

Type of publication: Technical Reports

Title: Genome-wide profiling of promoter methylation in human

Authors

Izuo Hatada^{1,2*}, Masayuki Fukasawa^{1,2}, Mika Kimura^{1,2}, Sumiyo Morita^{1,2}, Kazuo Yamada³, Takeo Yoshikawa³, Sumitaka Yamanaka^{4, 5}, Chiaki Endo⁴, Akira Sakurada⁴, Masami Sato⁴, Takashi Kondo⁴, Akira Horii⁵, Toshikazu Ushijima⁶, and Hiroyuki Sasaki⁷

Affiliations

1. Laboratory of Genome Science, Biosignal Genome Resource Center, Department of Molecular and Cellular Biology, Gunma University, Maebashi 371-8511 Japan
2. PRESTO, Japan Science and Technology Corporation (JST), 4-1 Honmachi, Kawaguchi 332-0012, Japan
3. Lab. for Molecular Psychiatry, RIKEN Brain Science Institute, 2-1 Hirosawa, Wako 351-0198, Japan
4. Department of Thoracic Surgery, Institute of Development, Aging and Cancer, Tohoku University, Sendai, 980-8575, Japan
5. Department of Molecular Pathology, Tohoku University School of Medicine, Sendai, 980-8575, Japan
6. Carcinogenesis Division, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan.
7. Division of Human Genetics, Department of Integrated Genetics, National Institute of Genetics, Research Organization of Information and Systems, and Department of Genetics, School of Life Science, Graduate University for Advanced Studies, Mishima 411-8540, Japan

* Author to whom correspondence should be addressed.

Tel: 81-27-220-8057 Fax: 81-27-220-8110 E-mail: ihatada@showa.gunma-u.ac.jp

Abstracts

DNA methylation in the promoter region of a gene is associated with a loss of that gene's expression and plays an important role in gene silencing. The inactivation of tumor-suppressor genes by aberrant methylation in the promoter region is well recognized in carcinogenesis (Baylin et al., 1997; Ushijima 2005). However, there has been little study in this area when it comes to genome-wide profiling of the promoter methylation. Here we developed a genome-wide profiling method called Microarray-based Integrated Analysis of Methylation by Isoschizomers (MIAMI) to analyze the DNA methylation of promoter regions of 8,091 human genes. With this method, resistance to both the methylation-sensitive restriction enzyme *Hpa* II and the methylation-insensitive isoschizomer *Msp* I was compared between samples by using a microarray with promoter regions of the 8,091 genes. The reliability of the difference in *Hpa* II resistance was judged using the difference in *Msp* I resistance. We demonstrated the utility of this method by finding epigenetic mutations in cancer. Aberrant hypermethylation is known to inactivate tumour suppressor genes. Using this method, we found that frequency of the aberrant promoter hypermethylation in cancer is higher than previously hypothesized. Aberrant hypomethylation is known to induce activation of oncogenes in cancer. Genome-wide analysis of hypomethylated promoter sequences in cancer demonstrated low CG/GC ratio of these sequences, suggesting that CpG-poor genes are sensitive to demethylation activity in cancer.

Text

Microarray-based methods of comparing differences in DNA methylation in the genome of two samples using methylation-sensitive restriction enzymes (Yan et al., 2001; Hatada et al., 2002) have two problems. The first is that the microarrays contain clones from libraries of CpG islands. CpG islands are CpG-rich regions of the genome originally thought to be associated with the 5' region of genes. There were

several approaches using CpG islands libraries for microarrays (Yan et al., 2001; Hatada et al., 2002; Heisler et al. 2005; Weber et al.). Although 60% of human genes have CpG islands in the promoter or first exon, more than 80% of all CpG islands have no relation to genes and are unlikely to regulate gene expression (Takai and Jones; 2002). To solve this problem, we used a microarray with 60-mer oligonucleotides derived from promoter regions of 8,091 human genes. DNA methylation in promoter regions is most important for the regulation of gene expression. The second problem is the risk of false positives resulting from restriction site polymorphisms and/or incomplete digestion of DNA. To resolve this issue, we developed a new method called Microarray-based Integrated Analysis of Methylation by Isoschizomers (MIAMI). We utilized resistance to a methylation-insensitive restriction enzyme, *Msp* I, to judge the false positive results for resistance to the methylation-sensitive isoschizomer *Hpa* II (Fig.1a). If two samples have a restriction site polymorphism at a *Hpa* II site and/or one of the samples has incomplete digestion at a *Hpa* II site, they will differ in resistance to *Hpa* II. However, in this case the resistance to methylation-insensitive *Msp* I at this site will also differ between samples because both enzymes recognize the same recognition site, CCGG. Therefore, we can treat such changes as false positives based on *Msp* I resistance.

We constructed a 60-mer-oligonucleotide microarray containing portions of *Hpa* II fragments located in promoter regions of 8,091 genes. We targeted the region from 600 base pairs upstream to 200 base pairs downstream of the transcriptional start sites for genes whose start sites were characterized on the basis of the National Center for Biotechnology Information (NCBI) annotation and/or Database of Transcriptional Start Sites (DBTSS). The probe nearest to a transcriptional start was selected on the condition that it doesn't have self complementarity (Primer 3 program, Rozen et al., 2000) and homology to the human genome (megaBlast program, Altschul et al., 1990). Microarrays were made using an ink-jet oligonucleotide synthesizer as described

(Hughes et al., 2001). Average position of the 8,091 probes was 36 base pairs upstream of the transcription start sites. Average GC content of the probes was 65%. All probes were included in the *Hpa* II fragments less than 600 base pairs. Average fragment length of the probe containing *Hpa* II fragments was 194 base pairs.

