target for testing the interdependence of neighboring DNA methylation patterns, including, for example, theories of long-range epigenetic silencing (15). According to this hypothesis, genes located in the same neighborhood are coordinately silenced by epigenetic modifications. Our methylation profile analysis does show that most CpG islands were methylated within the *HOXA* cluster (Fig. 1). However, not all CpG islands were methylated. Highly methylated and poorly or nonmethylated CpG islands could be next to each other within the cluster (Figs. 1 and 2; [SI Tables 1 and 2](#)). To scrutinize this theory in more detail, we focused on two neighboring CpG islands that reside just 4 kb apart in the middle of the *HOXA* cluster. COBRAs (Fig. 1) and bisulfite sequencing data (Fig. 2) confirmed that the methylation status of these two neighboring CpG islands is opposite. Hypermethylation of the *HOXA5* promoter suggests the formation of inactive chromatin, whereas the unmethylated status of the neighboring *HOXA6* promoter should correlate with an active chromatin configuration. These expectations were confirmed by comparison with chromatin immunoprecipitation (ChIP)-on-chip experiments previously performed with the GM06990 cell line with anti-histone H3K4me3 antibody (from the ENCODE database) and by tiling expression array data for the two exons of *HOXA5* and *HOXA6* (from the ENCODE database) (Fig. 2). The active gene-associated histone methylation mark was not detected at the methylated *HOXA5* promoter. The region upstream of the *HOXA6* gene, in particular CpG island 16, showed a strong signal for trimethylated K4 on histone H3 that defines an active chromatin state. Using 5' RACE, we detected an alternatively spliced transcript emanating from CpG

island 16 and containing parts of the two *HOXA6* exons (data not shown). The 3' end of CpG island 14 was methylated, indicating that a methylation boundary is between the 3' end of CpG island 14 and CpG island 15 (Fig. 2).

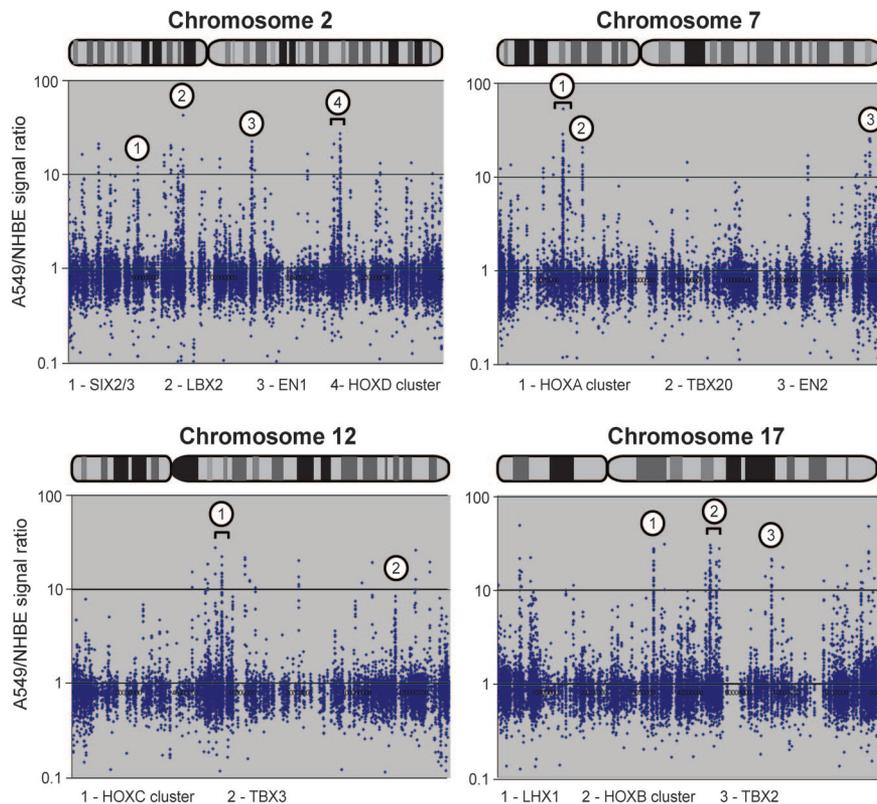
**The Four *HOX* Gene Clusters Are Preferred Methylation Targets in Lung Cancer Cells.** Cancer-associated DNA methylation changes commonly affect CpG islands that are often associated with gene promoters and undergo hypermethylation in tumors. We next used Agilent CpG island arrays that contain 27,800 CpG islands covering in a tiling fashion 21 Mb of the human genome. Combining MIRA with Agilent arrays, we compared the methylation status of CpG islands genome-wide between the lung cancer cell line A549 and normal human bronchial epithelial (NHBE) cells. Although the four *HOX* gene clusters reside on different chromosomes, we found that each of them is the target of extensive *de novo* methylation in lung cancer cells (Fig. 3 and [SI Table 2](#)). Fig. 3 shows that besides the *HOX* clusters there are several other cancer-related hypermethylation hot spots showing >10-fold enrichment by MIRA in the lung cancer cell line A549. Strikingly, many of them define other homeobox genes, for instance the engrailed homologues *EN1* and *EN2*, on chromosomes 2 and 7, respectively, and the *SIX3/2* gene pair, as well as *LBX2* and *LHX1*. Several T box genes (*TBX2*, *TBX3*, and *TBX20*) also were preferred chromosomal methylation sites (Fig. 3). In addition to those, several paired-box homologues (*PAX* genes 1–9, except *PAX1* and *PAX4*) and many other homeobox genes were methylated in this tumor cell line ([SI Table 3](#)). The occurrence of apparent methylation hot spots as displayed along the entire length of a chromosome is most often related to the simultaneous methylation of several CpG islands within a chromosomal segment. This observation is illustrated for the *HOX* gene clusters (Figs. 1 and 2 and [SI Tables 1 and 2](#)) but also can be seen for other chromosomal methylation hot spots. As an example, we illustrate the methylation of several CpG islands within or near the



