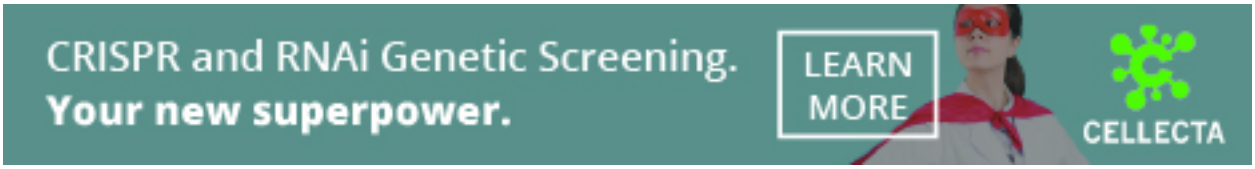
An *Asc*I Boundary Library for the Studies of Genetic and Epigenetic Alterations in CpG Islands

Zunyan Dai, Dieter Weichenhan, Yue-Zhong Wu, et al.

Genome Res. 2002 12: 1591-1598
Access the most recent version at doi:10.1101/gr.197402



Resource

An *Asc*I Boundary Library for the Studies of Genetic and Epigenetic Alterations in CpG Islands

Zunyan Dai,[1,2] Dieter Weichenhan,[4] Yue-Zhong Wu,[1] Julia L Hall,[1] Laura J. Rush,[1,3] Laura T. Smith,[1] Aparna Raval,[1] Li Yu,[1] Daniela Kroll,[5] Joerg Muehlisch,[5] Michael C. Frühwald,[5] Pieter de Jong,[6] Joe Catanese,[6] Ramana V. Davuluri,[1] Dominic J. Smiraglia,[1] and Christoph Plass[1,7]

[1]Division of Human Cancer Genetics, Department of Molecular Virology, Immunology and Medical Genetics, [2]Department of Pathology, [3]Department of Veterinary Biosciences, and the Comprehensive Cancer Center, The Ohio State University, Columbus, Ohio 43210, USA; [4]Medizinische Universität zu Lübeck, Medizinische Klinik II, Ratzeburger Allee 160 23538 Lübeck, Germany; [5]Klinik und Poliklinik für Kinderheilkunde, Pädiatrische Hämatologie/Onkologie, Universitätsklinikum Münster, 48149 Münster, Germany; [6]Children's Hospital, Oakland Research Institute, Oakland, California 94609, USA

Knudson's two-hit hypothesis postulates that genetic alterations in both alleles are required for the inactivation of tumor-suppressor genes. Genetic alterations include small or large deletions and mutations. Over the past years, it has become clear that epigenetic alterations such as DNA methylation are additional mechanisms for gene silencing. Restriction Landmark Genomic Scanning (RLGS) is a two-dimensional gel electrophoresis that assesses the methylation status of thousands of CpG islands. RLGS has been applied successfully to scan cancer genomes for aberrant DNA methylation patterns. So far, the majority of this work was done using *Not*I as the restriction landmark site. Here, we describe the development of RLGS using *Asc*I as the restriction landmark site for genome-wide scans of cancer genomes. The availability of *Asc*I as a restriction landmark for RLGS allows for scanning almost twice as many CpG islands in the human genome compared with using *Not*I only. We describe the development of an *Asc*I–*Eco*RV boundary library that supports the cloning of novel methylated genes. Feasibility of this system is shown in three tumor types, medulloblastomas, lung cancers, and head and neck cancers. We report the cloning of 178 *Asc*I RLGS fragments via two methods by use of this library.

[Supplemental material is available online at http://www.genome.org.]

Multiple genome scanning approaches have been developed in the past years to study genetic and epigenetic alterations in cancer (Gray and Collins 2000). The majority of those techniques target genetic alterations such as deletions, insertions, and copy number changes. Restriction Landmark Genomic Scanning (RLGS), a highly reproducible two-dimensional gel electrophoresis, allows scanning of genomes for DNA polymorphisms, DNA amplification, and DNA methylation (Hayashizaki et al. 1994a,b; Plass et al. 1996; Costello et al. 2000). The use of RLGS to study human cancers resulted in the identification of several novel genes that were amplified and overexpressed in malignant tissues (Costello et al. 1997; Frühwald et al. 2000). Furthermore, the use of methylation-sensitive restriction enzymes as landmark enzymes makes scanning of genomes for changes in the DNA methylation patterns possible (Dai et al. 2001; Frühwald et al. 2001b; Rush et al. 2001; Rush and Plass 2002). This is of particular interest in cancer genetics, because promoter methylation has been shown to be involved in the silencing of tumor suppressor genes (Jones and Laird 1999; Baylin et al. 2001; Costello and Plass 2001). The methylation-sensitive restriction landmark enzyme *Not*I has a GC-rich recognition sequence, which is preferentially located in CpG islands sequences, found mainly in promoter regions of genes (Costello et al. 2000). In normal tissue DNAs, these sites are unmethylated (Bird 1986). However, in tumors, methylation of a *Not*I site results in the absence of an RLGS fragment in the respective profile.

Although RLGS profiles can be generated from any high-quality genomic DNA without prior sequence information, subsequent cloning of RLGS fragments is essential for future studies. Several PCR-based protocols have been developed allowing the identification of RLGS sequences (Ohsumi et al. 1995). More efficient, however, is a cloning strategy that uses an arrayed human library of *Not*I–*Eco*RV clones and RLGS mixing gel catalogs (Smiraglia et al. 1999). This protocol circumvents the need for PCR-based amplification, which could be problematic with GC-rich sequences. Successful use of this library system resulted in the identification of many methylation targets in several human tumors (Smiraglia and Plass 2002).

The use of the *Not*I–*Eco*RV boundary library as a cloning tool for RLGS is restricted to RLGS profiles that use the enzyme combination *Not*I and *Eco*RV as the first and second restriction enzymes. To increase the potential coverage of CpG islands, we developed reaction conditions for the use of *Asc*I as the restriction landmark enzyme in RLGS. In addition, we prepared an *Asc*I–*Eco*RV library and RLGS mixing gels that allow the efficient recovery of cloned RLGS fragments. We estimate that this novel resource, together with the *Not*I–*Eco*RV library, will greatly increase the utility of RLGS and, in

[7]Corresponding author.
E-MAIL Plass-1@medctr.osu.edu; FAX (614) 688-4761.
Article and publication are at http://www.genome.org/cgi/doi/10.1101/gr.197402.

addition, provide access to up to 15,000 of the estimated 29,000 CpG islands in the human genome (Venter et al. 2001).

