Genomics & Informatics Vol. 4(1) 1-10, March 2006

Cloning of *NotI*-linked DNA Detected by Restriction Landmark Genomic Scanning of Human Genome

Jeong-Hwan Kim[1], Kyung-Tae Lee[1], Hyung-Chul Kim[1], Jin-Ok Yang[2], Yoonsoo Hahn[1], Sangsoo Kim[2,5], Seon-Young Kim[1], Hyang-Sook Yoo[3] and Yong Sung Kim[1,4*]

[1]Human Genomics Laboratory, Genome Research Center, [2]National Genome Information Center, [3]The Center for Functional Analysis of Human Genome, 21st Century Frontier R&D Program, [4]Department of Functional Genomics, University of Science & Technology, Korea Research Institute of Bioscience and Biotechnology, 52 Eoeun-dong, Yuseong-gu, Daejeon 305-333, Korea, [5]Department of Bioinformatics, Soongsil University, Seoul 156-743, Korea.

Abstract

Epigenetic alterations are common features of human solid tumors, though global DNA methylation has been difficult to assess. Restriction Landmark Genomic Scanning (RLGS) is one of technology to examine epigenetic alterations at several thousand *NotI* sites of promoter regions in tumor genome. To assess sequence information for *NotI* sequences in RLGS gel, we cloned 1,161 unique *NotI*-linked clones, compromising about 60% of the spots in the soluble region of RLGS profile, and performed BLAT searches on the UCSC genome server, May 2004 Freeze. 1,023 (88%) unique sequences were matched to the CpG islands of human genome showing a large bias of RLGS toward identifying potential genes or CpG islands. The cloned *NotI*-loci had a high frequency (71%) of occurrence within CpG islands near the 5' ends of known genes rather than within CpG islands near the 3' ends or intragenic regions, making RLGS a potent tool for the identification of gene-associated methylation events. By mixing RLGS gels with all *NotI*-linked clones, we addressed 151 *NotI* sequences onto a standard RLGS gel and compared them with previous reports from several types of tumors. We hope our sequence information will be useful to identify novel epigenetic targets in any types of tumor genome

Keywords: RLGS, epigenetics, cancer, CpG island

*Corresponding author: E-mail yongsung@kribb.re.kr
Tel +82-42-879-8110, Fax +82-42-879-8119
Accepted 20 February 2006

Introduction

CpG islands are short stretches (500-2000 bp) of genomic DNA enriched for the dinucleotide, 5'-CpG-3', which is the substrate for methylation by DNA methyltransferases. While most CpG sites in the human genome are methylated, those in CpG islands are typically unmethylated in normal tissue. In human cancers, *de novo* methylation of CpG island sequences is accompanied by gene silencing and can serve as an alternative to mutation or deletion in the inactivation of tumor suppressor and other genes. In fact, over the past several years many studies have demonstrated that epigenetic alterations are responsible for the development and progression of the human cancers (Jones and Baylin, 2002).

The ubiquity of DNA methylation changes in tumorigenesis has led to a host of innovative diagnostic and therapeutic strategies. Recent advances attest to the promise of DNA methylation markers as powerful tools for the clinic in the future. For many epigenetically silenced genes, re-expression in tumor cells can lead to suppression of cell growth or altered sensitivity to existing anticancer therapies, and small molecules that reverse epigenetic inactivation are now undergoing clinical trials in cancer patients (Momparler *et al.*, 1997; Pohlmann *et al.*, 2002). Epigenetic changes have been detected in the peripheral blood for almost every organ in cancer patients (Laird, 2003). Thus, epigenetic changes are not only potential therapeutic targets because of their reversibility, but also potential biomarkers that can be used to detect and diagnose cancer in its earliest stage and to accurately assess individual risk (Brown and Strathdee, 2002).

Restriction landmark genomic scanning (RLGS) is a highly reproducible two-dimensional gel electrophoresis of genomic DNA that allows the assessment of over 2000 loci simultaneously, when a methylation sensitive enzyme *NotI* is used as the landmark enzyme (Hatada *et al.*, 1991). The technique has been used for various purposes including genetic mapping, identification of novel imprinted genes, genomic amplifications, regions of hypomethylation or hypermethylation, candidate tumor suppressor genes, and measuring the degree of CpG island hypermethylation in cancer (Costello *et al.*, 2000; Hayashizaki *et al.*, 1993; Kim *et al.*, 2000; Miwa *et al.*, 1995; Nagai *et al.*, 1994; Okazaki *et al.*, 1996; Okuizumi *et al.*, 1995; Plass *et al.*, 1996; Shibata *et al.*, 1995).

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 *NotI*/*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 (Costello *et al.*, 2000; Rush *et al.*, 2004; Smiraglia and Plass 2002).

The use of the *NotI*/*Eco*RV boundary library as a cloning tool for RLGS has a limitation in covering all *NotI*-linked clones because the spots on RLGS gel originates from *NotI*/*NotI* clones as well as *NotI*/*Eco*RV clones. To increase the potential coverage of CpG islands, we prepared a *NotI*/*NotI* library in addition to a *NotI*/*Eco*RV library and RLGS mixing gels that allow the efficient recovery of the cloned RLGS fragments. We hope that this novel resource, together with the sequence information of the previous *NotI*/*Eco*RV library, will greatly increase the utility of RLGS.

Methods

RLGS run

We isolated high molecular weight DNA from a normal stomach mucosa according to the standard protocol and performed RLGS as previously described (Hatada *et al.*, 1991). Briefly, five μg of genomic DNA was blocked in a 25 μl reaction by adding nucleotide analogs ([α-^{35}S]-dGTP, [α-^{35}S]-dCTP, ddATP, and ddTTP) with 2.5 U of DNA polymerase I (TAKARA, Japan) for 20 min at 37°C, followed by inactivation of the enzyme at 65°C for 30 min. The DNA was then digested with 50 units of *NotI* (New England BioLabs Inc., Ipswich, MA). We used Sequenase Ver. 2.0 (USB, Cleveland, OH) to fill in the *NotI* ends with [α-^{32}P]-dGTP (6,000 Ci/mmole; NEN) and [α-^{32}P]-dCTP (6,000 Ci/mmole; NEN) for 30 min at 37°C and stopped the reaction by heating for 30 min at 65°C. The labeled DNA was then digested with 20 units of *Eco*RV (New England BioLabs Inc.) and separated at 8 V/cm for 12 h in 0.8% SeaKem agarose gel (FMC) for first-dimensional separation. The DNA-containing agarose strip was in gel digested with 1,500 units of *HinfI* (New England BioLabs Inc.) for 2 h at 37°C. The agarose strip equilibrated in buffer was then placed across the top of a non-denaturing 5% polyacrylamide gel, connected by molten agarose, and followed by second-dimensional electrophoresis at 8 V/cm for 7 h. After separation, the gels were dried and exposed to X-ray film (Kodak X-OMAT) in the presence of intensifying screens for 2-10 days. To find out which spot originated from the *NotI*/*Eco*RV or the *NotI*/*NotI* fragments of the 1st dimension, DNA digested only with *NotI* enzyme and the mixed DNA containing both *NotI* and *NotI*/*Eco*RV digested DNA were used for RLGS analyses.