**Fig. 2.** DNA and histone methylation profile at *HOXA* cluster genes in GM06990 cells. (A) DNA and histone methylation profile analysis of the entire *HOXA* cluster in GM06990 cells. The MIRA-enriched methylated DNA fraction and input fraction were mixed and hybridized onto NimbleGen ENCODE arrays. Note that in this experiment, the MIRA-enriched DNA and the input DNA were processed without amplification and were directly labeled and hybridized to the array to make these data methodologically comparable with the ChIP data. ChIP with anti-histone H3K4m3 antibody was performed, and the DNA was hybridized onto Sanger ENCODE3.1.1 DNA microarrays (data were obtained from the ENCODE database). The pink shading indicates the promoter-associated CpG islands. (B) (Upper) DNA and histone methylation and mRNA expression profile analysis of the *HOXA5* and *HOXA6* genes. Data for H3K4 methylation and mRNA expression are from the ENCODE database. (Lower) Bisulfite sequencing verification of the DNA methylation status of the indicated CpG island regions.

*EN1* and *TBX20* genes in SI Fig. 6. In all, 104 of 194 (54%) of the homeobox genes associated with CpG islands in the lung cancer cell line A549 were methylated (SI Tables 2 and 3). We used rather stringent criteria for making this assignment, and the percentage of methylated homeobox genes is likely an underestimation because of the elimination of low signal intensity spots, for instance caused by large spacing of adjacent restriction sites and low amplification efficiency. Also, homeobox genes without CpG islands are not represented on the array, and some may be methylated already in normal cells. Taken together, these findings suggest that the inactivation of clusters of CpG islands near homeobox genes is a common event in this lung cancer cell line.