RESULTS AND DISCUSSION

*Asc*I and *Not*I Restriction Sites Predicted in the Human Genome

The majority of RLGS gels generated to study DNA methylation profiles in human malignancies have used *Not*I as the restriction landmark enzyme. Many of these studies were supported by a *Not*I–*Eco*RV library (Smiraglia et al. 1999). To develop an additional landmark enzyme for the purpose of CpG island identification by RLGS, we analyzed the human genome sequence for the frequency and location of restriction sites for rare cutting, methylation-sensitive restriction endonucleases. *Asc*I is a restriction enzyme that recognizes the target sequence GG**CGCG**CC and does not cut the methylated recognition sequence. Some characteristics of the loci cut by *Asc*I obtained from the human genomic sequence (August 6, 2001 draft assembly of UCSC) are listed in Table 1 and compared with those cut by *Not*I. Surprisingly, the human genome possesses only half the number of *Asc*I sites (4935) as compared with *Not*I sites (9628), although both recognition sequences are composed of four guanines and four cytosines each, and both contain 2 CG dinucleotides. Nevertheless, *Not*I and *Asc*I are highly comparable in terms of the types of loci they assess. Of particular note is the fact that 86% and 83% of these sites, respectively, are found in CpG islands, whereas only 5% and 7% are found in repetitive elements not associated with CpG islands. This strong bias of representation of CpG islands over repetitive elements is a major strength of RLGS using these two enzymes. Furthermore, 86% and 83% of these CpG islands, respectively, are associated with known genes or ESTs. These data indicate that *Asc*I is an excellent choice of landmark enzyme to complement RLGS studies performed using *Not*I. In addition, because *Not*I and *Asc*I sites colocalize in only 3.7% of CpG islands, by using *Asc*I as a second landmark enzyme, we are able to almost double the number of CpG islands whose methylation status can be analyzed.

RLGS Profiles Using *Asc*I as the Landmark Enzyme Display up to 2000 Distinct CpG Islands

We established the reaction conditions for the use of *Asc*I as a restriction landmark site (see Fig. 1 for an outline of the procedure). RLGS profiles with *Asc*I show a lower fragment density than *Not*I profiles (Fig. 2A), as expected from the genome sequence survey that identified fewer *Asc*I restriction sites in the human genome (Table 1). An *Asc*I master profile was prepared using total genomic DNA from three donors to maximize coverage of polymorphic spots. The master profile was labeled with a coordinate system of spot numbering (a portion is shown in Fig. 2D) as was done for the *Not*I master profile described previously (Costello et al. 2000). The lower density of RLGS fragments in an *Asc*I profile allows the scoring of more fragments in the higher molecular weight sections. These sections are difficult to score in the *Not*I profiles due to the high density of spots and are frequently excluded from the analysis. Thus, although the number of fragments on an *Asc*I profile is less than on a *Not*I profile, a similar number of ~2500 fragments can be analyzed on both.

Because *Asc*I is methylation sensitive, we compared methylation frequencies detected by *Not*I and *Asc*I in the same samples to determine whether both recognition sequences are equal targets for aberrant methylation in human malignancies. Table 2 summarizes the data obtained for nine lung cancers, six medulloblastomas, and three head and neck cancers. The number of methylated sequences detected with both restriction enzymes is not statistically different ($P \leq 0.05$). These data indicate that although these enzymes assess different loci, they are similar in their abilities to detect aberrant methylation in human malignancies.

Establishment and Initial Characterization of an *Asc*I–*Eco*RV Library

The initial step in the construction of the *Asc*I–*Eco*RV boundary library was the purification of *Asc*I–*Eco*RV fragments from total genomic DNA using the *Bss*HII/*Asc*I restriction trapper. This procedure results in the enrichment of *Asc*I–*Eco*RV fragments and eliminates *Eco*RV–*Eco*RV fragments (see Methods for details). The quality of the purified *Asc*I–*Eco*RV fragments was tested by using an aliquot of these fragments for RLGS separation. A portion of the purified *Asc*I–*Eco*RV fragments was labeled and subjected to two-dimensional separation in the RLGS system. The resulting RLGS profile showed the same set of fragments as the original profile without prior purification (Fig. 2A,B), indicating that the purification did not result in loss or gain of certain fragments. The remaining purified material was used for cloning into pBluescript KS-*Asc*I. The *Asc*I–*Eco*RV library (A-RV-1) consists of 19,200 clones picked into fifty 384 well plates. The average insert size was 2.48 kb ($n = 75$) ranging from 0.3 to 10 kb. Accordingly, the library has an expected bias toward smaller fragments, reflecting the

Table 1. Characteristics of *Not*I and *Asc*I Restriction Sites

	*Not*I	*Asc*I
Total number of restriction sites in the human genome[a]	9628	4935
Number of restriction sites in CpG islands	8239 (86%)	4071 (83%)
Number of restriction sites in repetitive elements, not CpG islands	520 (5%)	332 (7%)
Frequency of restriction sites near 5′ end of a known gene	3357 (34.9%)	1612 (32.7%)
Frequency of restriction sites near 3′ end of a known gene	1328 (13.8%)	725 (15%)
Frequency of restriction sites inside a gene	1392 (14.5%)	738 (15%)
Frequency of restriction sites falling near ESTs	2.221 (23.1%)	1001 (20.3%)
Frequency of CpG islands with both *Not*I and *Asc*I sites	1100 (3.7%)	

[a]August 6, 2001 draft assembly of UCSC.

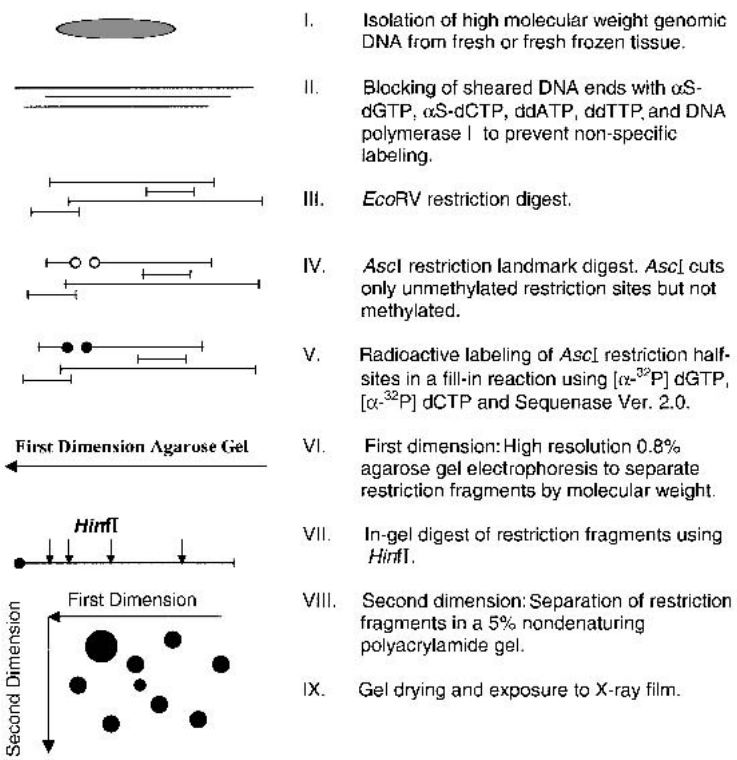


Figure 1 Outline of the RLGS procedure using *Asc*I as a restriction landmark enzyme.

cloning bias of the plasmid vector. Clones from 48 plates were spotted onto filters for hybridization-based screening, providing an additional resource for studies of CpG islands. Each filter contains the entire set of clones from the 48 plates spotted in duplicate. The availability of these filters allows for rapid identification of plasmid clones with 5′ end sequences for known genes. In addition, these clones provide a unique resource for array-based studies.