RLGS profile analysis

Two RLGS profiles from DNA digested only with *NotI* enzyme and mixed DNA containing both *NotI* and *NotI*/*Eco*RV digested DNA were overlaid, and the differences between the two profiles were detected by visual inspection and independently validated by two investigators. To exclude a difficulty due to high density or low resolution of spots and to allow the uniform comparison of RLGS profiles from different samples, we compared spots only on the central portion of the RLGS profile.

NotI-linked DNA library construction.

To generate *NotI*/*Eco*RV and *NotI*/*NotI* libraries as a resource to facilitate the analysis of RLGS loci, one hundred μg of normal mucosa tissue was doubly digested with *NotI*/*Eco*RV and the resulting fragments were purified by using the phenol/ethanol extraction method. First, to exclude *Eco*RV-*Eco*RV DNA fragments in DNA solution, λ-ZAPII DNA (22.12 kb) was digested with *NotI* enzyme and its *NotI* end was treated with CIP (TAKARA). The λ-ZAPII/*NotI* DNA was ligated with genomic DNA and then digested with *Eco*RV to separate fragments ligated at *Eco*RV sites. The mixtures were run on 0.8% LMP agarose (FMC) gel and the portion of the gel over 22 kb was eluted and purified with beta-agarase I (New England BioLabs Inc.). This DNA was digested with *NotI*, run on 1% LMP agarose gel, and fragments ranging 0.7 to 4 kb were eluted by usinga Gel Extraction kit (Qiagen). The resulting DNA solution was then divided into two tubes: DNA fragments from one tube were ligated into pBluescript KS (+) DNA digested with *NotI* and *Eco*RV and the other fragments were ligated into pBluescript KS (+) digested only with *NotI*. The solutions were transformed into DH5α cells by electroporation. Finally, we prepared two kinds of *NotI*-linked DNA library, *NotI*/*NotI* DNA library ('NN') and *NotI*/*Eco*RV DNA library ('NE') of 0.7 to 4 kb fragments comprising the central portion of the RLGS gel (Fig. 2).

Isolation of unique *NotI*-linked DNA and BLAT search

Plasmid DNAs from all clones of NN and NE libraries were isolated by using a MWG plasmidprep 96 (MWG Biotech.). Sequencing reactions were performed on a

GeneAmp PCR System 9700 thermal reactor (Perkin-Elmer Corp.) by using a BigDye Terminator Sequencing kit with T7 sequencing primer to get *Not*I end sequences. After removing the unincorporated dye terminator, the reaction products were run on an ABI prism 3700 DNA analyzer (Perkin-Elmer Corp.). DNA sequences were assembled to isolate unique *Not*I-linked clones by using DNASTAR software. BLAT search for unique sequences were performed on the UCSC genome server, May 2004 Freeze (http://genome.ucsc.edu/cgi-bin/hgBlat) to determine whether the sequences fall in a known CpG island or reported gene region.

Addressing of unique *Not*I-linked clones onto RLGS gel

Plasmid DNAs with unique *Not*I-linked sequences were arrayed into 12 96-well microtiter plates for their use in the RLGS mixing gels. The DNAs were pooled in three different ways: by plate, by row, and by column, as described previously (Smiraglia *et al.*, 1999). The 96 plasmid DNAs from each of the 12 plates were pooled into 12 microtubes. The 8 rows (A-H) from each of the 12 plates were also pooled into new microtiter plates by using a 12-channel pipette and finally transferred into 8 microtubes. The 12 columns (1-12) from each of the 12 plates were similarly pooled by using an 8-channel pipette and finally transferred into 12 microtubes. In total, 32 pooled DNAs were labeled according to the standard RLGS procedure and mixed with the labeled genomic DNA when the first dimensional agarose gel was loaded. The standard RLGS procedure described above was continued.

Results and Discussion

We generated a standard RLGS profile for DNAs from a normal mucosa and divided the profile into 30 sections to allow a uniform comparison of RLGS profiles from different samples. Then each fragment was given a three-variable designation (Y coordinate, X coordinate, fragment number). The central region of the RLGS profile had 30 sections (18 vertically and A-D horizontally), containing 1,948 spots (Fig. 1A). The 1A and 2A sections were excluded because there was no fragment in those sections. A standard RLGS profile, on which individual spot number was assigned and is available at http://21cgenome.kribb.re.kr/html/2004_new/RLGS_master/image01.html.