We defined resistance as reciprocal of sensitivity. Therefore, *Hpa* II-sensitive (cleavable) DNA and *Msp* I-sensitive (cleavable) DNA were amplified and used for calculating the *Hpa* II resistance and *Msp* I resistance, respectively. For *Hpa* II resistance, *Hpa* II-cleavable unmethylated DNA was amplified (I). *Hpa* II-cleaved DNA fragments were ligated to an adaptor and subjected to first PCR (Fig.1a). At this stage, only DNA fragments which had methylated internal *Hpa* II sites before the PCR retained *Hpa* II (*Msp* I) sites. Therefore, *Msp* I digestion made it impossible to amplify these methylated fragments. In the second main PCR, only unmethylated DNA fragments were amplified. Amplified unmethylated *Hpa* II-cleaved DNA fragments from two samples were labeled with Cy3 and Cy5, respectively, and cohybridized to the microarray with 60-mer oligonucleotides from promoter regions of 8,091 genes. After hybridization, the microarray was scanned and fluorescence intensities on a scanned image were quantified, corrected for background noise, and normalized with the software DNASIS Array (Hitachi Software Engineering). Spots with both Cy3 and Cy5 signals less than 0.001% of total signals were removed before analysis. *Hpa* II resistance (HR) was defined as $1/(\text{normalized } Hpa \text{ II intensity})$. Therefore, the ratio of *Hpa* II resistance of two samples (HR_B/HR_A) can be represented by $(\text{normalized } Hpa \text{ II intensity})_A/(\text{normalized } Hpa \text{ II intensity})_B$. For *Msp* I resistance, all *Msp* I -cleavable DNA (unmethylated plus methylated) was amplified (II). *Msp* I-cleaved DNA fragments were amplified and labeled the same as *Hpa* II-cleaved DNA fragments then cohybridized to another microarray with the same 8,091 genes. *Msp* I resistance (MR) was defined as $1/(\text{normalized } Msp \text{ I intensity})$. Therefore, the ratio of *Msp* I resistance of two samples (MR_B/MR_A) can be represented by $(\text{normalized } Msp \text{ I intensity})_A/(\text{normalized } Msp \text{ I intensity})_B$.

intensity)^A/(normalized *Msp* I intensity)^B. Details for all procedures are described in Supplementary information 1.

We applied the MIAMI method to a lung cancer cell line (1-87, abbreviated as LC) and a normal lung (abbreviated as C). Values for $\log(\text{HR}_{\text{LC}}/\text{HR}_{\text{C}})$ and $\log(\text{MR}_{\text{LC}}/\text{MR}_{\text{C}})$ are plotted on the x and the y-axis, respectively, of Fig. 1b. Various genes whose *Hpa* II resistance (HR) changed more than 5-fold ($\text{abs}[\log(\text{HR}_{\text{LC}}/\text{HR}_{\text{C}})] > \log 5$, areas more than $\log 5$ of the horizontal distance from the y axis) were selected as candidates (indicated by red and green circles in Fig. 1b and Supplementary Fig. 1). These genes were confirmed to differ in methylation between the cancer and the normal lung by combined bisulfite restriction (COBRA) analysis, with the genes indicated by red circles hypermethylated in the cancer cells and the genes indicated by green circles having no methylation-based changes (Fig. 2a). To characterize these false positives without changes in methylation (green circles), PCR was conducted followed by digestion with *Hpa* II to test for site polymorphisms. We found these false positives have site polymorphisms between the cancer and the normal lung (Fig. 2b). All these false positives were close to the regression line (yellow line in Fig. 1b and Supplementary Fig. 1) where ideal changes in *Hpa* II resistance and *Msp* I resistance are postulated to be equal. Therefore, we made threshold criteria with which to judge points located more than $\log 5$ of the horizontal distance from the regression line as altered genes. (Fig. 1b and Supplementary Fig. 1). Points located more than this distance right of the regression line were judged as hypermethylated and points located more than this distance left of the regression line were judged as hypomethylated. Using our criteria, we could neglect all false positives (green circles) and all genes meeting the criteria (red circles) had methylation changes, indicating our threshold is quite reasonable for selecting methylation-changed genes. Next we chose six genes which were located more than $\log 5$ of the horizontal distance from the regression line but less than $\log 5$ from the y-axis (indicated by orange and blue circles in Fig. 1b and

Supplementary Fig. 1). These genes can be judged as hypermethylated using our criteria but their changes in *Hpa* II resistance (HR) are less than 5-fold. COBRA analysis indicated that five of the six had actually methylation-based changes (Fig. 2a), indicating again our threshold criteria is useful for selecting methylation-changed genes (orange circles indicate positives and blue circles indicates false positive). Conventional, independent COBRA experiments using gene-specific primers confirmed 17 of 18 hypermethylations that were identified by integrated analysis of *Hpa* II resistance and *Msp* I resistance at a threshold of log5. This suggests that our empirical rate of false positives is 6%. We used our threshold criteria to calculate the ratio of changes and found that 5.7% of the promoters of the genes were hypermethylated and 0.6% were hypomethylated in lung cancer (Fig. 1c and Supplementary information 2). This frequency is much higher than a previous result in lung cancers (Yan et al., 2001), suggesting high sensitivity. Further improvement such as using linear amplification could make this method more efficient because it is expected that a proportion of fragments will not amplify and give no signal by PCR. Actually we removed 14% of spots with both Cy3 and Cy5 signals less than 0.001% of total signals for analysis to get reproducible results.

