Direct Determination of NotI Cleavage Sites in the Genomic DNA of Adult Mouse Kidney and Human Trophoblast Using Whole-Range Restriction Landmark Genomic Scanning

Hirosumi IMOTO, Shinji HIROTSUNE, Masami MURAMATSU, Kiyoji OKUDA, Osamu SUGIMOTO, Verne M. CHAPMAN, and Yoshihide HAYASHIZAKI

Genome Science Laboratory, Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305, Japan, Department of Obstetrics and Gynecology, Osaka Medical College, 2-7 Daigakucho, Takatsuki, Osaka 569, Japan, and Molecular and Cellular Biology Department, Roswell Park Cancer Institute, Elm and Carton Streets, Buffalo, NY 14263, USA

(Received 29 August 1994)

Abstract

Restriction landmark genomic scanning (RLGS) is a method for visualizing restriction landmarks, employing direct labeling of restriction sites of genomic DNA and high-resolution two-dimensional electrophoresis. We determined the conditions for both the first and second dimensions of RLGS that define all of the restriction fragments which carry the NotI landmark. Using this system, we determined the number of cleavable NotI sites of genomic DNA from the mouse kidney (C57BL/6) and from the human placenta. The mouse and human genomes were cleaved at 2,380±80 sites (4,760±160 spots) and 3,240±110 sites (6,480±220 spots), respectively with NotI.

Key words: whole-range RLGS; NotI landmark; CpG island

Restriction landmark genomic scanning (RLGS) is based upon the direct labeling of restriction landmark sites in genomic DNA. These end-labeled fragments are reduced in size by digestion with a second endonuclease and the double-digest fragments are separated in the first dimension using agarose gel electrophoresis.1,2 The separated fragments are digested in-gel with a third restriction enzyme and the resulting fragments are separated in a polyacrylamide gel. The final distribution of end-labeled landmark fragments is analyzed by autoradiography. The total number of labeled spots and their distribution are determined by the enzyme combinations used and the electrophoretic conditions. However, it has been possible to identify routinely 1,500-2,000 landmark sites in mouse or human genomic DNA digested with combinations such as NotI, EcoRI, Mbol or NotI, PvuII, PstI. In the standard conditions used for many analyses, there could be fragments separated in the first dimension that were not cleaved with the third enzyme. Under these conditions, the end-labeled fragments would not be separated in the size range of the second-dimensional gel. Similarly, there could be some fragments that were not separated in the first-dimensional gel in the size range that could detect electrophoretic differences. As a consequence, the use of RLGS analyses of genomic DNA is confined to size fragments that could be readily resolved in a single gel.1,2

An estimate of the total number of landmark sites that are present in the genome is an important issue in the use of RLGS for several genome analysis applications such as genetic mapping3 and the identification of imprinted landmark sites that show parent allele-specific methylation.4 These determinations are critical for making a direct estimate of the proportion of these sites that have been examined either as genetic loci or as sites of genomic imprinting. Analysis of the number of CpG islands that may be associated with the promoter regions of functional genes indicates that there are approximately 3.7×10^4 islands in the mouse and 4.5×10^4 islands in the human genome.5 The NotI recognition sequence is CpG-rich and most likely is associated with CpG islands.6 There are an estimated 3,000 NotI sites in the mammalian genome based upon the frequency of NotI sites in genomic DNA sequence databases.

We have used a modification of the RLGS method to estimate the number of NotI landmarks directly in mouse kidney genomic DNA and in the genomic DNA from human placenta. The first- and second-dimensional separation of genomic fragments was optimized using different gel concentrations and third enzyme combinations. Using this approach, all of the NotI sites were identified.
We estimate that there were $2,380 \pm 80$ NotI sites in the mouse genome (4,760 RLGS spots) and 3,240$\pm 110$ NotI sites in human placental DNA (6,480 RLGS spots).

1. RLGS Profiles in Five Windows

Figure 1 shows RLGS patterns, representing the entire mouse genome. The RLGS profile was sub-divided into five windows (I-V) to capture all of the NotI landmarks. The first and second dimensions (1-D, 2-D) of the five windows cover the following molecular sizes: I (2700-470 bp, origin-90 bp), II (origin-2500 bp, origin-90 bp), III (500-40 bp, 500-10 bp), IV (2700-470 bp, 210-10 bp) and V (origin-2500 bp, 210-10 bp).