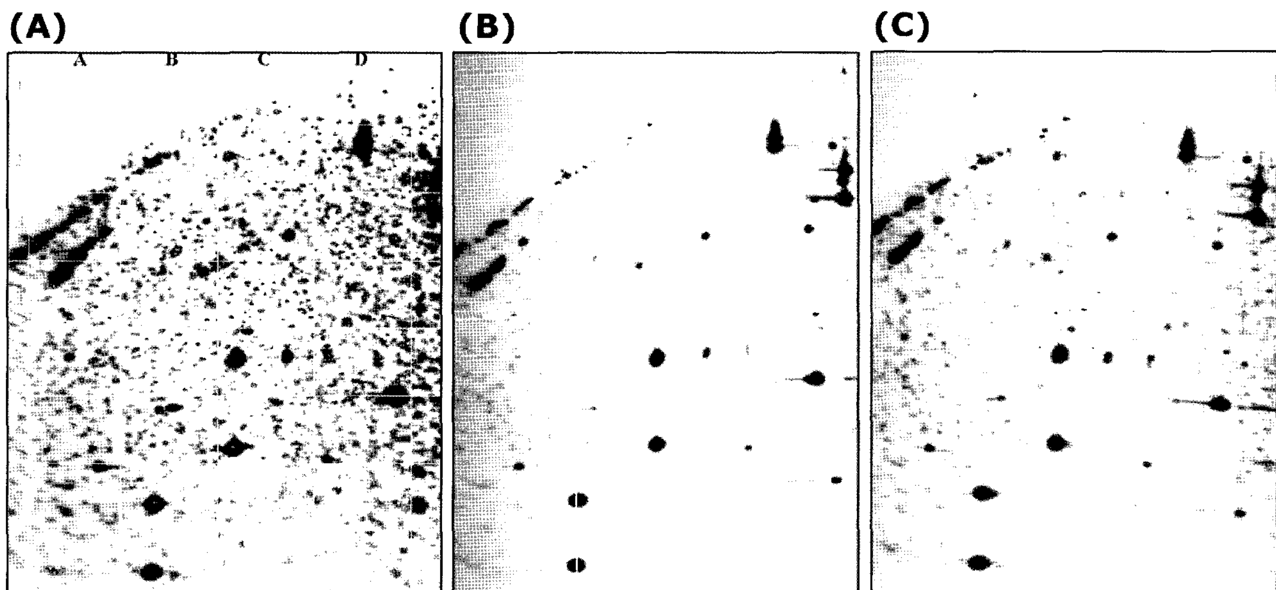

Fig. 1. RLGS profile using *Not*I/*Eco*RV/HinfI restriction enzymes in human genomic DNA. (A) A standard RLGS profile from a normal mucosa DNA displaying nearly 2,300 *Not*I fragments. For the comparisons of RLGS profiles from different samples, each spot is given a three-variable designation (Y coordinate, X coordinate, spot number). The enlarged RLGS profile is accessible at http://www.21cgenome.kribb.re.kr/standardRLGS. To know how many spots were derived from *Not*I/*Not*I DNA fragments of the first dimension, two RLGS profiles were compared with each other. (B) A RLGS profile from DNA digested with only *Not*I enzyme showing over 400 spots. (C) A profile from mixed DNA of equal amounts with *Not*I/*Eco*RV and *Not*I digestion. Spots (B) from the *Not*I digestion can be distinguished as enhanced spots compared to those (C) from the *Not*I/*Eco*RV digestion.*

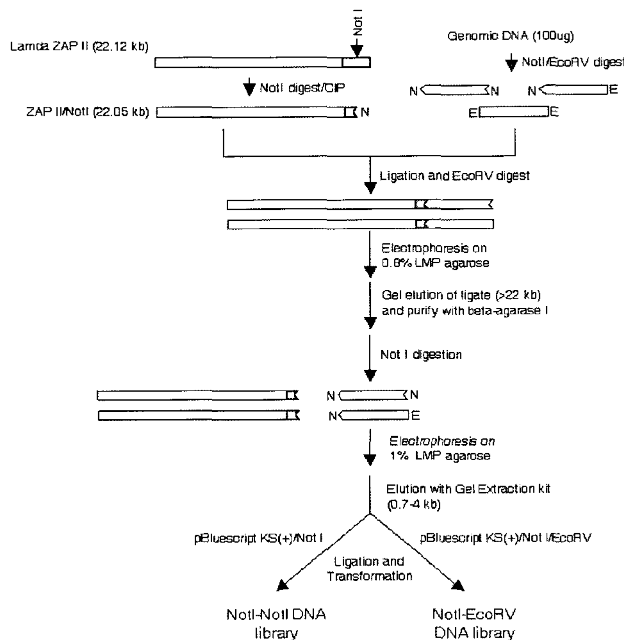

Fig. 2. Cloning strategy of *Not* I-linked DNA fragments and three-dimensional addressing onto RLGS gel. *Not*I-linked DNA fragments (*Not*I/*Eco*RV or *Not*I/*Not*I) ranging from 0.7 to 4 kb were selected and ligated into a pBluescript KS (+) vector to establish *Not*I/*Not*I or *Not*I/*Eco*RV libraries

We found that 404 spots in the central portion of the standard RLGS profile originated from *NotI*/*NotI* DNA fragments (Fig. 1B). Most of the *NotI*/*NotI* DNA fragments reappeared in the RLGS profile originated from *NotI*/*Eco*RV DNA fragments (Fig.1C). This result indicates that 20.7% (404 of 1948 spots) on the standard RLGS profile originated from *NotI*/*NotI* DNA fragments and that *NotI*/*NotI* DNA fragments are also important parts of the *NotI*/*Eco*RV DNA fragments. To identify *NotI*/*NotI* as well as *NotI*/*Eco*RV DNA fragments as a resource to facilitate the analysis of RLGS loci, we established an 'NN library' and an 'NE library' containing *NotI*/*NotI* and *NotI*/*Eco*RV fragments ranging from 0.7 to 4 kb (Fig. 2). We collected a total of 5,255 white colonies, of which 4,328 were from the NE library and 927 from the NN library, and determined their *NotI*-end sequences. All *NotI*-end sequences were assembled to select unique sequences by using DNASTAR software. Totally 1,161 unique sequences (or clones) were isolated; 1,051 sequences were isolated from the NE library and 110 from the NN library. The sequence data were submitted to NCBI GenBank as accession numbers CG464575-CG465835.

The average insert size for unique clones was approximately 1.5 kb (n = 860) ranging from 0.7 to 4 kb and average read length was 423 bp. The unique clones were estimated to cover 60% of the 1,948 RLGS spots (Table 1). To identify potential genes or CpG islands linked with the above 1,161 unique sequences, we performed BLAT searches on the UCSC genome server, May 2004 Freeze (http://genome.ucsc.edu/cgi-bin/hgBlat). 1,023 (88%) unique sequences were matched to UCSC CpG islands, of which 919 (87%) were derived from NE clones and 104 (95%) from NN clones. We also found that 646 (70%) sequences matched to UCSC CpG islands were in the 5'end of genes, 136 (15%) were in the 3' end of genes, and 27 (3%) were intragenic regions, and 110 (12%) were not linked to any gene. This result is similar to the previous report that 8,239 (86%) of 9,628 *NotI* sites estimated from human genome draft sequence were linked to CpG islands (Dai *et al.*,