**Methylation Analysis of *HOX* Gene Paralogues.** *HOX* proteins belonging to a given paralogous group can be functionally similar and partially interchangeable (16–18). To gain more information on the methylation profile of the individual genes that constitute a paralogous group, we analyzed the Agilent CpG island arrays harbor most of the 39 *HOX* gene promoters. We determined the hypermethylation profile of each *HOX* gene promoter for all of the *HOX* gene clusters in the lung cancer cell line A549 versus NHBE cells (SI Table 2). The data show that many CpG islands in the *HOX* clusters were hypermethylated in A549 lung cancer cells relative to NHBE cells. We focused our attention on the *HOX6* and *HOX7* paralogues. We



**Fig. 3.** DNA methylation profiles of chromosomes 2, 7, 12, and 17 that carry the *HOXA*, *HOXB*, *HOXC*, and *HOXD* gene clusters. MIRA-enriched fractions from A549 lung cancer cells and from NHBE cells were mixed and hybridized onto Agilent CpG island arrays. The ratio (fold difference) plotted indicates individual probe signals (blue diamonds) for hypermethylated CpG islands in A549 cells (A549/NHBE signal ratio). The numbers indicate selected areas of densely hypermethylated CpG island clusters for the genes indicated at the bottom.

found that the *HOXA6* and *HOXB6* promoters were not methylated in NHBE cells, but both promoters were subject to dense DNA methylation in the tumor cell line (SI Fig. 7). Although *HOXC6* belongs to the same paralogous group, there is no CpG island near its promoter region, and it is not represented on the array. The two *HOX7* paralogous genes (*HOXA7* and *HOXB7*) are methylation-free in normal cells, but in the lung cancer cell line one of them (*HOXB7*) kept the methylation-free status, whereas the other one (*HOXA7*) became hypermethylated (SI Fig. 7). Noncoordinate methylation was seen for other paralogues as well (SI Table 2), indicating that DNA methylation in tumors does not necessarily simultaneously affect all members of a paralogous group.

**Methylation of *HOXA* Genes in Primary Lung Tumors.** To investigate whether methylation of *HOX* cluster genes is present in human primary lung tumors, we first analyzed pairs of normal lung tissue and adjacent stage 1 adenocarcinoma or squamous cell carcinoma (SI Figs. 8 and 9). We found that all four *HOX* clusters, as well as several other homeobox genes, were hypermethylated in the early-stage adenocarcinoma. Data for the squamous cell carcinoma show that the *HOXA* and *HOXD* clusters were highly methylated and the *HOXB* and *HOXC* clusters were methylated to a lesser extent.

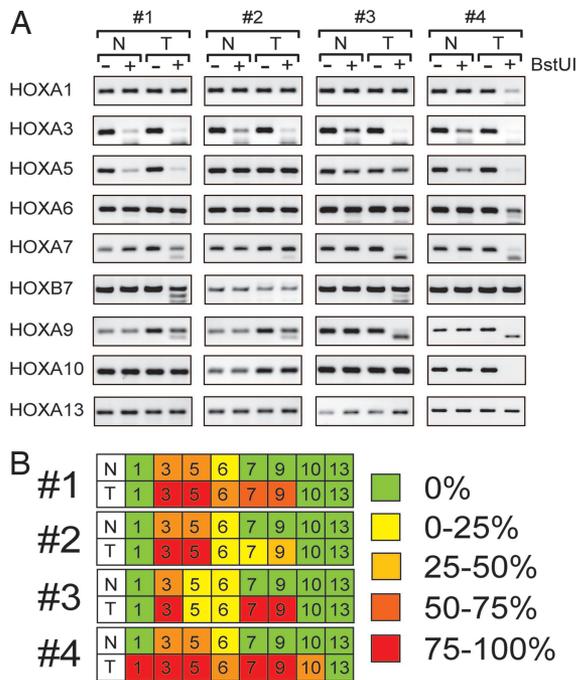
We then analyzed *HOXA* cluster genes in four stage 1 lung squamous cell carcinoma samples and their matching normal tissues by COBRA methylation assays (Fig. 4). In pilot experiments, we found that these early-stage tumor samples were not methylated at the promoters of the following well known lung cancer DNA methylation marker genes: *RASSF1A* (19), *PAX5 $\alpha$*  (20), *DLEC1* (14), and *RARB2* (21). The *p16* (22) and *TCF21* (23) promoters showed partial methylation (data not shown). We next systematically checked the methylation status of the CpG islands in the *HOXA* cluster. The *HOXA* genes near the 3' end of the cluster