*Asc*I–*Eco*RV Library Clone Sequence Characteristics are Similar to Predicted

CpG islands are mainly located in the promoter region of genes and are less frequently found in the body or 3′ end of genes. The survey of the human genome for *Asc*I sites described above indicated that the recognition sequence of *Asc*I (GG**CGCG**CC) has a preferential localization to CpG islands. To determine whether our library has a similar representation, we sequenced 178 *Asc*I–*Eco*RV fragments cloned from this library by the two methods described below. A total of 158 sequences (89%) showed CpG island features (see Methods). We mapped all 178 sequences to the human genome draft sequence (August 6, 2001 draft assembly of UCSC) and found that 137 (77%) mapped to known genes or ESTs. In 84 cases in which the CpG island could be mapped within the context of a known gene, 66 (79%) were found in the 5′ end of a gene, 12 (14%) in the body, and 6 (7%) in the 3′ end (Table 3). This further supports the assumption that *Asc*I sites are preferentially located in CpG islands near genes and assures that our library is a faithful representation of this.

Establishment of Mixing Gels as a Cloning Tool for RLGS Fragments

To generate a novel tool that will aid the cloning of RLGS fragments from profiles generated with *Asc*I as the landmark enzyme and *Eco*RV as the second restriction enzyme, we prepared RLGS mixing gels from plates 1 to 32 of the A-RV-1 library. In addition, the rows and columns from these 32 plates were individually pooled to produce 16-row pool (A–P) and 24 column pool (1–24) mixing gels. The procedure followed the strategy used previously for the generation of the *Not*I–*Eco*RV mixing gel catalog (Smiraglia et al. 1999). In RLGS mixing gels, fragments for which a corresponding clone is present in the pool of clones mixed with the genomic DNA will show enhancement. Determination of the plate, row, and column mixing gels in which the RLGS spot of interest is enhanced indicates the unique library address in which the corresponding RLGS fragment is cloned. An example of an RLGS mixing gel is shown in Figure 2C, with clones from a 384-well plate of the library. The average number of enhanced RLGS fragments per plate is 153, as many of the clone insert sizes fall outside of the window of resolution of a standard RLGS profile. In the 32-plate mixing gels, there are 1468 unique RLGS fragments represented.

Use of the *Asc*I–*Eco*RV Cloning Gel Catalog to Identify Hypermethylated Sequences in Various Cancers

*Asc*I is preferentially located within CpG islands and is methylation sensitive. Thus, *Asc*I is useful as a restriction landmark enzyme in RLGS studies to identify methylation changes in two different samples. We used the methylation scanning properties of *Asc*I to determine methylation changes in medulloblastoma (MB), lung cancer, and head and neck cancer (HNSCC) primary tumors relative to adjacent normal tissue, as well as cell lines representing all three tumor types and a leukemia cell line. RLGS fragment loss in tumor profiles and cell line profiles is the most prominent observation and is indicative of hypermethylation of those fragments. Less frequently, newly appearing RLGS spots are found on the tumor profiles that may represent hypomethylation. Unfortunately, however, such rare RLGS spots cannot be cloned using this library, because they are not present in the RLGS profiles of the DNAs used to create the library. Figure 3 shows an example of an RLGS fragment (RLGS fragment A2E54; in which A indicates the *Asc*I profile and 2E54 indicates spot No. 54 in section 2E) that is present in normal adjacent lung tissue DNA, but absent from the lung tumor and two lung cancer cell lines. The corresponding library clone was identified in the plate 4, row I, and column 14 mixing gels. Clone 4I14 was isolated from the library and used in a single-clone mixing gel to confirm that the clone represents the intended RLGS spot (Fig. 3A). Insert DNA from this clone was used as a hybridization probe for Southern blot analysis to confirm methylation of the *Asc*I site. Lung tumor and normal genomic DNA was digested with both *Eco*RV and *Asc*I. Control DNA in lane 1 of Figure 3B was digested with *Eco*RV only. In the Southern analysis, the probe detects either a small (*Asc*I–*Eco*RV) fragment or a larger (*Eco*RV–*Eco*RV) fragment. The presence of the large fragment indicates that the

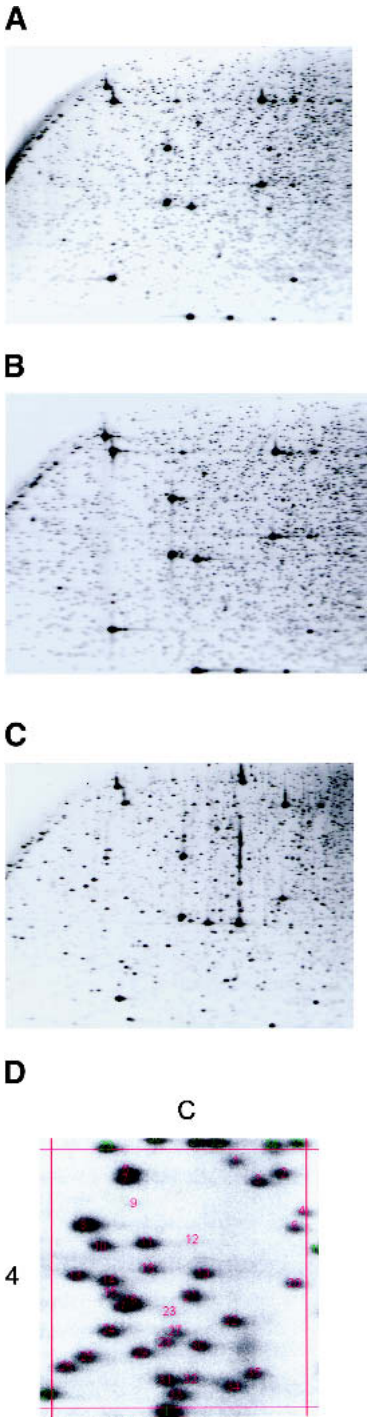


Figure 2 RLGS profiles using *Asc*I–*Eco*RV–*Hin*fI restriction enzyme combination. (*A*) RLGS profile of normal lung DNA. (*B*) RLGS profile using restriction trapper purified *Asc*I–*Eco*RV fragments derived from peripheral blood lymphocyte DNAs. (*C*) RLGS mixing gel generated with normal lung DNA as the genomic background and clones from plate 3 pool in the A-RV1 library. (*D*) Section 4C of the *Asc*I Master RLGS profile, showing the numbers assigned to each *Asc*I fragment.