Table 1. Characteristics of cloned *Not*I/*Eco*RV and *Not*I/*Not*I fragment

	*Not*I-*Eco*RV	*Not*I-*Not*I	Total
Total number of DNA spot on central region of RLGS profile[a]	1544	404	1948
Total number of cloned *NotI*-linked fragments (% coverage)[b]	1051 (68%)	110 (27%)	1161 (60%)
Matched sequences with UCSC CpG islands (% CpG)[c]	919 (87%)	104 (95%)	1023 (88%)
Frequency of CpG islands near 5' end of a known gene	646 (70%)	78 (75%)	724 (71%)
Frequency of CpG islands near 3' end of a known gene	136 (15%)	14 (13%)	150 (15%)
Frequency of CpG islands inside a known gene	27 (3%)	3 (3%)	30 (3%)
Frequency of CpG islands not linked to any gene	110 (12%)	9 (9%)	119 (12%)
Unmatched sequences with UCSC CpG islands	132 (13%)	6 (5%)	138 (12%)
Frequency of sequences near 5' end of a known gene	11 (8%)	1 (17%)	12 (9%)
Frequency of sequences near 3' end of a known gene	18 (14%)	0	18 (13%)
Frequency of sequences inside a known gene	2 (2%)	1 (17%)	3 (2%)
Frequency of sequences not linked to any gene	39 (30%)	2 (33%)	41 (30%)
Frequency of no hit sequence	62 (47%)	2 (33%)	64 (46%)

[a]Indicates DNA spot numbers counted in 30 sections of Fig. 1.

[b]Indicates the number of unique *NotI*-linked clones from NE and NN libraries, respectively.

[c]Indicates the correspondence with UCSC CpG islands by BLAT search on http://genome. ucsc. edu/cgi-bin/hgBlat, May 2004 Freeze

Table 2. Cloned *Not*I-linked spots

Chromosome[a]	Present study	Master RLGS[b]	Not I-linked sequence			Gene matched[e]			Methylation in tumors[g]
			Accn-ID in GenBank	Library[c]	CpG island[d]	Symbol	Accession no.	Direction[f]	
1p21.2	4C06	3D60	CG465078	NE	I	*GPR88*	NM_022049	p	CLN, GLI, AML (1)
1p22.2	7D47	6F08	CG465415	NE	5'		BC036441	m	
1p34.3	4C90	-	CG465039	NE	5'	*ZNF258*	NM_145310	m	
1p36.22	2D16	-	CG465807	NN	5'	*LOC374946*	NM_198545	p	
					5'		AK094316	m	
1q21.3	2C44	-	CG464955	NE	5'		AK125876	p	
					5'	*MRPL9*	NM_031420	m	
1q23.3	4C55	4E07	CG464950	NE	3'		BI490420	m	HCC (2)
1q25.2	5D87	4F52	CG465167	NE	5'	*FLJ35530*	NM_207467	p	
1q25.3	7C35	6E06	CG465365	NE	5'	*IVNS1ABP*	NM_006469	m	
1q42.11	3A23	3B02	CG464589	NE	5'	*TP53BP2*	NM_005426	m	
1q43	3C48	-	CG465234	NE	5'	*FLJ40773*	NM_152666	m	
1q43	6D27	5E25	CG464911	NE	5'	*FMN2*	NM_020066	p	CLN, LNG (3)
2p13.2	5C10	4D22	CG464853	NE	5'	*EMX1*	NM_004097	p	HCC (2)
2p16.1	3D27	2F75	CG464591	NE	N		no match		
2q11.1	5A46	4B48	CG465301	NE	5'	*ZNF514*	NM_032788	m	
2q11.2	2C20	2E16	CG464730	NE	5'	*INPP4A*	NM_001566	p	
2q11.2	2D67	2F28	CG465243	NE	I	*MGC42367*	NM_207362	m	
2q12.1	3D83	3E53	CG465812	NN	I		AK096498	m	
2q14.2	4A61	-	CG464797	NE	N		no match		
2q14.3	4B37	3B44	CG465101	NE	5'	*LIMS2*	NM_017980	m	
2q24.3	7C34	6D07	CG464587	NE	5'	*LASS6*	NM_203463	p	
2q31.1	5C17	-	CG464673	NE	5'	*SP3*	NM_003111	m	
2q31.1	6C18	5D36	CG465245	NE	5'	*HOXD3*	NM_006898	m	
2q36.1	2B24	2C07	CG464913	NE	5'		BC008048	p	
2q36.1	4A50	4B01	CG465268	NE	5'	*SGPP2*	NM_152386	p	
2q36.3	4D12	3F45, 46	CG465169	NE	5'	*FLJ20701*	NM_017933	m	
3p14.1	3D84	3E54	CG464616	NE	Y		no match		
3p21.1	5A54	4B49	CG465121	NE	5'	*SFMBT1*	NM_016329	m	
3p21.31	4C67	-	CG465075	NE	5'	*DHX30*	NM_138614	p	
3q12.1	5B13	-	CG465490	NE	5'	*DCBLD2*	NM_080927	m	
3q27.1	7C02	5B38	CG465798	NN	3'	*CAM-KIIN*	NM_033259	m	
4p14	3C61	-	CG465023	NE	5'	*FLJ11017*	BC037906	p	
4p16.1	6B47	5C29	CG464976	NE	5'		BC040874	m	
4p16.3	5C30	4D34	CG464576	NE	5'	*SH3BP2*	NM_003023	p	HCC (2)
4p16.3	8B01	-	CG464965	NE	W	*HPX-153*	X76978	m	
4q11	2C46	2D26	CG465017	NE	5'	*SGCB*	NM_000232	m	HCC (2)
4q13.1	5A61	-	CG465347	NE	5'	*LPHN3*	NM_015236	p	
4q24	3B37	3C07	CG464713	NE	5'	*CXXC4*	NM_025212	m	
4q24	6C10	5D30	CG464728	NE	5'	*LOC255743*	NM_198278	p	
4q31.21	7D05	-	CG464813	NE	5'	*KIAA0882*	AB020689	m	
4q35.1	7D24	-	CG465462	NE	5'	*ACSL1*	NM_001995	m	
5p13.1	4C17	3D37	CG464996	NE	5'	*MGC39830*	BC029608	m	
5p15.33	3B71	3C02	CG464841	NE	W	*IRX1*	BC029160	p	CLN, HNSCC, AML(1)
5q11.2	6C72	-	CG464924	NE	N		no match		
5q15	6B57	5C30	CG464628	NE	5'		BC010610	m	
5q23.3	3D04	2E49	CG465126	NE	5'	*CSS3*	NM_175856	p	
5q23.3	7B40	6C02	CG464893	NE	N		no match		
6p21.31	2D10	-	CG464794	NE	5'	*TEAD3*	NM_003214	m	
					3'	*TULP1*	NM_003322	m	
6q12	3C22	2D57	CG464928	NE	5'	*BAI3*	AB011122	p	
					5'		AL833240	m	
6q13	3B80	3C17	CG464977	NE	5'	*C6orf155*	NM_024882	m	CLN (1)
6q16.1	3C75	3E48	CG465060	NE	5'	*POU3F2*	NM_005604	p	
6q21	4D07	3E30	CG465809	NN	5'	*SLC16A10*	NM_018593	p	
6q23.2	3C58	2D62	CG465370	NE	5'	*ENPP1*	NM_006208	p	
6q25.1	6B51	5C22	CG465020	NE	Y		no match		
7q11.2	4B81	4B11	CG464818	NE	Y		no match		