(*HOXA3* and *HOXA5*) were partially methylated in the normal lung samples, and the methylation levels were moderately or strongly elevated in the matching tumor pairs (Fig. 4), suggesting the possibility that the preexisting low levels of methylation made these promoters more susceptible to cancer-associated hypermethylation. The *HOXA6* gene is a methylation boundary region in normal tissue samples (Fig. 4). Distinct from the 3' located genes, DNA methylation can only weakly be detected at the *HOXA6* promoter by COBRAs in normal lung cells. Upstream of the *HOXA6* gene, the other promoters are not methylated in normal lung samples, but some of them are cancer-specifically methylated in tumor samples. The *HOXA7* and *HOXA9* promoters show high levels of DNA methylation in stage 1 squamous cell carcinomas.

To extend the analysis of *HOX* gene methylation to an additional series of primary lung tumors, we analyzed the methylation status of the CpG islands associated with the *HOXA7* and *HOXA9* genes in primary lung squamous cell carcinomas (SI Fig. 10). No methylation of *HOXA7* or *HOXA9* was detected in any of the 18 normal lung tissues. However, the *HOXA7* CpG island was methylated in 10 of 22 tumors (45%), and the *HOXA9* CpG island was methylated in 15 of the 22 tumors (68%). Remarkably, *HOXA9* was methylated in 8 of 10 (80%) of the stage 1 tumors.

### Discussion

To characterize the DNA methylation status of the human *HOX* gene clusters we combined the MIRA method with region-specific genomic and genome-wide CpG island tiling arrays. This technology proved to be highly reproducible, and methylation differences could generally be confirmed by standard bisulfite-based methylation assays. *HOX* gene clusters containing multiple CpG islands were preferential methylation targets in lung cancer cells and primary lung tumors. Although many of the CpG islands within the



**Fig. 4.** Methylation of *HOXA* genes in normal lung tissue and matching stage 1 squamous cell carcinoma samples. (A) Methylation differences between squamous cell carcinomas and matching normal pairs (pairs 1–4) were detected by COBRAs of *HOXA* gene targets. T, tumor; N, normal tissues; –, control digestion with no BstUI; +, BstUI-digested samples. (B) Summary of the methylation status of promoters in the *HOXA* gene cluster.

*HOX* clusters were methylated, we found that nonmethylated and highly methylated CpG islands can still be next to each other in the gene-rich *HOX* clusters (Figs. 1 and 2 and SI Tables 1 and 2). For example, the *HOXA5* and *HOXA6* promoters are located just 4 kb apart and are embedded in CpG islands. We found that the methylation status of the two promoters is the opposite in the lymphoblastoid cell line GM06990 (Figs. 1 and 2). These findings suggest that cancer-associated CpG island methylation still operates on an island-by-island basis within the *HOX* clusters rather than by a long-range effect encompassing all CpG islands within a specific chromosomal domain as reported for the 2q.14.2 chromosomal region in colon cancer (15). ChIP-on-chip data with anti-histone H3 trimethylated K4 antibody (ENCODE database) negatively correlated with the MIRA-assisted microarray data. DNA hypermethylation and histone H3 K4 trimethylation were mutually exclusive along the *HOXA* cluster (Fig. 2).

The human *HOX* genes are organized in four clusters, and based on homeobox sequence similarities, the individual *HOX* genes are classified into 13 paralogous groups. Our data suggest that paralogues do not always undergo simultaneous methylation in tumor cells.