In windows I, II, IV and V, DNA fragments separated by the 1-D electrophoresis span from the origin to about 470 bp, with an average of several kilobase pairs, because the loaded sample was treated by NotI and BamHI. It is expectedly that there are no undetected spots located in the high-molecular region of 1-D since a 0.8% agarose gel electrophoretic separation in 1-D covers the origin from about 500 bp. A 4 or 10% polyacrylamide gel was used in the second dimension of these windows and the electrophoretic pattern of DNA markers showed that DNA fragments of 10 bp or less could be detected clearly without stacking (data not shown). The shortest fragments that are theoretically possible in the second dimension are $5'\text{NTCGGCCG}3'/3'\text{AGCGCCGGG}5'$ (underline represents $\gamma$P-labeled residues incorporated by filling reaction with Sequenase Ver 2.0), because HindIII was used as restriction enzyme C. Thus, DNA fragments less than 8 bp is not produced. Therefore, we concluded that all of the DNA fragments which were located in the region of more than 500 bp in the first dimension could be detected.

Window III was selected for the analysis of fragments in the first dimension of less than 500 bp, using a 1.5% agarose gel. A pilot experiment revealed that DNA fragments of less than 40 bp were separated in 60-cm long agarose disc gels (data not shown). CpG islands are preferentially cleaved when enzymes with GC-rich recognition sequences, such as NotI, are employed. HindIII ($5'\text{GANTC}3'$) was used to produce the RLGS patterns in Windows I, II, IV and V, avoiding the use of GC-rich 4-bp cutters such as HaeIII and HpaI for the in-gel digestion step. These enzyme sites are also preferentially located close to NotI sites on CpG islands and their use would result in the production of short DNA pieces of labeled NotI fragments in the second dimension. On the other hand, HaeIII and HpaI were used as restriction enzyme C in the analysis of Window III. The size of the DNA fragments separated in the first dimension of Window III were considerably shorter than the fragment sizes separated in the other windows, spanning 500-40 bp. Therefore, the cleavage by in-gel digestion with the enzyme recognizing 4-bp sequences including A and T was expected to be very infrequent in the DNA fragments located in Window III, resulting in insufficient separation of these fragments with almost all spots remaining on the diagonal parabola line. Furthermore, HaeIII is sensitive to methylation at the 5-position of the first cytosine residue of GGCC but not at the second cytosine residue. Almost all of the methylation in the mammalian genome occurs at the CpG dinucleotide. Thus, HaeIII is the most suitable enzyme for separation of spots in Window III and is least affected by methylation.

It was difficult to superimpose the patterns of the adjacent Windows I and IV with Window III as we used different enzymes for the in-gel digestion. To identify the border of these regions, we adopted genomic DNAs mixed with a 489-bp marker DNA which shows enhanced spots in these windows. In Fig. 1, these enhanced spots are indicated by arrowheads. Thus, we could avoid redundant counting of the same spots.
2. Total Number of NotI Sites in Human and Mouse Genome

The whole-range RLGS reported in this paper enables us to detect nearly all the NotI landmarks in the RLGS spots. NotI cleaved the mouse and human genomes at 2,380±80 and 3,240±110 sites, respectively. There are about 50 NotI sites on the 50-Mbp-long human chromosome 21, although this figure was based on cell line data, whose level of DNA methylation might differ from that of non-cultured tissues 5,10% of Cpg sites; (ref 9 and our unpublished data). These results provide an estimate of 3,000 NotI sites for the entire human genome which is in agreement with our results. The number of NotI sites in the mouse genome is 72% as many as that in the human genome, reflecting the expectation reported by Antequera et al. that about 20% of human CpG islands are absent from the homologous mouse genes.5

3. Direct Enumeration of NotI Sites in Mouse and Human Genomes by Whole-range RLGS

Our data indicated that almost all NotI sites throughout the genome were detected in five windows which were designed to cover all ranges in 1-D and 2-D and that the mouse and human genomes were cleaved at 2,380±80 and 3,240±110 sites with NotI. Conventional data for the number of NotI sites were compiled from sequence databases, such as GENBANK and EMBL. However, such sequence data do not reflect tissue-specific differences in methylation, which might affect cleavage by restriction enzymes. Whole-range RLGS enables the direct detection of those restriction sites which can be digested in a particular tissue.

4. Population of CpG Islands with NotI Sites

Our data demonstrated that the total numbers of NotI sites were 2,380±80 and 3,240±110 per haploid genome in the mouse and human genomes, respectively. It has been reported that 89% of all NotI sites are located in CpG islands.6 Therefore, it is calculated that the numbers of NotI sites located in CpG islands are about 2,140 and 2,850. On the other hand, Antequera et al. reported that the number of CpG islands could be calculated to be 37,000 and 45,000, respectively, from the data of the ratio of radioactivity incorporated into HapI tiny DNA fragments compared with those of total DNA.5 These data lead us to estimate that 5.7% and 6.4% of all CpG islands carry NotI sites that are unmethylated and cleaved in mouse kidney and human placenta.