Table 2. Continued

Chromosome[a]	Present study	Master RLGS[b]	Not I-linked sequence			Gene matched[e]			Methylation in tumors[g]
			Accn-ID in GenBank	Library[c]	CpG island[d]	Symbol	Accession no.	Direction[f]	
7p13	3B43	2C58	CG465180	NE	5'	*FLJ00315*	AK090415	p	HCC (2)
7p13	3B73	3C04	CG465340	NE	5'	*BLVRA*	NM_000712	p	
7p15.3	6D91	5E27	CG465464	NE	5'	*DNAH11*	NM_003777	p	
7p22.3	4B63	3C64	CG464849	NE	Y		no match		AML (4)
7q21.11	7B03	-	CG464791	NE	Y		no match		
7q21.3	5D57	-	CG464652	NE	5'		AK096514	p	
7q31.33	6B34	5B28	CG465150	NE	Y		no match		
7q36.3	3C17	2E56	CG465441	NE	Y		no match		
8p12	2B16	2C13	CG465783	NN	5' (no CpG)	*WRN*	NM_000553	p	GLI (5)
					5'	*PURG*	NM_013357	m	
8q13.1	4D21	-	CG465336	NE	5'	*PDE7A*	NM_002604	m	
8q24.11	6B26	-	CG465122	NE	5'	*THRAP6*	NM_080651	p	
8q24.12	7B38	6C06	CG464964	NE	5'	*FLJ39458*	NM_207506	m	
8q24.21	4B66	4C04	CG465779	NN	5'	*MYC*	NM_002467	p	
8q24.22	7A32	6B16	CG464757	NE	5'	*ADCY8*	NM_001115	m	
9q21.32	5A27	-	CG464989	NE	5'	*UBQLN1*	NM_053067	m	
9q32	2C22	2E02	CG465345	NE	5'		AK096613	p	
					5'		AK131020	m	
9q33.3	6C26	5D13	CG465292	NE	I	*LMX1B*	NM_002316	p	CLN (1), CLL (6)
9q34.12	3B26	2C48	CG464971	NE	5'	*ABL1*	NM_005157	p	
9q34.13	3B74	-	CG464594	NE	5'	*C9orf58*	NM_031426	p	
10p12.1	3D21	3E68	CG465756	NN	5'	*C10orf48*	NM_173576	m	
10p12.2	5D04	4E23	CG465066	NE	5'	*COMMD3*	NM_012071	p	
10p13	3C04	2D40	CG464803	NE	5'		L36531	m	
10q11.21	3C45	2D37	CG464603	NE	5'	*ALOX5*	NM_000698	p	
10q11.22	7C17	6D02	CG465797	NN	I	*ARHGAP22*	NM_021226	m	
10q21.2	6B14	-	CG464979	NE	5'	*RHOBTB1*	NM_014836	m	
10q21.3	2B49	-	CG465452	NE	N		no match		
10q21.3	4B51	3C67	CG465236	NE	5'	*MGC14425*	NM_032903	p	AML (1)
					5'		AK027280	m	
10q21.3	4D69	3E06	CG465063	NE	5'	*CXXC6*	NM_030625	p	
10q24.32	3B04	2B26	CG464851	NE	N		no match		
11p11.2	7D01	-	CG465162	NE	5'	*MADD*	NM_003682	p	
11p13	3B61	3C25	CG464972	NE	I	*WIT-1*	NM_015855	p	
11p15.4	4C65	3E02	CG464715	NE	5'	*AMPD3*	NM_000480	p	
11q12.1	4A36	3B22	CG464657	NE	5'	*UBE2L6*	NM_004223	m	
11q12.1	4A47	3A25	CG465322	NE	3'	*RTN4RL2*	NM_178570	p	
11q21	3B79	3C18	CG464599	NE	I		FLJ25393	p	HCC (2)
11q23.3	8B25	-	CG464973	NE	5'	*FXYD6*	NM_022003	m	
12p11.21	7C30	6D03	CG464689	NE	5'	*IPO8*	NM_006390	m	
12q13.11	8C22	-	CG464864	NE	5'	*SLC38A1*	NM_030674	m	
12q13.13	3C72	3D22	CG464748	NE	5'		AK095191	m	
12q13.3	5C33	4D33	CG464771	NE	5' (no CpG)	*LOC283377*	NM_207344	p	
12q21.1	8A17	-	CG465082	NE	5'	*GPR49*	NM_003667	p	
12q21.2	2B42	1B02	CG464943	NE	5'	*E2F7*	NM_203394	m	
12q24.12	4A18	-	CG465730	NN	5'	*RUNX2*	NM_004348	p	
12q24.21	3A12	2B36	CG464758	NE	Y		no match		
12q24.23	6A21	5B05	CG464646	NE	I	*MSI1*	NM_002442	m	
12q24.31	4B61	-	CG464683	NE	5'	*ATP6V0A2*	NM_012463	p	
13q12.2	3D85	3E55	CG464742	NE	3'	*IPF1*	NM_000209	p	LNG(7), AML(4), CLL(6)
13q14.2	2C05	1D19	CG465793	NN	5'		AK056507	p	
					5'		AF380424	m	
13q32.1	2C04	-	CG464676	NE	Y		no match		
14q12	7C10	-	CG465266	NE	5'	*PRKD1*	NM_002742	m	
14q24.3	3C118	2E63	CG464619	NE	5'	*CHX10*	NM_182894	p	
14q32.32	5B52	4C36	CG465757	NN	5'	*AMN*	NM_030943	m	
15q11.2	3B32	-	CG465146	NE	Y		no match		