Many homeobox genes are aberrantly expressed in a variety of hematological malignancies and solid tumors, most commonly showing up-regulation and less commonly down-regulation (24, 25). However, many of these genes are not expressed in adult tissues, including the lung (24, 26). In some instances, the lack of their expression may promote *de novo* methylation during malignant progression. Several homeobox genes are methylated in tumors of various histological origins. For example, methylation of *HOXB13* occurs in 30% of renal cell carcinomas (27), and methylation of genes in the *HOXA* and *HOXD* clusters was reported in lung cancer (28). The *HOXA5* promoter region was methylated in 16 of 20 p53-negative breast tumor specimens (29). While this manuscript was in preparation, methylation of the *HOXA* gene cluster in breast

cancer was reported (30). Our data indicate that dense methylation of the four *HOX* gene clusters and many other homeobox genes located on different chromosomes occurs in lung cancer.

Homeodomain-containing proteins are highly conserved in their DNA-binding motif and can bind to similar cis elements in *in vitro* experiments (31). However, by interacting with other transcription factors they can diversify their binding targets. Interacting partners for *HOX* proteins can be other homeodomain-containing factors (32–35), and they synergistically govern the expression of downstream target genes. We noticed that nonclustered homeobox genes are also frequent targets of hypermethylation (Fig. 3 and SI Table 3). Hypermethylation of clustered and nonclustered homeobox genes may cause misregulation of a finely tuned regulatory network. However, it remains to be determined whether methylation-induced silencing of homeobox genes contributes to tumorigenesis directly or is merely associated with it.

The widespread and pervasive nature of homeobox gene methylation in lung cancer (104 of 194 genes analyzed in A549 cells) suggests a common mechanistic pathway that promotes *de novo* methylation of these genes during tumorigenesis. Polycomb complexes are involved in silencing of homeobox genes (36–38). A recent genome-wide analysis of the localization of Polycomb components including SUZ12, which is required for the histone methyltransferase activity and silencing function of the EED–EZH2 complex, has identified a large fraction of the targets as homeobox genes (39–41). EZH2 expression is increased in tumors of different histological types, including precancerous tissues (42, 43). The Polycomb component EZH2 associates with DNA methyltransferase activity and can promote DNA hypermethylation (44–46), and a speculative mechanism can be proposed that links Polycomb silencing with tumor-associated DNA methylation. This mechanistic link has received support from recent studies finding good concordance between Polycomb occupancy of genes in noncancerous cells and tissues including ES cells with cancer-associated DNA hypermethylation events (47–49). Many of the Polycomb targets are homeobox genes and other key developmental regulatory genes. They overlap with those genes we find hypermethylated in lung cancer, lending strong support to the Polycomb connection.

Alternatively, a regional breakdown in activating factors that maintain active chromatin domains at homeobox genes, such as MLL1, a member of the mammalian trithorax group, and H3K4 methyltransferase, may ultimately lead to a shift in histone modification patterns and *de novo* methylation in cancer cells (50–52). A complete understanding of the range and patterns of CpG island methylation within tumors and premalignant lesions will be an important step toward understanding of the mechanistic pathways leading to hypermethylation of genes during tumorigenesis.

Finally, the extent and tumor specificity of homeobox gene methylation provide a potentially useful source of DNA methylation markers for detecting early stage disease (8, 53–56). With relevance to lung cancer, diagnostic tests of the population at risk (heavy smokers) employing noninvasive technology such as sputum analysis could supplement high-resolution imaging technologies to provide a sensitive and more specific way for assessing the presence of early-stage tumors. In particular, methylation of *HOXA9* in a high percentage of early-stage squamous cell carcinomas may be one such promising marker.

### Materials and Methods

**MIRA.** MIRA was done essentially as described previously (14) with minor modifications. Briefly, purified GST-tagged MBD2b protein was eluted from a glutathione–Sepharose CL-4B matrix (Amersham Biosciences, Piscataway, NJ) with elution buffer (50 mM Tris-HCl, pH 8.5/150 mM NaCl/20 mM glutathione/0.1% Triton X-100) for 4 h at 4°C. The eluted GST-MBD2b fraction was dialyzed against PBS for 5 h and then overnight against 50 mM Hepes, pH 7.4/150 mM NaCl/5 mM 2-mercaptoethanol/50% (vol/vol) glycerol. GST-tagged MBD2b protein was kept at –20°C.

