DNA RESEARCH 2, 27–35 (1995)

Comparative Analysis of Mouse *Not*I Linking Clones with Mouse and Human Genomic Sequences and Transcripts

Christoph PLASS,[1] Jun KAWAI,[2] Iveta KALCHEVA,[1] Leslie DAVIS,[1]
Sachihiko WATANABE,[2] Yoshihide HAYASHIZAKI,[3] and Verne CHAPMAN[1,*]

Roswell Park Cancer Institute, Department of Molecular and Cellular Biology, Elm and Carlton Streets, Buffalo New York 14263-0001, USA,[1] Shionogi Institute for Medical Science, Shionogi & Co., Ltd, 5-1 Mishima 2-chrome, Settsu-shi 566, Japan,[2] and Genome Science Laboratory, Tsukuba Life Science Center, RIKEN, The Institute of Physical and Chemical Research, 3-1-1 Koyadai, Tsukuba City, Ibaraki 305, Japan[3]

(Received 9 February 1995)

Abstract

*Not*I cleavage sites are frequently associated with CpG islands that identify the 5′ regulatory sites of functional genes in the genome. Therefore we analyzed a sample of 22 *Not*I linking clones prepared from mouse brain DNA, to determine whether these mouse *Not*I site associated clones could be used for comparative analysis of mouse and human genomes by cross-reaction with both mouse and human genomic DNA and RNA in Southern and Northern hybridization. We further examined whether we could establish the identity of these clones with known genes by comparing the nucleotide sequences surrounding the *Not*I site with the GenBank database. We observed that 70% of the clones cross-hybridized with human DNA and that 4 of 11 tested clones (36%) detected a transcript in human HeLa cells RNA whereas 73% clones (8/11) detected transcripts in mouse RNAs from one or more organs. Single pass sequence analysis was successful on 16 of 19 clones. The GC content in these sequences was very high (48.8% to 73.8%) suggesting that 12 of 16 sequenced clones contained a CpG island. Three out of 19 clones showed significant similarity with previously analyzed mouse gene sequences in GenBank, including the mouse rRNA gene family, cathepsin and the scip POU-domain genes. In addition, two sequences showed significant similarity to the human and rabbit protein phosphatase 2A-β subunit and the human transforming growth factor -β. Thus, 5 of 16 clones showed homology with identified genes. These results and the recent work of using RLGS methods for genetic mapping indicate that *Not*I linking clones can be used to efficiently cross reference a comparative analysis of the mouse and human genomic maps.

Key words: *Not*I linking clone; mouse; comparative mouse/human mapping

1. Introduction

*Not*I restriction sites (GCGGCCGC) are preferentially located in CpG islands. These CpG islands are dispersed throughout the genome and they are generally associated with the 5′ region of genes.[1–3] Most housekeeping genes have a CpG island upstream of the transcription start region and as many as 40% of the tissue-specific genes may also have CpG islands associated with the 5′ region of the gene.[4] There are an estimated 37,000 to 45,000 CpG islands per haploid mouse and human genome, respectively.[3] The total number of *Not*I sites in the mouse genome has been estimated using whole-range RLGS analysis of *Not*I restriction landmarks that are identified by end labeling.[5] These authors estimated that there are 2350–2500 *Not*I sites in the mouse genome that were identified by *Not*I cleavage of kidney genomic DNA. A slightly greater number of sites were identified in human placental genomic DNA. These results suggest that approximately 10% of the CpG islands have a *Not*I site which is slightly lower than that suggested by others.[6]

The *Not*I restriction enzyme is methylation sensitive and distinguishes between methylated and unmethylated sequences. DNA methylation in mammals is restricted to the cytosine in the sequence 5′-CG-3′.[7] Normally, CpG islands associated with ubiquitously expressed housekeeping genes are unmethylated. However, CpG islands may be methylated as part of the repression of tissue-specific gene expression in different cell lineages[8–9] and in the case of genomic imprinting[10–12] or X-chromosome inactivation.[13]

*Not*I linking clones, containing the flanking sequences of a *Not*I site, are a useful tool in genetic and phys-

Communicated by Mituru Takanami

* To whom correspondence should be addressed. Tel. +716-845-5840, Fax. +716-845-8169

ical mapping because of their preferential location in CpG islands and their relatively infrequent occurrence in the genome. Several recent studies were performed using *Not*I linking clones derived from libraries specific for human chromosome 3,[6] human chromosome 21,[14] human chromosome 11[15] and human chromosome segment 22q11.[16] These clones are a useful tool in the analysis of functional sequences and the identification of useful candidates for positional cloning.