Next we analyzed the character of 5' sequences (from 1000 base pairs upstream to 200 base pairs downstream of the transcriptional start sites) for these hypermethylated and hypomethylated genes. Average ratio of CpG contents to GC contents (CG/GC) were calculated for hypermethylated, unchanged, and hypomethylated genes (Fig. 1d). We found hypomethylated genes had a low CG/GC ratio compared to genes without methylation change ($P=4.0 \times 10^{-15}$, Fig. 1d). However, the AT/TA ratio showed no such tendency (Fig. 1d). This suggests that CpG-poor genes are easily demethylated compared to CpG-rich genes. In other words, CpG-poor genes are more sensitive to demethylation activity than to CpG-rich genes. This could be explained by protection of demethylation activity by a methyl-CpG binding protein.

Promoters with a low density of methyl-CpGs bind MeCP-1 less strongly than those with a high density of methyl-CpGs (Boyes and Bird, 1992). Therefore, it is intriguing to speculate that CpG-poor genes are less protected by MeCP-1 from demethylation activity. Aberrant hypomethylation is related to the activation of oncogenes. Therefore, our finding of the unique character of hypomethylated genes will help us to understand the mechanism of carcinogenesis.

Aberrant hypermethylation is known to inactivate tumour suppressor genes. Among the hypermethylated genes we identified, further analysis of CIDEB and MLH3 were performed. CIDEB, Cell death-inducing DFFA-like effector b, activates apoptosis in mammalian cells (Inohara et al., 1998) and is located at 14q11 where LOH frequently occurs in lung cancers (Abujiang et al., 1998). MLH3, MutL Homolog 3, is a DNA mismatch repair gene associated with mammalian microsatellite instability (Lipkin et al., 2000). MLH1, from the same family, was frequently mutated in hereditary nonpolyposis colon cancer (Papadopoulos et al., 1994) and was involved in microsatellite instability DNA in colon cancers (Jager et al., 1997). A methylation-based analysis of an additional five lung cancer cell lines using COBRA revealed hypermethylation in five of six for CIDEB and two of six for MLH3 (Fig. 2c). RT-PCR analysis showed that expression was reduced in all hypermethylated cancers (Fig. 2c), indicating that the expression profile of the genes completely correlated with the methylation profile of the genes. Further methylation analysis were performed for CIDEB in primary tumours using COBRA. We found 71% (15/21) of primary lung cancers were hypermethylated in the promoter of CIDEB (Fig. 2d).

In conclusion, we conclude MIAMI is a powerful method for genome-wide profiling of promoter methylation in the human genome. This method is useful for epigenetic studies of cancers.

Acknowledgments

This work was supported in part by grants from the Japanese Science and Technology Agency (I.H.) and the Ministry of Education, Culture, Sports, Science and Technology of Japan (I.H.). We thank the Cancer Cell Repository (Institute of Development, Aging and Cancer, Tohoku University) for providing cancer cell lines and Miss. Asano for technical assistance.

Legends to figures

Fig. 1 Microarray-based Integrated Analysis of Methylation by Isoschizomers (MIAMI). (a) Schematic flowchart for the MIAMI method for comparison of sample A and sample B. Details were described in the text. (b) Application of the MIAMI method to a lung cancer cell line (1-87, abbreviated as LC) and a normal lung (abbreviated as C). Values for $\log(\text{HR}_{\text{LC}}/\text{HR}_{\text{C}})$ are plotted on the x-axis and $\log(\text{MR}_{\text{LC}}/\text{MR}_{\text{C}})$ are plotted on the y-axis. Green lines are located \log_5 of the horizontal distance from the y-axis. The regression line is in yellow and red lines are located \log_5 of the horizontal distance from this line. Points located more than this distance right of the regression line are judged as hypermethylated. Points located more than \log_5 of the horizontal distance left of the regression line are judged as hypomethylated. Genes indicated by red and green circles are located more than this distance from the y-axis. Hypermethylation was confirmed for genes indicated by red circles which were found to meet the criteria. Hypermethylation was not confirmed for genes indicated by green circles which did not meet the criteria although they were located \log_5 of the horizontal distance from the y-axis. Orange and blue circles meet the criteria but are less than \log_5 of the distance from the y-axis. Hypermethylation was confirmed for genes indicated by orange circles but not blue circles. (c) Summary of methylation changes in a lung cancer cell line (1-87, abbreviated as LC) compared to a normal lung (abbreviated as C). Methylation change (horizontal distance from the regression line) for each gene is

plotted on the y-axis. Red broken lines indicate threshold we used (\log_5). Genes are placed in order of position along X-axis. (d) Average ratio of CpG contents to GC contents (CG/GC) and average ratio of AT contents to TA contents (AT/TA) were calculated for hypermethylated, unchanged, and hypomethylated genes.

Fig. 2 Characterization of genes detected by MIAMI in a lung cancer. The PCR primers used are indicated in Supplementary information 3. (a) COBRA analysis of indicated genes. Genes indicated by red, orange, and blue circles met our criteria whereas indicated by green circles did not. COBRA analysis confirmed hypermethylation for genes indicated in red and orange and not genes indicated by green and blue. U indicates bands originating from unmethylated DNA. Other bands originated from methylated DNA. (b) Characterization of genes not meeting our criteria. *Msp I* (*Hpa II*) polymorphisms were detected by PCR followed by digestion with *Msp I*. All the genes not meeting the criteria have *Msp I* (*Hpa II*) polymorphisms. (c) COBRA and RT-PCR analysis of CIDEB and MLH3 genes in 6 lung cancer cell lines (1-87, RERF-LCMS, EBC-1, LK-2, VMRC-LCP, and LK79). U indicates bands originating from unmethylated DNA. Hypermethylation was observed for five of six lung cancer cell lines for CIDEB and two of six for MLH3. RT-PCR analysis showed that expression was reduced in all these hypermethylated cell lines. G3PDH was used for a control. (d) COBRA analysis of CIDEB genes in primary tumours. Eight adenocarcinomas, eight squamous cell carcinomas, and five small cell carcinomas were used for analysis. 71% (15/21) of primary tumours were hypermethylated. The present study was approved by the Ethics Committees of Tohoku University School of Medicine and Gunma University. Following a complete description of the research protocol, written informed consent was obtained from each participant.