Table 2. Continued

Chromosome[a]	Present study	Master RLGS[b]	Not I-linked sequence: Accn-ID in GenBank	Library[c]	CpG island[d]	Gene matched[e]: Symbol	Accession no.	Direction[f]	Methylation in tumors[g]
15q15.3	7C52	6D26	CG464921	NE	5'	*FLJ12973*	NM_024908	p	
15q21.3	2B43	1B01	CG464755	NE	5'	*NEDD4*	NM_006154	m	
15q22.2	6D05	5E12	CG465416	NE	Y		no match		BRE, GLI, AML (1)
15q23	3C85	3D05, 06	CG465160	NE	Y		no match		
15q23	4A24	—	CG465233	NE	5' (no CpG)		BC033162	m	
15q23	6C31	—	CG464852	NE	5'	*LOC196993*	BC048128	m	
15q23	6D57	—	CG464745	NE	Y		no match		
15q24.1	2C42	—	CG465398	NE	5'	*CLK3*	BC002555	p	
15q24.3	2C08	1E10	CG464759	NE	5'	*FLJ14594*	BC011057	m	
15q25.2	3D75	—	CG464743	NE	5'	*BTBD1*	NM_025238	m	
16p13.3	5B15	4C24	CG465073	NE	I	*MMP25*	NM_022718	p	
16q11.2	2C51	2D33	CG464639	NE	5'	*GPT2*	NM_133443	p	
16q12.2	7B26	6C01	CG465012	NE	I		AK092208	p	
16q22.1	4B22	—	CG464655	NE	5'	*CTCF*	NM_006565	p	
16q23.1	5B60	4C44	CG465764	NN	5'	*MAF*	NM_005360	m	
16q24.3	2B22	2C08	CG465178	NE	5' (no CpG)	*CBFA2T3*	NM_005187	m	
16q24.3	5B06	4C16	CG464685	NE	5'	*AFG3L1*	NM_001132	p	CLL (6)
					5'	*MGC16385*	NM_145039	m	
17p13.3	4A44	3B60	CG464723	NE	5'	*ABR*	NM_021962	m	RCC (8)
17q11.1	6C13	5D24	CG464954	NE	I (no CpG)	*WSB1*	NM_015626	p	
17q21.1	5A55	4B50	CG464880	NE	I	*RAPGEFL1*	NM_016339	p	
17q21.31	4A41	3B50	CG464687	NE	5'		BC045575	p	
17q21.32	5D103	4E57, 58	CG465324	NE	Y		no match		
17q21.32	7C15	6D09	CG465051	NE	5'		AK127707	m	
17q23.2	3C56	2D66	CG464941	NE	Y		no match		
17q23.2	6B49	5C34	CG465777	NN	5'	*NOG*	NM_005450	p	
18q21.2	6B54	5C23	CG464927	NE	N	*TCF4*	NM_003199	I	
19p13.11	3C42	—	CG465484	NE	5'	*CRLF1*	NM_004750	m	
19p13.12	4B14	3C47	CG464802	NE	3'	*PRKCL1*	NM_002741	p	
19q13.12	6A51	—	CG465327	NE	I		AK124779	m	
21q21.1	5D84	4F55	CG464578	NE	5'	*NCAM2*	NM_004540	p	CLN (1)
21q22.11	6B05	5C10	CG464640	NE	5'	*SYNJ1*	NM_003895	m	
21q22.3	3B13	2B55	CG465742	NN	5'	*POFUT2*	NM_015227	m	HCC (2), RCC (8)
M	3D19	—	CG465296	NE	I		AY029066	p	
Xp22.22	5A28	4B38	CG465077	NE	5'	*MID1*	NM_000381	m	
Xq24	4B84	4B05	CG465395	NE	5'	*WDR44*	NM_019045	p	BRE, CLN, GLI (1)

[a]Indicates chromosomal localization by BLAT search on http://genome.ucsc.edu/cgi-bin/hgBlat, May 2004 Freeze.
[b]Indicates spot name corresponded on Master RLGS described in previous work (Costello *et al.*, 2000). Blank (-) indicates that the spot position is obscure in master profile.
[c]NE indicates the sequence originated from *NotI*-*Eco*RV DNA library and NN from *NotI*-*NotI* DNA library.
[d]*NotI*-linked sequences were compared to CpG islands in http://genome.ucsc.edu/cgi-bin/hgBlat. 5' indicates when a matched CpG island spans the region upstream of transcription start site and/or 1st exon 1; 3', when a matched CpG island is found in the last exon and/or 3' UTR; W, when a matched CpG island covers the whole gene; I, when a matched CpG island occurs in internal exons and introns; Y, when a matched CpG island is found, but it is not matched to genes or mRNAs known; N, when any matched CpG island is not found in the region; 5' (no CpG) or I (CpG), when any matched CpG island is not found, but NotI-linked sequence localizes 5' upstream or internal region of the gene.
[e]Indicates RefSeq or mRNA matched in http://genome.ucsc.edu/cgi-bin/hgBlat.
[f]P or M indicate each gene sequence to be 'plus strand' or 'minus strand' along each chromosome.
[g]Indicates methylation in various tumors described in previous literatures: (1) Costello *et al.*, 2000, (2) Nagai *et al.*, 1994, (3) Smiraglia *et al.*, 2001, (4) Rush *et al.*, 2001, (5) Nakamura *et al.*, 1997, (6) Rush *et al.*, 2004, (7) Dai *et al.*, 2001, (8) Cho *et al.*, 1998. AML, acute myeloid leukemia; BRE, breast carcinoma; CLL, chronic lymphocytic leukemia; CLN, colon carcinoma; GLI, glioma; HCC, hepatocellular carcinoma, HNSCC, head and neck squamous cell carcinoma; LNG, lung carcinoma; RCC, renal cell carcinoma.