5. Applications of the Direct Detection of Total Enzyme Cleavage Sites

RLGS methods have been used for genetic mapping and for detecting alterations in cancer.10 The methylation-sensitive properties of NotI have also been used to identify loci that change during differentiation of neuroepithelial cells9 and to identify allelic differences that show imprinted patterns of transmission.4 NotI landmark enzyme cleavage was used in each of these studies and it is critical to know the relative abundance of NotI sites when planning additional studies in each of these areas. For example, more than 1,000 loci have been identified and mapped in various mouse crosses. It is increasingly likely that the proportion of independent loci will diminish significantly in the analyses of additional enzyme combinations or different parental crosses. Similarly, the analysis of additional crosses and enzyme combinations for imprinted NotI sites may yield only a modest increase in the estimated number of imprinted loci since the total number of predicted imprinted genes is around 100.

Finally, the direct estimate of the number of NotI cleavage sites is important for establishing NotI linking libraries of either the mouse or human genome. We have previously used the NotI restriction trap method to isolate genomic NotI sites for the construction NotI linking and boundary libraries.12 The estimates from this study will guide the formation of arrayed NotI-linking libraries that can be used for the co-localization of NotI-linking clones with specific RLGS spots by an ordered mixing of clones with genomic DNA that is subsequently subjected to RLGS analyses.

Thus, whole-range RLGS provides us a very powerful tool to analyze the accessible restriction sites throughout the whole genome.

Acknowledgments: This work was supported by a Research Grant for the Human Genome Project from the Science and Technology Agency, by a Grant-in-Aid from the Human Genome Program and a Grant-in-Aid for Scientific Research on Priority Areas from The Ministry of Education, Science and Culture Japan to Y.H. This work was also supported a Research Grant for Aging and the Comprehensive 10-Year Strategy for Cancer Control from The Ministry of Health and Welfare to Y.H. This work was supported in part by U.S. Public Health Service Grant GM33160 to V.M.C.

References

1. Hatada, I., Hayashizaki, Y., Hirotsune, S., Komatsubara, H., and Mukai, T. 1991, A genomic scanning method for higher organisms using restriction sites as landmarks, Proc. Natl. Acad. Sci. USA, 88, 9523–9527.
2. Hayashizaki, Y., Hirotsune, S., Okazaki, Y. et al. 1993,
Restriction landmark genomic scanning method and its various applications, *Electrophoresis*, 14, 251–258.

3. Hayashizaki, Y., Hirotsune, S., Okazaki, Y. et al. 1994, A genetic linkage map of the mouse using Restriction Landmark Genomic Scanning (RLGS), *Genetics*, in press.

4. Hayashizaki, Y., Shibata, H., Hirotsune, S. et al. 1994, Identification of an imprinted U2af binding protein related sequence on mouse chromosome 11 using the RLGS method, *Nature Genetics*, 6, 33–40.

5. Antequera, F. and Bird, A. 1993, Number of CpG islands and genes in human and mouse, *Proc. Natl. Acad. Sci. USA*, 90, 11995–11999.

6. Lindsay, S. and Bird, A. P. 1987, Use of restriction enzymes to detect potential gene sequences in mammalian DNA, *Nature*, 327, 336–338.

7. Vanyushin, B. F., Tscher, S. G., and Belozersky, A. N. 1970, Rare bases in animal DNA, *Nature (London)*, 225, 948–949.

8. Ichikawa, H., Hosoda, F., Arai, Y. et al. 1993, A NotI restriction map of the entire long arm of human chromosome 21, *Nature Genetics*, 4, 361–366.

9. Kawai, J., Hirotsune, S., Hirose, K. et al. 1993, Methylation profiles of genomic DNA of mouse developmental brain detected by restriction landmark genomic scanning (RLGS) method, *Nucleic Acids Res.*, 21, 5604–5608.

10. Hirotsune, S., Hatada, I., Komatsubara, H. et al. 1992, New approach for detection of amplification in cancer DNA using restriction landmark genomic scanning, *Cancer Res.*, 52, 3642–3647.

11. Hirotsune, S., Shibata, H., Okazaki, Y. et al. 1993, Molecular cloning of polymorphic markers on RLGS gel using the spot target cloning method, *Biochem. Biophys. Res. Commun.*, 194, 1406–1412.

12. Hayashizaki, Y., Hirotsune, S., Hatada, I. et al. 1992, A new method for constructing NotI linking and boundary libraries using a restriction trapper, *Genomics*, 14, 733–739.