References

Abujiang P, Mori TJ, Takahashi T, Tanaka F, Kasyu I, Hitomi S and Hiai H. (1998). *Oncogene.*, **17**, 3029-3033.

Altschul SF, Gish W, Miller W, Myers EW and Lipman DJ. (1990). *J. Mol. Biol.*, **215**, 403-410.

Baylin SB, Herman JG, Graff JR, Vertino, PM and Issa J-P. (1997). In *Advances in Cancer Research*. Vande Woude and Klein G (ed). Academic Press: San Diego, pp.141-196.

Boyes J and Bird A. (1992). *EMBO J.*, **11**, 327-333.

Hatada I, Kato A, Morita S, Obata Y, Nagaoka K, Sakurada A, Sato M, Horii A, Tsujimoto A and Matsubara K. (2002). *J. Hum. Genet.*, **47**, 448-451.

Heisler LE, Torti D, Boutrous PC, Watson J, Chan C, Winegarden N, Takahashi M, Yau P, Huang TH, Farnham PJ *et al.* (2005). *Nucleic Acids Res.*, **33**, 2952-2961.

Hughes TR, Mao M, Jones AR, Burchard J, Marton MJ, Shannon KW, Lefkowitz SM, Ziman M, Schelter JM, Meyer MR *et al.* (2001). *Nat. Biotechnol.*, **19**, 342-347.

Inohara N, Koseki T, Chen S, Wu X and Nunez, G (1998). *EMBO J.*, **17**, 2526-2533.

Jager AC, Bisgaard ML, Myrholm T, Bernstein I, Rehfeld JF and Nielsen FC. (1997). *Am. J. Hum. Genet.*, **61**, 129-138.

Lipkin SM, Wang V, Jacoby R, Banerjee-Basu S, Baxevanis AD, Lynch HT, Elliott RM and Collins FS (2000). *Nat Genet.*, **24**, 27-35.

Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, Haseltine

WA, Fleischmann RD, Fraser CM, Adams MD *et al.* (1994). *Science*, **263**, 1625-1629.

Rozen S and Skaletsky HJ. (2000). *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Krawetz S and Misener S (ed). Humana Press: Totowa, pp.365-386.

Takai D and Jones PA. (2002). *Proc. Natl. Acad. Sci. USA.*, **99**, 3740-3745.

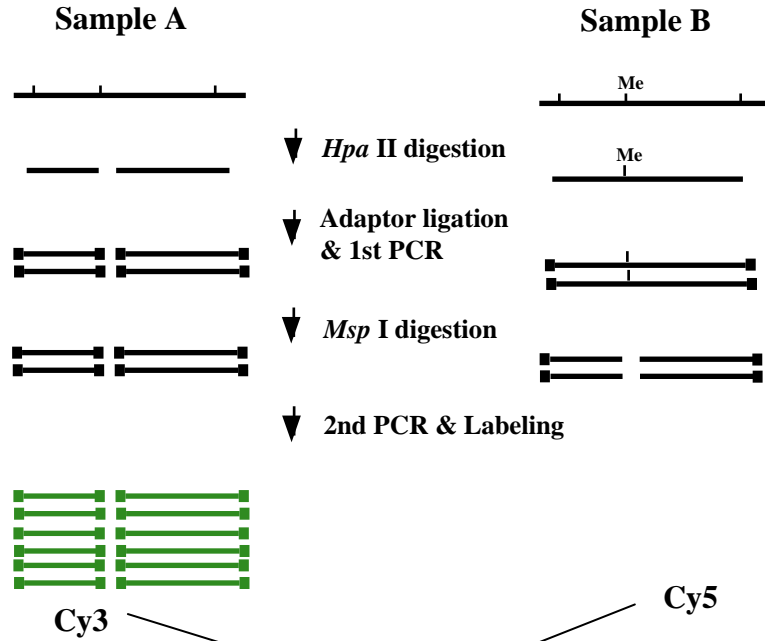
Ushijima T. (2005). *Nat Rev Cancer.*, **5**, 223-231.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL and Schubeler D. (2005). *Nat Genet.*, **37**, 853-862.

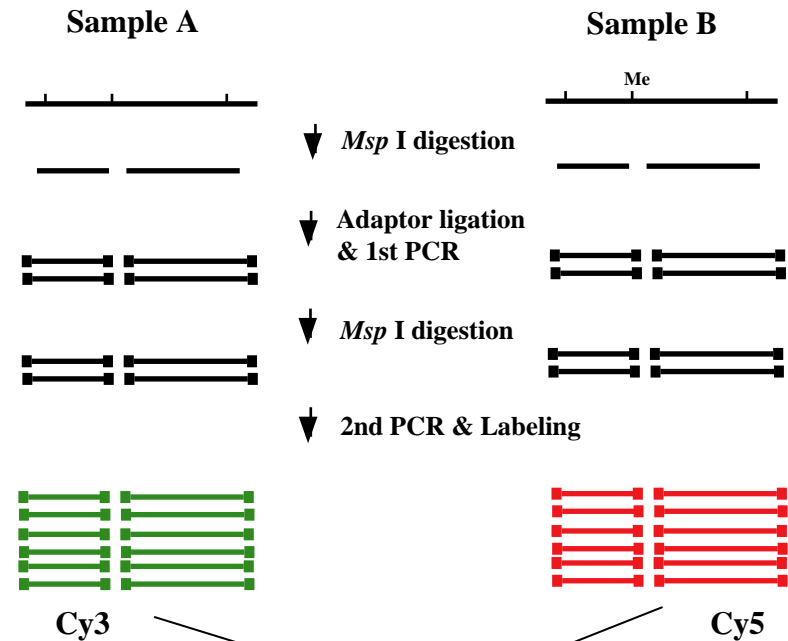
Yan PS, Chen CM, Shi H, Rahmatpanah F, Wei SH, Caldwell CW and Huang TH. (2001). *Cancer Res.*, **61**, 8375-8380.