*Asc*I site was protected from restriction digestion by DNA methylation. The Southern data for patient 14 confirms the RLGS result. Similarly, 17 additional RLGS fragments have been cloned using this targeted approach and are shown in bold in Table 3.

Large-Scale Identification of RLGS Fragment Sequences

A second nontargeted strategy to identify the sequences of RLGS fragments in *Asc*I gels was also used. *Asc*I–*Eco*RV clones derived from the A-RV-1 were sequenced from the *Asc*I end. This sequence was used to determine the full-length *Asc*I–*Eco*RV and the *Asc*I–*Hin*fI restriction fragment sizes from the August 2001 freeze of the human genome. DNAs of clones with insert sizes of 0.5–5 kb and an *Asc*I–*Hin*fI fragment >100 bp were pooled into groups of up to 15 plasmid DNAs and used for RLGS mixing gels. These mixing gels resulted in the enhancement of exactly the same number of RLGS fragments as the number of plasmid clones in the pool. Because the sequence, and thus the predicted mobility of these clones in RLGS gels was known, it was possible to unambiguously identify which clone and sequence corresponds with each RLGS fragment. By use of this strategy, it was possible to identify 160 additional RLGS fragments and their sequences. This data is summarized in Table 3, which is sorted by methylation status and chromosomal location. Of the total number (178) of *Asc*I fragments cloned by the two methods described, we found 70 that were methylated in primary tumors and 119 methylated in cancer cell lines (Table 3; see supplemental data online for the complete list of cloned RLGS fragments).

Conclusion

We have developed a valuable resource for isolating and studying CpG-rich regulatory human sequences using *Asc*I as a restriction landmark enzyme. *Asc*I is as suitable as *Not*I to determine methylation patterns in human malignancies and nearly doubles the set of loci that can be studied. The *Asc*I recognition sequence occurs less frequently in the genome than the *Not*I sequence, but its location is similarly biased toward CpG islands. The cloning gel catalogs now available for *Not*I–*Eco*RV and *Asc*I–*Eco*RV allow for the targeted cloning of up to 3257 RLGS fragments [1789 from *Not*I (Smiraglia et al. 1999), 1468 from *Asc*I], >90% of which are expected to represent CpG islands (Smiraglia and Plass 2002). In addition, we have developed a nontargeted, but higher throughput strategy for using these libraries to clone RLGS fragments.

By applying these cloning strategies, we have created a novel resource, which when used in conjunction with RLGS analysis of tumor profiles, allows for the identification of large numbers of methylation targets. Even in this limited study of 18 analyzed primary tumor profiles, we have already identified 70 targets of hypermethylation in cancer. Three of the genes that were identified, *HOXA11*, *NELL1*, and *ALX3* have been identified previously by others as methylation targets in lung adenocarcinomas (Shiraishi et al. 2002a,b) or neuroblastomas (Wimmer et al. 2002), respectively. As we increase the number of tumor profiles analyzed and go through multiple iterations of the cloning strategies describe in this article, we will significantly increase the number of targets of hypermethylation that we can identify. This is a requisite step to begin to understand the mechanisms and consequences of such hypermethylation.

Together, there are a little over 15,000 *Not*I and *Asc*I restriction sites in the human genome. Considering the total number of 29,000 CpG islands in the human genome (Venter et al. 2001), these libraries provide access to nearly half of the

Table 2. Methylation Frequencies in Various Tumor Samples Determined by Either *Not*I or *Asc*I as a Restriction Enzyme

	*Not*I RLGS gels			*Asc*I RLGS gels			
Tumor type[a]	No. of methylated CpG islands	No. of RLGS fragments analyzed	Methylation frequency in *Not*I gels	No. of methylated CpG islands	No. of RLGS fragments analyzed	Methylation frequency in *Asc*I gels	Z-test statistic for proportions[b]
Lung	63	1184	5.3%	78	1614	4.8%	0.5833
Lung	59	1184	5.0%	62	1614	3.8%	1.4669
Lung	28	1184	2.4%	49	1690	2.9%	−0.8735
Lung	8	1184	0.7%	8	1008	0.8%	−0.3234
Lung	7	1184	0.6%	13	1205	1.0%	−1.3079
Lung	5	1184	0.3%	18	1734	1.0%	−1.8470
Lung	3	1184	0.3%	7	1734	0.4%	−0.6822
Lung	1	1184	0.1%	3	1380	0.2%	−0.8503
Lung	0	1184	0.0%	5	1614	0.3%	−1.9169
MB	53	1702	3.1%	20	923	2.1%	1.4091
MB	32	1768	1.8%	22	1246	1.8%	0.0903
MB	31	1825	1.7%	21	1093	1.9%	−0.4401
MB	32	2016	1.6%	19	1172	1.6%	−0.0735
MB	15	1741	0.9%	19	1327	1.4%	−1.4947
MB	14	2018	0.7%	9	1421	0.6%	0.2140
HNSCC	13	1703	0.8%	14	1739	0.8%	−0.1387
HNSCC	3	1839	0.2%	6	1009	0.6%	−1.9625
HNSCC	0	2126	0.0%	1	1243	0.1%	−1.3080

[a]HNSCC, head and neck squamous cell carcinomas; MB, medulloblastoma.
[b]Z-test static value for testing the significant difference in methylation frequencies in *Not*I and *Asc*I gels. All of the Z-static values are between −1.96 to +1.96, which suggest that there is no significant ($p \leq 0.05$) difference in methylation frequencies in *Not*I and *Asc*I gels.

CpG islands. Therefore, these libraries will prove to be excellent tools for the study of aberrant CpG island methylation when used in combination with various methylation-scanning techniques such as RLGS and differential methylation hybridization (Frühwald and Plass 2002; Smiraglia and Plass 2002). Although standard RLGS running conditions only resolve a set of ~2500 CpG islands with a first dimension size of 5 kb–500 bp, these conditions can be altered to resolve a similar number of fragments with first dimension size ranging from 10 to 5 kb (Hughes et al. 1998). Thus, by modifying RLGS electrophoresis conditions and by utilizing other technologies that do not require electrophoresis, the full potential of these libraries may be achieved.