His-tagged MBD3L1 was prepared as described previously (14). Genomic DNA was fragmented by MseI digestion, and linker ligation was done as described earlier (14). The linker-ligated fraction was incubated with GST-MBD2b and His-MBD3L1 proteins overnight as described previously (14). MagneGST beads (2.5  $\mu$ l) (Promega, Madison, WI) preblocked with JM110 bacterial DNA, were added to the binding reaction and incubated at 4°C for 45 min. Beads were washed three times with washing buffer (10 mM Tris-HCl, pH 7.5/700 mM NaCl/1 mM EDTA/3 mM MgCl<sub>2</sub>/0.1% Triton X-100), and the methylated CpG-enriched fraction was eluted by using Qiaquick PCR purification kits (Qiagen, Valencia, CA). Eluted fractions or MseI-digested and linker-ligated input fractions were PCR-amplified as described previously (14).

**Sample Labeling and Hybridization to NimbleGen and Agilent Arrays.** The labeling of dsDNA, microarray hybridization, and scanning were performed by the NimbleGen Service Laboratory (Madison, WI) as described previously (57). The NimbleGen ENCODE array, which contains  $\approx$ 385,000 50-mer oligonucleotides and covers the ENCODE regions at 38-bp spacing, was used. Data were extracted from scanned images by using NimbleScan 2.3 extraction software (NimbleGen Systems, Inc.).

Human CpG island microarrays, which contain 237,000 oligonucleotide probes covering 27,800 CpG islands, were purchased from Agilent Technologies (Santa Clara, CA). Genomic DNA was fragmented by MseI (5'TTAA) and Csp6I (5'GTAC) digestion and then subjected to linker ligation and MIRA as described above.

Two micrograms each of the amplicons from MIRA-enriched tumor DNA and control samples were labeled with BioPrime Array CGH Genomic Labeling kit (Invitrogen, Carlsbad, CA) with either Cy5-dCTP (tumor) or Cy3-dCTP (control) in 87.5- $\mu$ l reactions (both Cy3- and cy5-dCTP were obtained from GE Healthcare, Piscataway, NJ). The purified labeled samples were then mixed, and microarray hybridization was performed according to the Agilent ChIP-on-chip protocol (version 9.0). The hybridized arrays were scanned on an Axon 4000B microarray scanner (Molecular Devices, Sunnyvale CA), and the images were analyzed with Axon GenePix software version 5.1. Image and data analysis were done as described previously (14).

**DNA Methylation Analysis Using COBRA and Bisulfite Sequencing.** Stage 1 lung adenocarcinoma and squamous cell carcinoma samples and matching normal tissues removed with surgery were obtained from the frozen tumor bank of the City of Hope National Medical Center (Duarte, CA). The COBRAs were done according to the method of Xiong and Laird (58) using digestion with BstUI (5'-CGCG). DNA was treated and purified with an EpiTect Bisulfite kit (Qiagen). PCR primers for amplification of specific targets in bisulfite-treated DNA are listed in SI Table 4. For sequence analysis, the PCR products obtained after bisulfite conversion were cloned into the pDrive PCR cloning vector (Qiagen), and five individual clones were sequenced.

We thank Steven Bates for assistance with cell culture. This work was supported by National Institutes of Health Grant CA104967 (to G.P.P.).