2002), showing a large bias of RLGS toward identifying potential genes or CpG islands. It is worthy to note that the *NotI*-loci we cloned have a high frequency (71%) of occurrence within CpG islands near the 5' ends of known genes rather than within CpG islands near the 3' ends or intra-genic regions, thus making RLGS a potent tool for the identification of gene-associated methylation events. Several groups generated RLGS profiles from many types of cancer and reported interesting genes (Costello *et al.*, 2000; Smiraglia *et al.*, 1999; Zardo *et al.*, 2002). However, because previous studies used only *NotI*/*Eco*RV DNA fragments for mixing gel catalog, they have

limitations in identifying novel epigenetic targets. Thus, our *Not*I-linked sequences including *Not*I/*Not*I as well as *Not*I/*Eco*RV DNA fragments may be helpful to give additional information on RLGS study.

To identify each unique *Not*I-linked clone on the RLGS gel, all unique clones were arrayed into 12 96-well microtiter plates and RLGS mixing gels were prepared from plates 1 to 12. *Not*I/*Not*I clones from multi-copy rDNA were excluded in this step because the positions of *Not*I/*Not*I fragments derived from genomic rDNA are already known (Kuick *et al.*, 1996). The rows and columns from these 12 plates were individually pooled to produce mixing gels as previously described (Smiraglia *et al.*, 1999). In total, 32 RLGS mixing gels were produced. In the RLGS mixing gels, spots will be shown as enhanced if the corresponding clone is present in the pool of clones mixed with the genomic DNA. The determination of the plate, row, and column of the mixing gels in which the RLGS spot of interest is enhanced indicates the address of the unique clone in which the corresponding RLGS fragment was cloned.

Table 2. shows the sequence information for 151 unique clones identified by using the RLGS mixing gels. We compared the spot positions in this study to the master RLGS profiles in previous work (Costello *et al.*, 2000) and the methylated RLGS spots found in various tumor types in previous literatures. Costello *et al.* (2000) have shown that 4C06, 3B71, 6D05, and 4B84 (3D60, 3C02, 5E12, and 4B05 in Master RLGS) were CpG islands affected in at least three different tumor types: 4C06 in colon carcinoma, glioma, and acute myeloid leukemia; 3B71 in colon carcinoma, head and neck squamous cell carcinoma, and acute myeloid leukemia; 6D05 in breast carcinoma, glioma, and acute myeloid leukemia; 4B84 in breast carcinoma, colon carcinoma, and glioma. They have also shown that 3B80, 6C26, 4B51 and 5D84 (3C17, 5D13, 3C67 and 4F55 in Master RLGS) were affected at a high frequency in one tumor type but infrequently in others, thus suggesting that some CpG-island targets are methylated in a tumor-type specific manner while others are shared by multiple tumor types. In addition, the methylation of 4C55, 5C10, 5C30, 2C46, 3B43, 3B79 and 3B13 have been found in hepatocellular carcinoma (Nagai *et al.*, 1994); the methylation of 6D27 and 3D85 in lung carcinoma (Smiraglia *et al.*, 2001; Dai *et al.*, 2001); the methylation of 4B63 and 3D85 in acute myeloid leukemia (Rush *et al.*, 2001); the methylation of 2B16 in glioma (Nakamura *et al.*, 1997); the methylation of 6C26, 3D85 and 5B06 in chronic lymphocytic leukemia (Rush *et al.*, 2004); the methylation of 4A44 and 3B13 in renal cell carcinoma (Cho *et al.*, 1998). Aberrant methylation of CpG islands containing the promoters of cancer-related genes is often associated with transcriptional inactivation (Baylin *et al.*, 1998; Jones and Laird, 1999). The methylated CpG islands in various tumors cloned here may be novel epigenetic targets for the corresponding tumors, because no information on them has been found in literatures. Thus, it is required to determine whether the methylation events reported here have an impact on transcription of those genes.

In conclusion, we assessed sequence information for *Not*I sequences on a standard RLGS gel. We cloned 1,023 individual *Not*I-linked sequences matched to CpG islands of human genome, showing a high frequency of occurrence within CpG islands near the 5' ends of known genes rather than within CpG islands near the 3' ends or intra-genic regions. This information is available for the identification of gene-associated methylation events. We also provide 151 *Not*I sequences onto a standard RLGS gel with previous methylation events from several type of tumor. Therefore, our sequence information may be very useful to identify novel epigenetic targets in many tumor types.

Acknowledgments

This work was fully supported by FG05-21-01 of the 21C Frontier Functional Human Genome Project from the Ministry of Science & Technology of Korea.

References

Baylin, S.B., Herman, J.G., Graff, J.R., Vertino, P.M., and Issa, J.P. (1998). Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv. Cancer Res.* 72, 141-196.

Brown, R. and Strathdee, G. (2002). Epigenomics and epigenetic therapy of cancer. *Trends Mol. Med.* 8, S43-S48.

Costello, J.F., Fruhwald, M.C., Smiraglia, D.J., Rush, L.J., Robertson, G.P., *et al.* (2000). Aberrant CpG-island methylation has non-random and tumor-type-specific patterns. *Nat. Genet.* 24, 132-138.

Cho, M., Konishi, N., Yamamoto, K., Inui, T., Kitahori, Y., Nakagawa, Y., Uemura, H., Hirao, Y., and Hiasa, Y. (1998). Genomic Aberrations in Renal Cell Carcinomas Detected by Restriction Landmark Genomic Scanning. *Eur. J. Cancer.* 34, 2112-2118.

Dai, Z., Lakshmanan, R.R., Zhu, W.G., Smiraglia, D.J., Rush, L.J., Fruhwald, M.C., Brena, R.M., Li, B., Wright, F.A., Ross, P., Otterson, G.A., and Plass, C. (2001). Global Methylation Profiling of Lung Cancer Identifies Novel Methylated Genes. *Neoplasia* 3, 1-10.

Dai, Z., Weichenhan, D., Wu, Y.Z., Hall, J.L., Rush, L.J., Smith, L.T., Raval, A., Yu, L., Kroll, D., Muehlisch, J., Fruhwald, M.C., de Jong, P., Catanese, J., Davuluri,

R.V., Smiraglia, D.J., and Plass, C. (2002). An AscI boundary library for the studies of genetic and epigenetic alterations in CpG islands. *Genome Res.* 12, 1591-1598.

Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H., and Mukai, T. (1991). A genomic scanning method for higher organisms using restriction sites as landmark. *Proc. Natl. Acad. Sci. USA* 88, 9523-9527.

Hayashizaki, Y., Hirotsune, S., Okazaki, Y., Hatada, I., Shibata, H., Kawai, J., Hirose, K., Watanabe, S., Fushiki, S., and Wada, S. (1993). Restriction landmark genomic scanning method and its various applications. *Electrophoresis* 14, 251-258.

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

Jones, P.A. and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. *Nat. Rev. Genet.* 3, 415-428.

Kim, Y.S., Yoo, H.S., Lee, K.T., Goh, S.H., Jung, J.S., Oh, S.W., Baba, M., Yasuda, T., Matsubara, K. and Nagai, H. (2000). Detection of genetic alternations in the human gastric cancer cell lines by two-dimensional analysis of genomic DNA. *Int. J. Oncol.* 17, 297-308.

Kuick, R., Asakawa, J., Neel, J.V., Kodaira, M., Satoh, C., Thoraval, D., Gonzalez, I.L., and Hanash, S.M. (1996). Studies of the inheritance of human ribosomal DNA variants detected in two-dimensional separations of genomic restriction fragments. *Genetics* 144, 307-316.

Laird, P.W. (2003). The power and the promise of DNA methylation markers. *Nat. Rev. Cancer* 3, 253-266.

Miwa, W., Yashima, K., Sekine, T., and Sekiya, T. (1995). Demethylation of a repetitive DNA sequence in human cancers. *Electrophoresis* 16, 227-232.

Momparler, R.L., Bouffard, D.Y., Momparler, L.F., Dionne, J., Belanger, K., and Ayoub, J. (1997). Pilot phase I-II study on 5-aza-2'-deoxycytidine (Decitabine) in patients with metastatic lung cancer. *Anticancer Drugs* 8, 358-368.

Nagai, H., Ponglikitmongkol, M., Mita, E., Ohmachi, Y., Yoshikawa, H., Saeki, R., Yumoto, Y., Nakanishi, T., and Matsubara, K. (1994). Aberration of genomic DNA in association with human hepatocellular carcinomas detected by 2-dimensional gel analysis. *Cancer Res.* 54, 1545-1550.

Nakamura, M., Konishi, N., Tsunoda, S., Hiasa, Y., Tsuzuki, T., Takemura, K., Kobitsu, K., and Sakaki, T. (1997). Genomic alterations in human glioma cell lines detected by restriction landmark genomic scanning. *J. Neurooncol.* 34, 203-209.

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., Ohsumi, T., Nomura, O., Takada, S., Kamiya, M., Sasaki, N., Matsuda, Y., Nishimura, M., Tagaya, O., Muramatsu, M., and Hayashizaki, Y. (1996). A genetic linkage map of the Syrian hamster and localization of cardiomyopathy locus on chromosome 9qa2.1-b1 using RLGS spot-mapping. *Nat. Genet.* 13, 87-90.

Okuizumi, H., Okazaki, Y., Ohsumi, T., Hanami, T., Mizuno, Y., Muramatsu, M., Hayashizaki, Y., Plass, C., and Chapman, V.M. (1995). A single gel analysis of 575 dominant and codominant restriction landmark genomic scanning loci in mice interspecific backcross progeny. *Electrophoresis* 16, 253-260.

Plass, C., Shibata, H., Kalcheva, I., Mullins, L., Kotelevtseva, N., Mullins, J., Kato, R., Sasaki, H., Hirotsune, S., Okazaki, Y., Held, W.A., Hayashizaki, Y. and Chapman, V.M. (1996). Identification of Grf1 on mouse chromosome 9 as an imprinted gene by RLGS-M. *Nat. Genet.* 14, 106-109.

Pohlmann, P., DiLeone, L.P., Cancella, A.I., Caldas, A.P., Dal Lago, L., Campos, O.Jr., Monego, E., Rivoire, W., and Schwartsmann, G. (2002).Phase II trial of cisplatin plus decitabine, a new DNA hypomethylating agent, in patients with advanced squamous cell carcinoma of the cervix. *Am. J. Clin. Oncol.* 25, 496-501.

Shibata, H., Yoshino, K., Muramatsu, M., Plass, C., Chapman, V.M., and Hayashizaki, Y. (1995). The use of restriction landmark genomic scanning to scan the mouse genome for endogenous loci with imprinted patterns of methylation. *Electrophoresis* 16, 210-217.

Smiraglia, D.J., Fruhwald, 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., Fruhwald, M.C., Dai, Z., Held, W.A., Costello, J.F., Lang, J.C., Eng, C., Li, B., Wright, F.A., Caligiuri, M.A., and Plass, C. (2001). Excessive CpG island hypermethylation in cancer cell lines versus primary human malignancies. *Hum. Mol. Genet.* 10, 1413-1419.

Smiraglia, D.J., Szymanska, J., Kraggerud, S.M., Lothe, R.A., Peltomaki, P., and Plass, C. (2002). Distinct epigenetic phenotypes in seminomatous and nonseminomatous testicular germ cell tumors. *Oncogene* 21, 3909-3916.

Rush, L.J., Dai, Z., Smiraglia, D.J., Gao, X., Wright, F.A., Fruhwald, M., Costello, J.F., Held, W.A., Yu, L., Krahe, R., Kolitz, J.E., Bloomfield, C.D., Caligiuri, M.A., and Plass, C. (2001). Novel methylation targets in de novo acute myeloid leukemia with prevalence of chromosome 11 loci. *Blood* 97, 3226-3233.

Rush, L.J., Raval, A., Funchain, P., Johnson, A.J., Smith, L., Lucas, D.M., Bembea, M., Liu, T.H., Heerema, N.A.,

Rassenti, L., Liyanarachchi, S., Davuluri, R., Byrd, J.C., and Plass, C. (2004). Epigenetic profiling in chronic lymphocytic leukemia reveals novel methylation targets. *Cancer Res.* 64, 2424-2433.

Zardo, G., Tiirikainen, M.I., Hong, C., Misra, A., Feuerstein, B.G., Volik, S., Collins, C.C., Lamborn, K.R., Bollen, A., Pinkel, D., Albertson, D.G., and Costello, J.F. (2002). Integrated genomic and epigenomic analyses pinpoint biallelic gene inactivation in tumors. *Nat.Genet.* 32, 453-458.