METHODS

Tissue Samples and Cell Lines

Frozen non-small cell lung tumors paired with normal adjacent tissues were collected through the Cooperative Human Tissue Network (CHTN). Nine paired samples (patient nos. 2, 3, 5, 7, 10, 11, 13, 14, and 17) and clinical characteristics were described previously (Dai et al. 2001). Six medulloblastoma samples were described previously (Frühwald et al. 2001b). Three head and neck cancer tissues were collected at The Ohio State University through the CHTN. All sample collection was performed in accordance with NIH guidelines. Non-small cell-line lung cancer lines A549 (from ATCC), H125, H1299, and H2086, head and neck cancer cell line SCC-9, leukemia cell lines HL-60, ML-1, and K-562, medulloblastoma cell lines Daoy, D425 MED, MHH-MED-1, and MHH-PNET-5, used in this study were described previously (Dai et al. 2001; Frühwald et al. 2001a; Rush et al. 2001, 2002; Smiraglia et al. 2001).

Isolation of Plasmid and Genomic DNAs

High molecular weight DNA for the RLGS procedure was isolated according to our previously published protocol (Smiraglia et al. 1999). Plasmid DNA was isolated using QIAprep Spin Miniprep kit (QIAGEN) and the manufacturer's recommended protocols.

RLGS

RLGS was performed according to published protocols (Okazaki et al. 1994) with modifications for the use of *Asc*I as the restriction landmark enzyme. Briefly, to prevent nonspecific labeling, the sheared ends of ~7 µg of genomic DNA were blocked in a 10-µL reaction by the addition of nucleotide analogs (αS-dGTP, αS-dCTP, ddATP, ddTTP) using 2.5 U of DNA polymerase I (Boehringer Mannheim) (37°C, 20 min) followed by enzyme inactivation (65°C, 30 min). The DNA was digested (37°C, 2 h) with 20 U of *Eco*RV (New England Biolabs), followed by 20 U of *Asc*I (New England Biolabs) in NEB buffer 4 (37°C, 2 h). The resulting restriction sites from *Asc*I were labeled in a fill-in reaction using Sequenase Ver. 2.0 (USB) in the presence of [α-^{32}P]dGTP (6000 Ci/mmole, NEN Life Science Products) and [α-^{32}P]dCTP (3000 Ci/mmole, NEN) for 30 min and stopped by adding buffer that included dCTP and dGTP. A portion of the reaction was electrophoresed through a 60-cm long, 0.8 % agarose tube gel (first dimension separation). The agarose gel was equilibrated in restriction buffer and the DNA was digested in the gel with 750 U of *Hin*fI (New England Biolabs) at 37°C for 2 h. The agarose gel was placed horizontally across the top of a nondenaturing 5% polyacrylamide gel, the two gels were connected with molten agarose, and the DNA was electrophoresed in the second dimension. The gels were dried and exposed to Kodak X-OMAT AR film in the presence of one intensifying screen (Quanta 111, DuPont) for 2–10 d.

*Asc*I Restriction Trapper Purification

A mix of 500 µg of total human genomic DNA from three donors was digested with 1500 U of *Asc*I (37°C for 3 h), and subsequently with 500 U of *Eco*RV (37°C overnight), extracted with phenol/chloroform/isoamyl alcohol (PCI), precipitated, and resuspended in H_2O at a concentration of 2 mg/mL. Aliquots of 100 µg of restriction-digested DNA were ligated in

Table 3. Cloned *Asc*I Spots (See Supplemental Data Online for a Complete List of Cloned RGLS Fragments[f])

RLGS spot	Chr. position[a]	Chr. band	CpG island[b]	Methylation[c]		Gene or EST homology[d]	Context[e]
				Primary	Cell lines		
A2D16[g]	chrNA_rdm:9890705-9892163	?	Y	L	L HN LU MB	DRD4/PTDSS2	5′end/Body
A5E46	chrX:17627774-17629763	Xp22.13	Y	L	L HN LU	–	–
A4G14	chr1:256787415-256787746	1q41	Y	L	L LU	–	–
A2D08	chr1:68251887-68253584	1p32.1	Y	L	LU	JUN	Body
A3F22	chr1:108979505-108982150	1p22.1	Y	L	L	–	–
A3F38	chr1:201154777-201155644	1q24.2	N	L HN MB	L LU	PMX1	5′ end
A6E17	chr1:133792275-133794522	1p13.2	Y	L	L LU	–	–
A3F30	chr1:8129740-8132794	1p36.23	N	L	LU	EST	–
A5E43	chr1:127726863-127728875	1p13.3	Y	L HN MB	L HN LU MB	ALX3	5′ end
A4E32	chr1:22646645-22648732	1p36.32	Y	L	L LU	–	–
A3H43	chr1:2844351-2849754	1p36.32	Y	L	L HN LU	–	–
A2F45	chr2:117320895-117322735	2q13	N	L HN MB	L HN LU MB	EST	–
A3E38	chr2:179015764-179017423	2q31.1	N	L	L	–	–
A4D36	chr2:138670622-138671469	2q21.2	N	L	L LU	–	–
A4G44	chr3:220718143-220722411	3q29	Y	L	HN LU	EST	–
A5F32	chr3:153653591-153655489	3q22.1	N	L MB	L LU MB	–	–
A5G07	chr3:8664656-8665380	3p26.1	Y	L HN	L HN LU	GRM7	5′ end
A5E28	chr3:3902165-3903025	3p26.1	Y	L	L HN LU	–	–
A4B07	chr3:146695171-146695918	3q21.3	Y	L	HN LU	EST	–
A2E14	chr4:5349208-5351766	4p16.1	Y	L MB	L LU	HSA250839	5′ end
A2D23	chr4:171849449-171850978	4q32.1	Y	L HN	L HN LU	GLRB	5′ end
A3E31	chr4:181006489-181008945	4q32.3	Y	L	L LU	TLL1	5′ end
A6E16	chr5:49418697-49420952	5p11	Y	L	L LU	EST	–
A4F02	chr5:187652008-187654787	5q35.1	Y	L		EST	–
A6D30	chr5:195984026-195984768	5q35.2	Y	MB	MB	B4GALT7	5′ end
A3G33	chr5:83731413-83731663	5q13.3	Y	L	L HN LU	MGC15435	3′ end
A4C25	chr6:103613999-103614941	6q15	Y	L	LU	BACH2	5′ end
A3E39	chr6:86748700-86751107	6q14.1	Y	L HN	L HN LU	EST	–

[a]BLAT search results based on http://genome.ucsc.edu/cgi-bin/hgBlat, Aug. 2001 Freeze.
[b]Y indicates that the *Asc*I site is within a CpG island, N indicates that it is not.
[c]Indicates methylation found in a primary tumor or cell line as indicated: (L) Lung carcinoma; (MB) medulloblastoma; (HN) head and neck squamous cell carcinoma; (LU) leukemia.
[d]Known genes found in the Refseq database or spliced ESTs.
[e]5′ end indicates that the CpG island includes the region immediately upstream of exon 1 and/or exon 1. The 3′ end indicates that the CpG island is found in 3′ most exon. Body indicates that the CpG island is found within the genomic structure of the gene excluding the 5′-most and 3′-most known exons.
[f]Genes share same CpG island in their 5′ end and are transcribed in opposite directions.
[g]Bold indicates the RLGS fragments that were cloned directly using the mixing gels.