- Riggs AD (1975) *Cytogenet Cell Genet* 14:9–25.
- Holliday R, Pugh JE (1975) *Science* 187:226–232.
- Baylin SB, Ohm JE (2006) *Nat Rev Cancer* 6:107–116.
- Jones PA, Baylin SB (2002) *Nat Rev Cancer* 3:415–428.
- Costello JF, Fruhwald MC, Smiraglia DJ, Rush LJ, Robertson GP, Gao X, Wright FA, Feramisco JD, Peltomaki P, Lang JC, et al. (2000) *Nat Genet* 24:132–138.
- Esteller M, Corn PG, Baylin SB, Herman JG (2001) *Cancer Res* 61:3225–3229.
- Issa JP (2004) *Nat Rev Cancer* 4:988–993.
- Ushijima T (2005) *Nat Rev Cancer* 5:223–231.
- Egger G, Liang G, Aparicio A, Jones PA (2004) *Nature* 429:457–463.
- Garcia-Fernandez J (2005) *Nat Rev Genet* 6:881–892.
- Lewis EB (1978) *Nature* 276:565–570.
- Rauch T, Pfeifer GP (2005) *Lab Invest* 85:1172–1180.
- Jiang CL, Jin SG, Pfeifer GP (2004) *J Biol Chem* 279:52456–52464.
- Rauch T, Li H, Wu X, Pfeifer GP (2006) *Cancer Res* 66:7939–7947.
- Frigola J, Song J, Storzaker C, Hinshelwood RA, Peinado MA, Clark SJ (2006) *Nat Genet* 38:540–549.
- Tvrđik P, Capecchi MR (2006) *Dev Cell* 11:239–250.
- Greer JM, Puetz J, Thomas KR, Capecchi MR (2000) *Nature* 403:661–665.
- Wellik DM, Hawkes PJ, Capecchi MR (2002) *Genes Dev* 16:1423–1432.
- Dammann R, Li C, Yoon JH, Chin PL, Bates S, Pfeifer GP (2000) *Nat Genet* 25:315–319.
- Palmisano WA, Crume KP, Grimes MJ, Winters SA, Toyota M, Esteller M, Joste N, Baylin SB, Belinsky SA (2003) *Cancer Res* 63:4620–4625.
- Grote HJ, Schmiemann V, Geddert H, Rohr UP, Kappes R, Gabbert HE, Bocking A (2005) *Int J Cancer* 116:720–725.
- Bearzatto A, Conte D, Frattini M, Zaffaroni N, Andriani F, Balestra D, Tavecchio L, Daidone MG, Sozzi G (2002) *Clin Cancer Res* 8:3782–3787.
- Smith LT, Lin M, Brena RM, Lang JC, Schuller DE, Otterson GA, Morrison CD, Smiraglia DJ, Plass C (2006) *Proc Natl Acad Sci USA* 103:982–987.
- Grier DG, Thompson A, Kwasniewska A, McGonigle GJ, Halliday HL, Lappin TR (2005) *J Pathol* 205:154–171.
- Samuel S, Naora H (2005) *Eur J Cancer* 41:2428–2437.
- Golpon HA, Geraci MW, Moore MD, Miller HL, Miller GJ, Tuder RM, Voelkel NF (2001) *Am J Pathol* 158:955–966.
- Okuda H, Toyota M, Ishida W, Furihata M, Tsuchiya M, Kamada M, Tokino T, Shuin T (2006) *Oncogene* 25:1733–1742.
- Shiraishi M, Sekiguchi A, Oates AJ, Terry MJ, Miyamoto Y (2002) *Oncogene* 21:3659–3662.
- Raman V, Martensen SA, Reisman D, Evron E, Odenwald WF, Jaffee E, Marks J, Sukumar S (2000) *Nature* 405:974–978.
- Novak P, Jensen T, Oshiro MM, Wozniak RJ, Nouzova M, Watts GS, Klimecki WT, Kim C, Futscher BW (2006) *Cancer Res* 66:10664–10670.
- Catron KM, Iler N, Abate C (1993) *Mol Cell Biol* 13:2354–2365.
- Williams TM, Williams ME, Innis JW (2005) *Dev Biol* 277:457–471.
- Liu Z, Shi W, Ji X, Sun C, Jee WS, Wu Y, Mao Z, Nagy TR, Li Q, Cao X (2004) *J Biol Chem* 279:11313–11319.