Fig.1a

I. Amplification of unmethylated DNA (For *Hpa* II Resistance)



II. Amplification of all DNA (For *Msp* I Resistance)



$$\text{Resistance} = 1 / (\text{normalized Intensity})$$

Me: Methylation

↑ : *Hpa* II (*Msp* I) recognition sites

Scanning & Normalization

Scanning & Normalization

$$\frac{(\textit{Hpa II Resistance})_B / (\textit{Hpa II Resistance})_A}{(\text{normalized } \textit{Hpa II Intensity})_A / (\text{normalized } \textit{Hpa II Intensity})_B} =$$

$$\frac{(\textit{Msp I Resistance})_B / (\textit{Msp I Resistance})_A}{(\text{normalized } \textit{Msp I Intensity})_A / (\text{normalized } \textit{Msp I Intensity})_B} =$$

Fig.1b

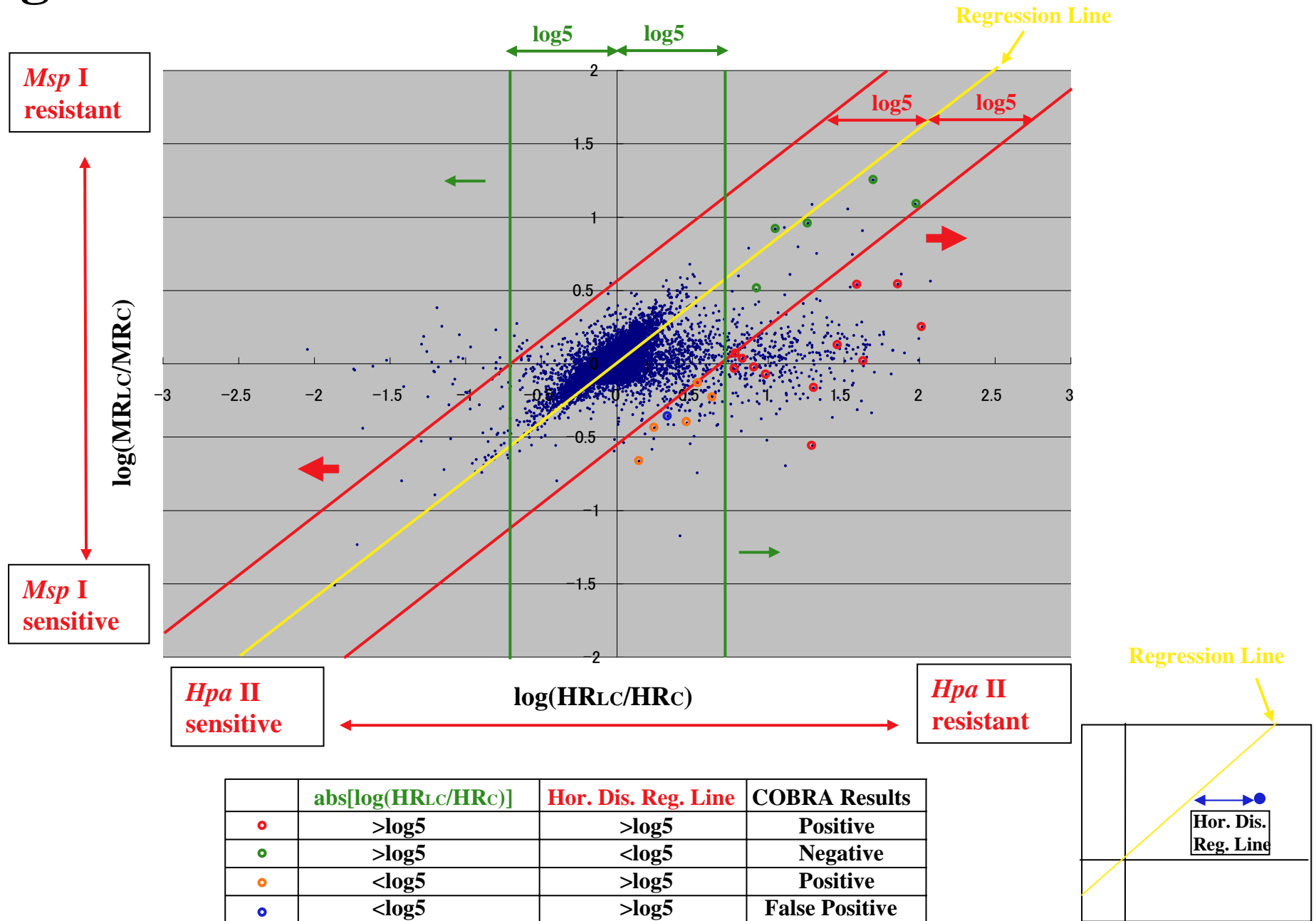


Fig.1c

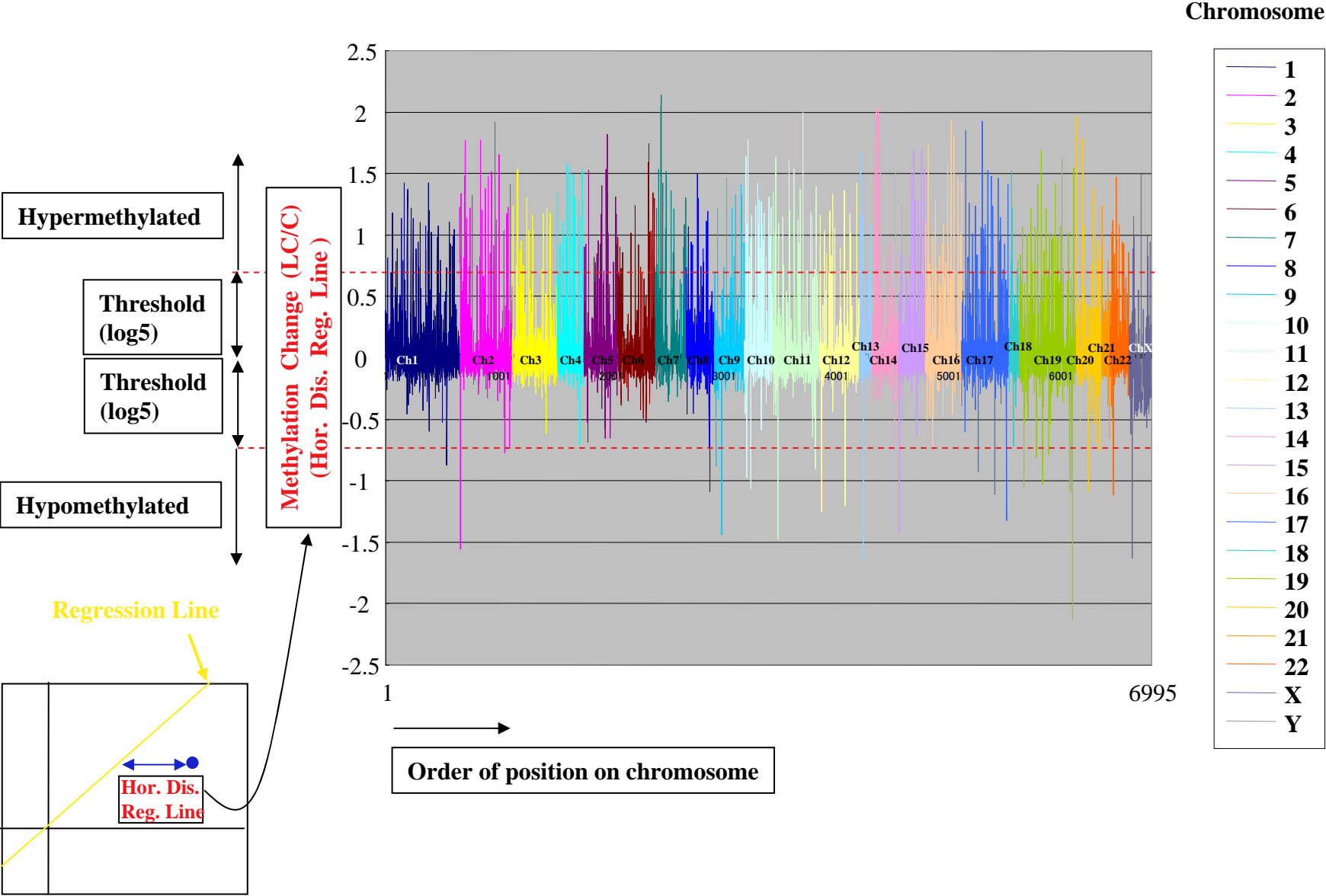


Fig.1d

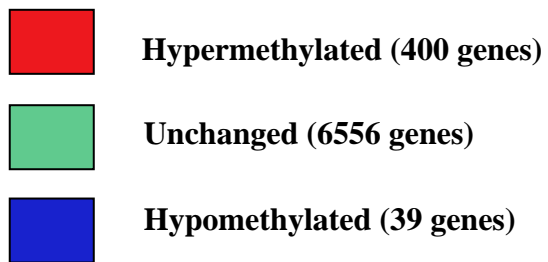
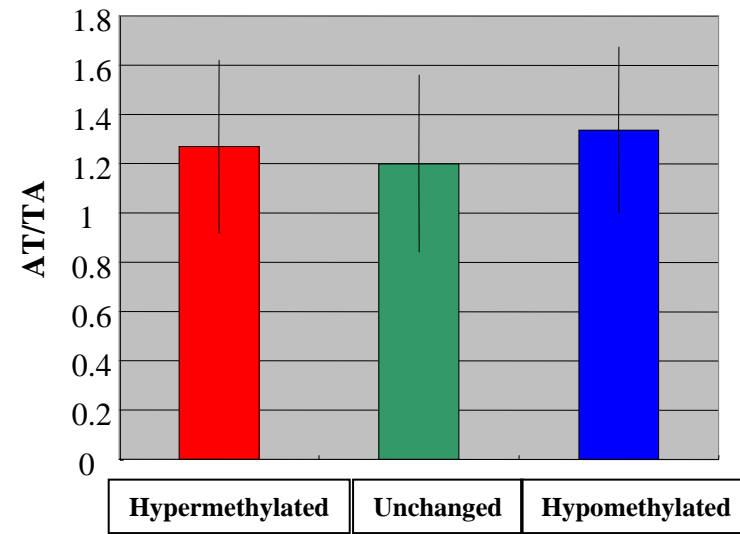
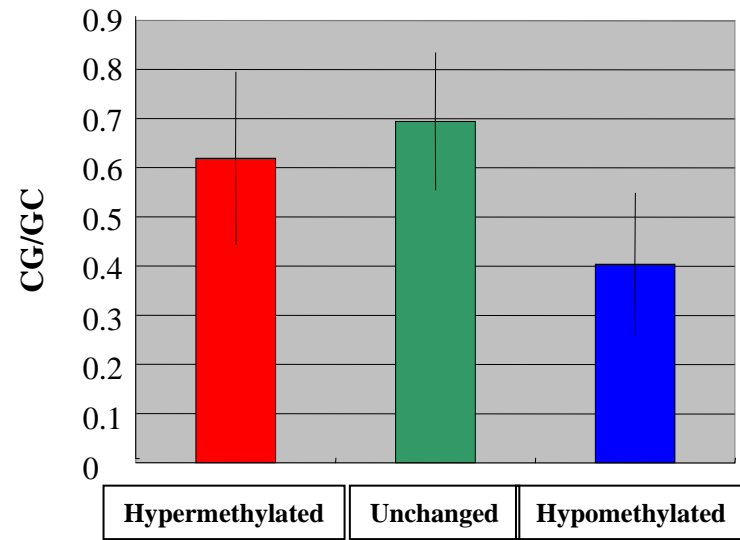