150-μL volume to a 0.67% (w/v) DNA Trapper R–*Bss*HII (Japan Synthetic Rubber Co.) in the presence of 10% PEG 6000 using 1400 U of T4 DNA ligase (New England Biolabs) at 18°C overnight. The DNA trapper-ligated DNA was digested twice with 100 U of *Eco*RV, centrifuged to remove nonligated *Eco*RV fragments, and then digested with 100 U of *Asc*I to release *Asc*I–*Eco*RV fragments. DNA fragments were PCI purified, precipitated in the presence of glycogen (Boehringer Mannheim), and dissolved in 13 μL of TE buffer. A total of 11.8 μg of purified DNA was recovered. To determine the quality and purity of the *Asc*I–*Eco*RV fragments, 1 μg was endlabeled with a fill-in reaction using Sequenase Ver. 2.0 in the presence of [α-^{32}P]dGTP (6000 Ci/mmole, DuPont) and [α-^{32}P]dCTP (3000 Ci/mmole, DuPont) and subjected to the two-dimensional separation in the RLGS system. The resulting profile was compared with the RLGS profile prepared from total genomic DNA.

Construction of Vector KSII$^+$–*Asc*I

To insert an *Asc*I site (GGCGCGCC) into vector Bluescript KSII$^+$ (Stratagene), 50 pmole of each primer *Asc*I-1 (CCACCGCGGTGGGCGCGCCT) and *Asc*I-2 (CTAGAGGCGCGCCCACCGCGGTGGAGCT) (custom made by MWG Biotech) were annealed and subsequently ligated with 100 ng of *Sac*I–*Xba*I cut vector DNA. Appropriate insertion of the annealed primers would not disturb the ORF of the multiple cloning site and, hence, the vector's capability for blue/white selection on agar medium containing X-GAL. *Escherichia coli* DH10B (Life Technologies) were transformed with the ligation mixture and plated onto LB agar containing ampicillin, IPTG, and X-GAL. Blue colonies were tested for the presence of KSII$^+$ harboring an *Asc*I site. One of such plasmids, designated KSII$^+$–AscI, was selected for subsequent library construction.

Library Construction

To facilitate reliable double digestions of vector KSII$^+$–*Asc*I with *Asc*I plus *Eco*V, we first shotgun cloned *Asc*I–*Eco*RV genomic fragments of mouse DNA into the vector. A clone with a 1.6-kb *Asc*I–*Eco*RV insert was then used to prepare the vector for library construction. Two micrograms of the recombinant plasmid were *Asc*I–*Eco*RV digested and separated on a gel. The vector band was sliced out and run a second time on a gel to improve purity. The band was eluted and dissolved in H_2O at a concentration of 10 ng/μL. Self ligation of 10 ng and subsequent electroporation of electrocompetent *E. coli* DH10B cells (Life Technologies; transformation efficiency ~7–× 10^9 transformants/μg pUC19) yielded in 45 clones. This figure

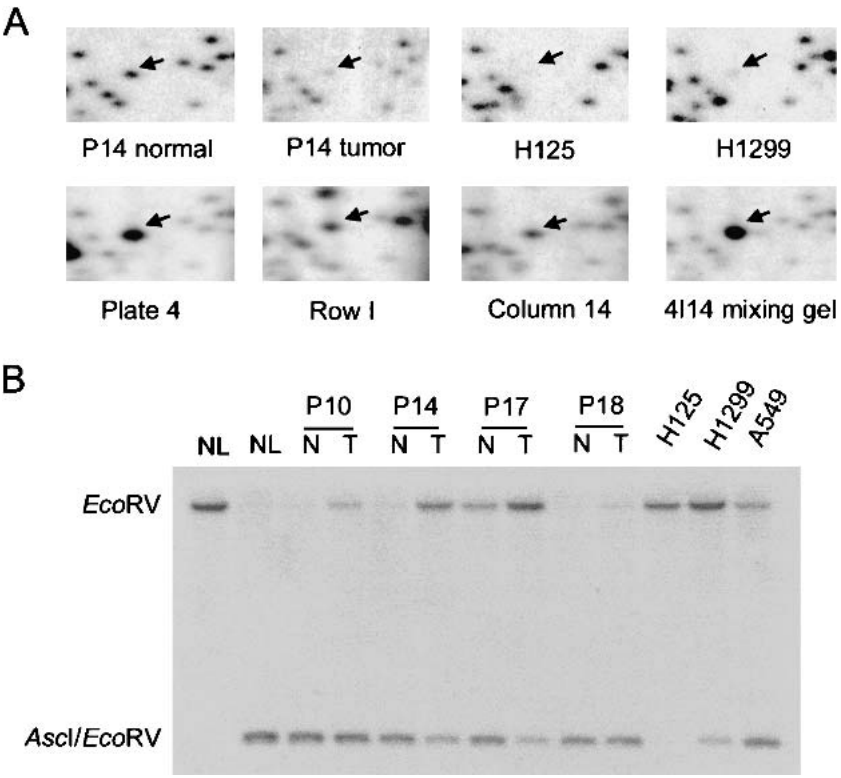


Figure 3 RLGS identifies DNA methylation in primary lung cancer. (*A*) Sections from RLGS profiles including RLGS fragment A2E54 (arrow). Sections from normal and tumor profiles from patient 14 as well as two lung cancer cell lines (H1299 and H125) are shown. The corresponding *Asc*I–*Eco*RV clone was found in plate 4, row I, and column 14, and this clone was confirmed by use in a mixing gel. (*B*) DNA from *Asc*I clone 4I14 corresponding to RLGS spot A2E54 was used for Southern analysis. DNAs from normal lung (NL), lung tumors (T), and adjacent normal tissue (N) from patients 10, 14, 17, and 18, as well as from three lung cancer cell lines H125, H1299, and A549 were digested with *Asc*I and *Eco*RV. DNA in the first lane was digested only with *Eco*RV and shows the size of the *Eco*RV fragment. In the double digests, hybridization to the large *Eco*RV band is indicative of protection of the *Asc*I site digestion by methylation. The smaller band is indicative of cutting by *Asc*I.

indicated the expected nonrecombinants when 10 ng of vector DNA were ligated with insert DNA at similar conditions. For library construction, two 10-µL ligation mixtures, each containing 10 ng of vector DNA, 3 µL of human restriction trapper purified DNA and 0.5 U of T4 DNA ligase (Roche Diagnostics) were incubated at 16°C for 16 h. After addition of 2.5 M NH_4-acetate (final concentration) and 1 µL of glycogen (stock: 20 mg/mL; Roche Diagnostics) as carrier, the DNA was precipitated and redissolved in a total of 5 µL of 0.5× TE. A total of 1 µL of ligated DNA was used per transformation.