- Saleh M, Rambaldi I, Yang XJ, Featherstone MS (2000) *Mol Cell Biol* 20:8623–8633.
- Shanmugam K, Green NC, Rambaldi I, Saragovi HU, Featherstone MS (1999) *Mol Cell Biol* 19:7577–7588.
- Cao R, Zhang Y (2004) *Mol Cell* 15:57–67.
- Kim SY, Paylor SW, Magnuson T, Schumacher A (2006) *Development (Cambridge, UK)* 133:4957–4968.
- Ehrhardt S, Su IH, Schneider R, Barton S, Bannister AJ, Perez-Burgos L, Jenuwein T, Kouzarides T, Tarakhovskiy A, Surani MA (2003) *Development (Cambridge, UK)* 130:4235–4248.
- Boyer LA, Plath K, Zeitlinger J, Brambrink T, Medeiros LA, Lee TI, Levine SS, Wernig M, Tajonar A, Ray MK, et al. (2006) *Nature* 441:349–353.
- Bracken AP, Dietrich N, Pasini D, Hansen KH, Helin K (2006) *Genes Dev* 20:1123–1136.
- Lee TI, Jenner RG, Boyer LA, Guenther MG, Levine SS, Kumar RM, Chevalier B, Johnstone SE, Cole MF, Isono K, et al. (2006) *Cell* 125:301–313.
- Ding L, Erdmann C, Chinnaiyan AM, Merajver SD, Kleer CG (2006) *Cancer Res* 66:4095–4099.
- Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C, Sanda MG, Ghosh D, Pienta KJ, Sewalt RG, Otte AP, et al. (2002) *Nature* 419:624–629.
- Vire E, Brenner C, Deplus R, Blanchon L, Fraga M, Didelot C, Morey L, Van Eynde A, Bernard D, Vanderwinden JM, et al. (2006) *Nature* 439:871–874.
- Hernandez-Munoz I, Taghavi P, Kuijl C, Neeffjes J, van Lohuizen M (2005) *Mol Cell Biol* 25:11047–11058.
- Reynolds PA, Sigaroudinia M, Zardo G, Wilson MB, Benton GM, Miller CJ, Hong C, Fridlyand J, Costello JF, Tlsty TD (2006) *J Biol Chem* 281:24790–24802.
- Ohm JE, McGarvey KM, Yu X, Cheng L, Schuebel KE, Cope L, Mohammad HP, Chen W, Daniel VC, et al. (2007) *Nat Genet* 39:237–242.
- Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M, Zimmerman J, Eden E, Yakhini Z, Ben-Shushan E, Reubinoff BE, et al. (2007) *Nat Genet* 39:232–236.
- Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G, Marth C, Weisenberger DJ, Campan M, Young J, Jacobs I, Laird PW (2007) *Nat Genet* 39:157–158.
- Guenther MG, Jenner RG, Chevalier B, Nakamura T, Croce CM, Canaani E, Young RA (2005) *Proc Natl Acad Sci USA* 102:8603–8608.
- Popovic R, Zeleznik-Le NJ (2005) *J Cell Biochem* 95:234–242.
- Terranova R, Agherbi H, Boned A, Meresse S, Djabali M (2006) *Proc Natl Acad Sci USA* 103:6629–6634.
- Belinsky SA (2004) *Nat Rev Cancer* 4:707–717.
- Laird PW (2003) *Nat Rev Cancer* 3:253–266.
- Ahrendt SA, Chow JT, Xu LH, Yang SC, Eisenberger CF, Esteller M, Herman JG, Wu L, Decker PA, Jen J, Sidransky D (1999) *J Natl Cancer Inst* 91:332–339.
- Belinsky SA, Nikula KJ, Palmisano WA, Michels R, Saccomanno G, Gabrielson E, Baylin SB, Herman JG (1998) *Proc Natl Acad Sci USA* 95:11891–11896.
- Selzer RR, Richmond TA, Pofahl NJ, Green RD, Eis PS, Nair P, Brothman AR, Stallings RL (2005) *Genes Chromosomes Cancer* 44:305–319.
- Xiong Z, Laird PW (1997) *Nucleic Acids Res* 25:2532–2534.