Fig.2a

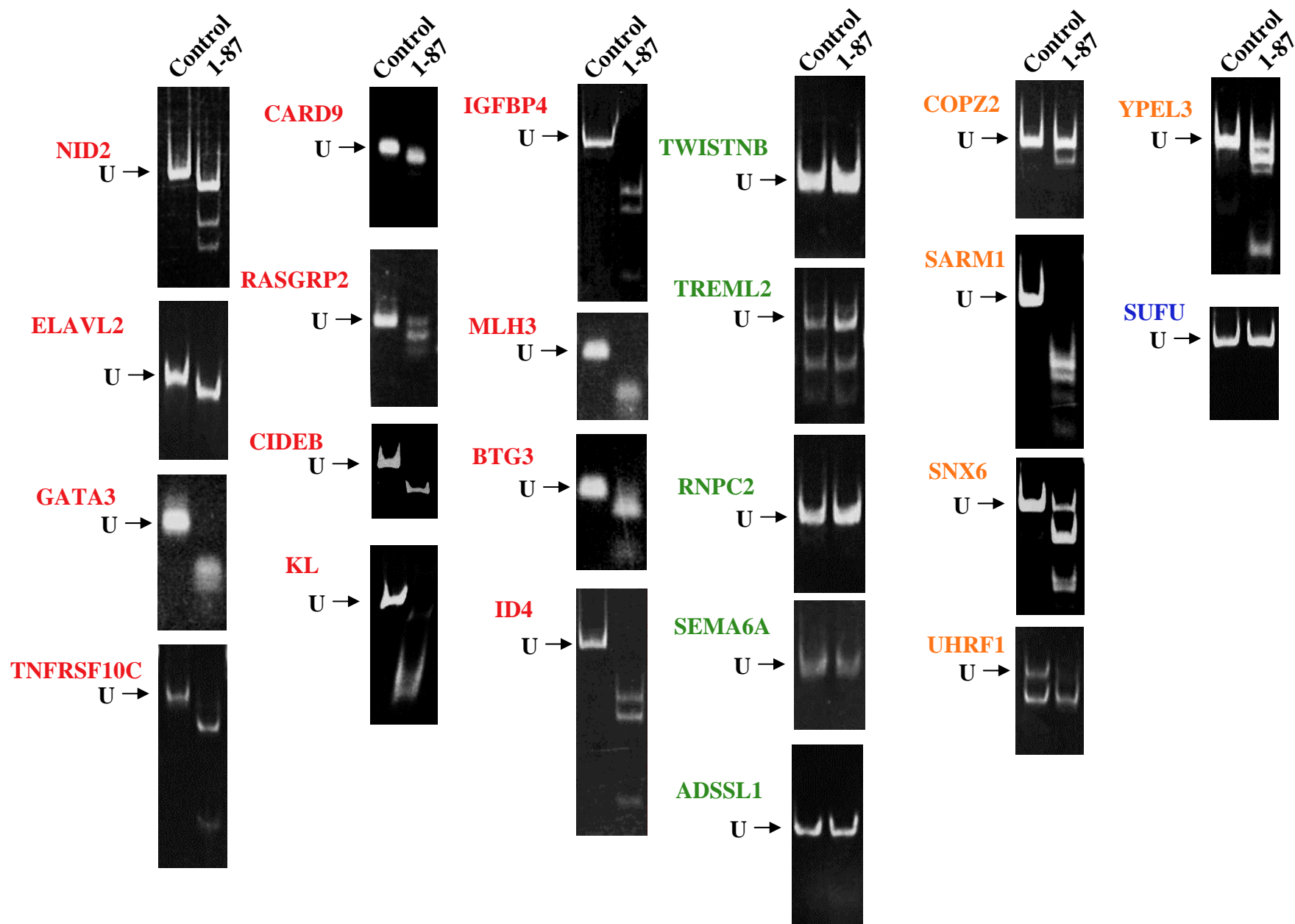


Fig.2b

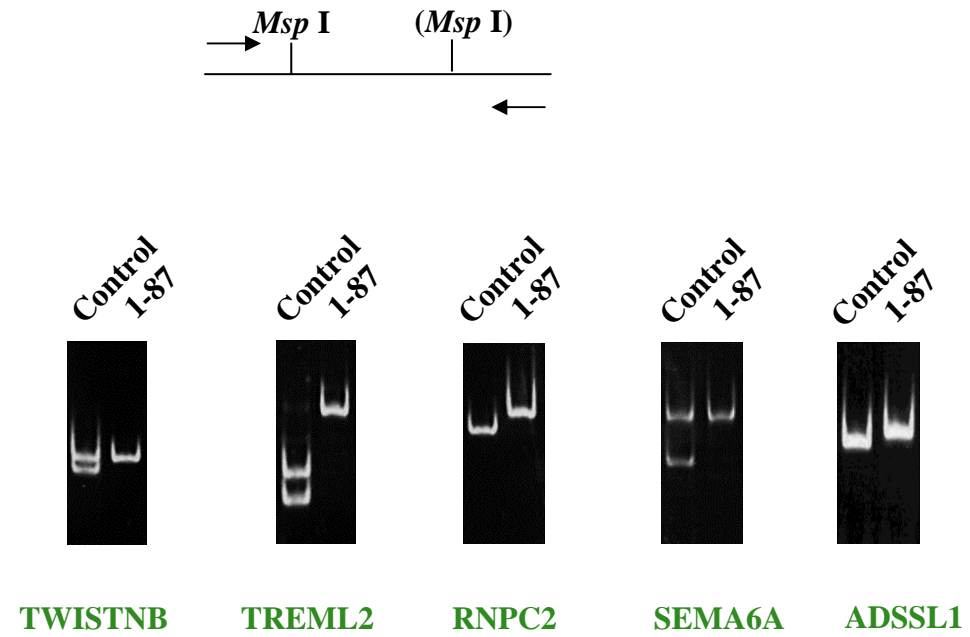