Library Picking, Replication, and Preparation of High-Density Hybridization Filters

Transformed cells were spread onto LB/Agar plates containing ampicillin (100 µg/mL), IPTG, and X-gal, and grown at 37°C for 18–20 h. Clones were picked manually and arrayed into 384-well microtiter plates containing LB/ampicillin (50 µg/mL)/glycerol (7.5%). The arrayed clones were incubated at 37°C for 18 h, and then frozen at −80°C. Five additional copies of each plate were made using a 384-pin replicating tool (V & P Scientific) for inoculation. The replicas were grown for 18 h at 37°C and then frozen at −80°C. High-density hybridization filters were prepared using a Q-Bot colony picker/high-density filter gridder. All clones from plate 1 to 48 were used to spot onto three 22.25 × 22.25-cm nylon membranes by use of protocols identical to the one used for BAC clones (Osoegawa et al. 2000).

RLGS Mixing Gels With Clones From the *Asc*I–*Eco*RV Library (A-RV-I)

Plates 1 to 32 from A-RV-1 were chosen for the RLGS mixing gels. Clone pool DNAs for each of the 32 plates, all 16 rows (A–P) and 24 columns (1–24) were prepared as described earlier (Smiraglia et al. 1999). Individual clones were grown in microtiter plates, overnight cultures were combined, and plasmid DNAs for each pool of clones were isolated using spin columns (QIAGEN). Genomic DNA from normal lung was labeled by a fill-in reaction using Sequenase Ver. 2.0 (USB) in the presence of [α-^{32}P]dCTP (6000 Ci/mmole, NEN) and [α-^{32}P]dGTP (3000 Ci/mmole, NEN) for 30 min. Pooled clone DNA was digested by *Eco*RV (Promega) and *Asc*I (NEB) sequentially and labeled following the same procedure for standard RLGS. Ten picogram DNA per clone of labeled pooled clone DNA was mixed with the appropriate amount of labeled genomic DNA and loaded on the first dimension RLGS agarose gel followed by the standard RLGS procedure. The amount of labeled genomic DNA was optimized to obtain a 4-day exposure of the RLGS gel on X-ray film.

Sequencing and Database Analysis

All sequence analyses were performed in the Core Facility of the Division of Human Cancer Genetics using an ABI PRISM 377 DNA sequencer. For CG-rich sequences high-annealing temperatures were employed using an ABI PRISM BigDye Terminator Cycle Sequencing kit. *Asc*I–*Eco*RV clones were sequenced with M13 forward primer. DNA sequence files were analyzed using DNAstar and Chromas software. For homology searches, sequences were submitted to the publicly available databases.

Bioinformatics

The standard two-sided Z-test was used to compare the methylation frequencies in *Not*I and *Asc*I test.

We downloaded the assembled sequences (August 6, 2001 draft assembly of UCSC) of the 24 chromosomes from the UCSC Human Genome Project working draft (http://genome.ucsc.edu). We scanned each of the chromosomes for *Not*I (GCGGCCGC) and *Asc*I (GGCGCGCC) sites, and retrieved the sequences that contain these sites. Each sequence is of a 1008-bp length (−500 to +500 of the site). We used a sliding window 201 bp in length, and counted the percentage of CpG dinucleotides (CpG score) and GC% for each window. The sequence is considered a CpG island if there exists a sliding window with CpG score ⩾ 60% and GC% ⩾ 50. We retrieved all of the 30,095 CpG islands mapped in the human genome and counted the number of CpG islands that have *Not*I and *Asc*I sites. To determine whether these sites fall in a gene region or not, we used the public human genome annotations available at UCSC genome server. We counted the number of sites that fall in and around (within the 5-kb region of the annotated gene ends) known genes and ESTs.

ACKNOWLEDGMENTS

We thank Barbara Swiatkiewcz for her help with the clone spotting. The authors thank Drs. Hayashizaki and Okazaki for the generous gift of the restriction trapper. This work was supported in part by grants P30 CA16058, RO1 CA93548, and R01 DE13123 (CP). L.J.R. was supported by a T32 training grant CA09338 (PI, Michael Caligiuri) and grant CA089317 from NCI. M.F., D.K., and J.M. were supported by the Deutsche Krebshilfe (10-1699-Fr 1), the Deutsche Forschungsgemeinschaft (FR1516/1-1). C.P. is a Leukemia and Lymphoma Society Scholar.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Baylin, S.B., Esteller, M., Rountree, M.R., Bachman, K.E., Schuebel, K., and Herman, J.G. 2001. Aberrant patterns of DNA methylation, chromatin formation and gene expression in cancer. *Hum. Mol. Genet.* **10:** 687–692.

Bird, A.P. 1986. CpG-rich islands and the function of DNA methylation. *Nature* **321:** 209–213.

Costello, J.F. and Plass, C. 2001. Methylation matters. *J. Med. Genet.* **38:** 285–303.

Costello, J.F., Fŕhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., Gao, X., Wright, F.A., Feramisco, J.D., Peltomaki, P., Lang, J.C., et al. 2000. Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat. Genet.* **24:** 132–138.

Costello, J.F., Plass, C., Arap, W., Chapman, V.M., Held, W.A., Berger, M.S., Su Huang, H.J., and Cavenee, W.K.1997. Cyclin-dependent kinase 6 (CDK6) amplification in human gliomas identified using two-dimensional separation of genomic DNA. *Cancer Res.* **57:** 1250–1254.

Dai, Z., Lakshmanan, R.R., Zhu, W.G., Smiraglia, D.J., Rush, L.J., Frühwald, M.C., Brena, R.M., Li, B., Wright, F.A., Ross, P., et al. 2001. Global methylation profiling of lung cancer identifies novel methylated genes. *Neoplasia* **3:** 314–323.

Frühwald, M.C. and Plass, C. 2002. Global and gene-specific methylation patterns in cancer: Aspects of tumor biology and clinical potential. *Mol. Genet. Metab.* **75:** 1–16.

Frühwald, M.C., O'Dorisio, M.S., Rush, L.J., Reiter, J.L., Smiraglia, D.J., Wenger, G., Costello, J.F., White, P.S., Krahe, R., Brodeur, G.M., et al. 2000. Gene amplification in PNETs/medulloblastomas: Mapping of a novel amplified gene within the MYCN amplicon. *J. Med. Genet.* **37:** 501–509.

Frühwald, M.C., O'Dorisio, M.S., Dai, Z., Rush, L.J., Krahe, R., Smiraglia, D.J., Pietsch, T., Elsea, S.H., and Plass, C. 2001a. Aberrant hypermethylation of the major breakpoint cluster region in 17p11.2 in medulloblastomas but not supratentorial PNETs. *Genes Chromosomes Cancer* **30:** 38–47.

Frühwald, M.C., O'Dorisio, M.S., Dai, Z., Tanner, S.M., Balster, D.A., Gao, X., Wright, F.A., and Plass, C. 2001b. Aberrant promoter methylation of previously unidentified target genes is a common abnormality in medulloblastomas—implications for tumor biology and potential clinical utility. *Oncogene* **20:** 5033–5042.

Gray, J.W. and Collins, C. 2000. Genome changes and gene expression in human solid tumors. *Carcinogenesis* **21:** 443–452.

Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Shibata, H., Akasako, A., Muramatsu, M., Kawai, J., Hirasawa, T., Watanabe, S., Shiroishi, T., et al. 1994a. A genetic linkage map of the mouse using restriction landmark genomic scanning (RLGS). *Genetics* **138:** 1207–1238.

Hayashizaki, Y., Shibata, H., Hirotsune, S., Sugino, H., Okazaki, Y., Sasaki, N., Hirose, K., Imoto, H., Okuizumi, H., Muramatsu, M., et al. 1994b. Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method. *Nat. Genet.* **6:** 33–40.

Hughes, S.J., Glover, T.W., Zhu, X.X., Kuick, R., Thoraval, D., Orringer, M.B., Beer, D.G., and Hanash, S. 1998. A novel amplicon at 8p22–23 results in overexpression of cathepsin B in esophageal adenocarcinoma. *Proc. Natl. Acad. Sci.* **95:** 12410–12415.

Jones, P.A. and Laird, P.W. 1999. Cancer epigenetics comes of age. *Nat. Genet.* **21:** 163–167.

Ohsumi, T., Okazaki, Y., Hirotsune, S., Shibata, H., Muramatsu, M., Suzuki, H., Taga, C., Watanabe, S., and Hayashizaki, Y. 1995. A spot cloning method for restriction landmark genomic scanning. *Electrophoresis* **16:** 203–209.

Okazaki, Y., Okuizumi, H., Sasaki, N., Ohsumi, T., Kuromitsu, J., Kataoka, H., Muramatsu, M., Iwadate, A., Hirota, N., Kitajima, M., et al. 1994. A genetic linkage map of the mouse using an expanded production system of restriction landmark genomic scanning (RLGS Ver.1.8). *Biochem. Biophys. Res. Commun.* **205:** 1922–1929.

Osoegawa, K., Tateno, M., Woon, P.Y., Frengen, E., Mammoser, A.G., Catanese, J.J., Hayashizaki, Y., and de Jong, P.J. 2000. Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res.* **10:** 116–128.

Plass, C., Shibata, H., Kalcheva, I., Mullins, L., Kotelevtseva, N., Mullins, J., Kato, R., Sasaki, H., Hirotsune, S., Okazaki, Y., et al. 1996. Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. *Nat. Genet.* **14:** 106–109.

Rush, L.J. and Plass, C. 2002. Restriction landmark genomic scanning for DNA methylation in cancer: Past, present, and future applications. *Analyt. Biochem.* (In press).

Rush, L.J., Dai, Z., Smiraglia, D.J., Gao, X., Wright, F.A., Frühwald, M., Costello, J.F., Held, W.A., Yu, L., Krahe, R., et al. 2001. Novel methylation targets in de novo acute myeloid leukemia with prevalence of chromosome 11 loci. *Blood* **97:** 3226–3233.

Rush, L.J., Heinonen, K., Mrozek, K., Wolf, B.J., Abdel-Rahman, M., Szymanska, J., Peltomaki, P., Kapadia, F., Bloomfield, C.D., Caligiuri, M.A., et al. 2002. Comprehensive cytogenetic and molecular genetic characterization of the TI-1 acute myeloid leukemia cell line reveals cross-contamination with K-562 cell line. *Blood* **99:** 1874–1876.

Shiraishi, M., Sekiguchi, A., Oates, A.J., Terry, M.J., and Miyamoto, Y. 2002a. HOX gene clusters are hotspots of de novo methylation in CpG islands of human lung adenocarcinomas. *Oncogene* **21:** 3659–3662.

Shiraishi, M., Sekiguchi, A., Terry, M.J., Oates, A.J., Miyamoto, Y., Chuu, Y.H., Munakata, M., and Sekiya, T. 2002b. A comprehensive catalog of CpG islands methylated in human lung adenocarcinomas for the identification of tumor suppressor genes. *Oncogene* **21:** 3804–3813.

Smiraglia, D.J. and Plass, C. 2002. The study of aberrant methylation in cancer via restriction landmark genomic scanning. *Oncogene* **21:** 5414–5426.

Smiraglia, D.J., Frühwald, M.C., Costello, J.F., McCormick, S.P., Dai, Z., Peltomaki, P., O'Dorisio, M.S., Cavenee, W.K., and Plass, C. 1999. A new tool for the rapid cloning of amplified and hypermethylated human DNA sequences from restriction landmark genome scanning gels. *Genomics* **58:** 254–262.

Smiraglia, D.J., Rush, L.J., Frühwald, M.C., Dai, Z., Held, W.A., Costello, J.F., Lang, J.C., Eng, C., Li, B., Wright, F.A., et al., 2001. Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. *Hum. Mol. Genet.* **10:** 1413–1419.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. 2001. The sequence of the human genome. *Science* **5507:** 1304–1351.

Wimmer, K., Zhu, X.X., Rouillard, J.M., Ambros, P.F., Lamb, B.J., Kuick, R., Eckart, M., Weinhausl, A., Fonatsch, C., and Hanash, S.M. 2002. Combined restriction landmark genomic scanning and virtual genome scans identify a novel human homeobox gene, ALX3, that is hypermethylated in neuroblastoma. *Genes Chromosomes Cancer* **33:** 285–294.

Web Site References

http://genome.ucsc.edu; Web site offers free access to the human genomic sequence.

Received February 20, 2002; accepted in revised form July 25, 2002.