Fig.2c

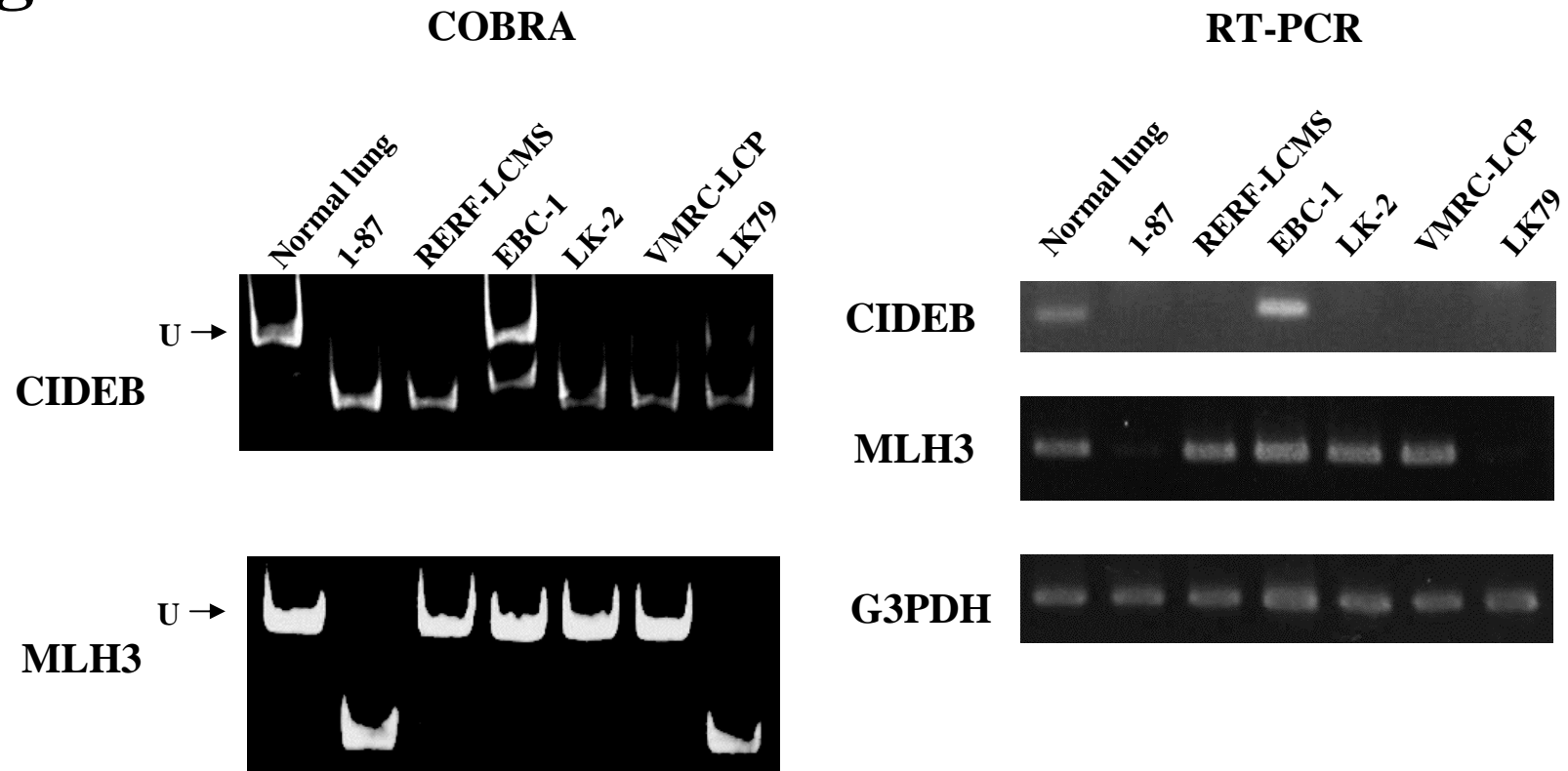


Fig.2d