For the construction of *Not*I linking clones a method has been recently reported that uses a *Not*I restriction trapper to purify the DNA fragments with a *Not*I site for the construction of *Not*I linking libraries.[17] This new method is based on the ligation of DNA fragments with *Not*I restriction sites to a oligonucleotide linker which is covalently bound to the surface of latex beads (*Not*I restriction trapper). The selectivity of this procedure is based on the specific *Not*I restriction digest and results in very efficient purification.

The restriction landmark genomic scanning method using a methylation-sensitive restriction enzyme (RLGS-M) is a new method to follow differences in methylation during development or in different tissues.[18–19] By analyzing specific RLGS spots that differ between genotypes,[20] thousands of loci can be scanned in one experiment. Recently, Kawai et al.[21] used the RLGS-M to analyze methylation changes between mouse brain tissue and three mouse cell lines. In this study the restriction enzyme *Not*I was used as a restriction landmark. It was shown that 5–14% DNA loci were newly methylated in the established cell lines and 0.5–1.2% were demethylated compared to the original tissue DNA.

Previous work has demonstrated that *Not*I linking clones can be mixed with genomic DNA in the RLGS analysis to determine the alignment of clones with specific RLGS spots in the overall profile.[17,34] In other studies, as many as 50% of the RLGS spots vary in inter specific crosses which can be genetically mapped very efficiently in back crosses and RI strain analyses.[20] Thus, *Not*I linking clones provide a means for recovering physical clones of RLGS loci. We have selected a sample of these *Not*I linking clones to determine the utility of using these as comparative mapping probes between the mouse and human genomes that will ultimately tie the genetic mapping of RLGS spots to the identification of functional genes in both mouse and human genomic DNA. To accomplish this, we have made a single-pass sequence analysis on each end of the *Not*I site to estimate the proportion of *Not*I sites that are associated with previously characterized genes. We have also examined the hybridization of these clones by both Southern and Northern analysis to determine how well these clones will identify single copy genomic fragments and whether they identify expressed sequences.

2. Materials and Methods

2.1. Not*I linking clones*

The *Not*I linking library was constructed using the previously described method by Hayashizaki et al.[17] Briefly, genomic DNA from brain of C3H/HeN mice was isolated, digested with *Bam*HI and circularized with T4 DNA ligase. The resulting DNA was then digested with *Not*I and fragments containing *Not*I overhang were purified with the *Not*I restriction trapper.[17] The purified genomic fragments were cloned into Lambda ZapII vector (Stratagene). Inserts were excised with helper phage (R408) and recircularized to generate the subclones in pBluescript plasmid vector (Stratagene). Plasmid DNA was isolated using the Qiagen tip-100 columns (Qiagen Inc.) according to the manufacturer's instructions.

2.2. *Restriction landmark genomic scanning using methylation-sensitive restriction enzymes (RLGS-M)*

The RLGS-M was performed as described by Hayashizaki et al.[19] Briefly 7.5 μg genomic DNA were digested with the restriction enzyme *Not*I. The resulting restriction sites were end-labeled in a fill-in reaction using Sequenase Ver. 2.0 in the presence of [α-^{32}P]dCTP (6000 Ci/mmol) and [α-^{32}P]dGTP (3000 Ci/mmol). The end-labeled DNA was digested with *Bam*HI and separated in 1% agarose thin layer gels (first dimension). After electrophoresis, a *Hin*fI restriction digest of the DNA was performed in the gel. The DNA fragments were electrophoresed in the second dimension in 6% acrylamide gels. The gels were dried and exposed to Kodak XAR X-ray films at -70°C for several days.

To determine the correspondence of a *Not*I linking clone with a spot in the RLGS profile, DNA of the *Not*I linking clones was added separately to the genomic sample. The amount of the cloned DNA added was equivalent to 10 copies per haploid genome which produced a tenfold enhancement of the spots corresponding to the *Not*I linking clones. A sample of 22 *Not*I linking clones were used in this study: c5, c7, c19, c21, c25, c30, c59, c94, c103, c116, c156, c159, c171, c195, c198, c201, c211, c223, c233, c235, c255 and c265 (clones c94, c159 and c195 were used only for hybridization experiments).

2.3. *Automated sequencing*

Sequence reactions were carried out using the Applied Biosystems Taq DyeDeoxy Terminator Cycle Sequencing Kit. Reaction buffer and PCR profile were changed to accomplish the sequencing of the expected high GC content of the sequences. The reaction premix was modified by replacing the original buffer with the Vent polymerase buffer (New England Biolabs) supplemented with 3% DMSO (10x buffer: 100 mM KCl, 100 mM $(NH_4)_2SO_4$,

200 mM Tris pH 8.8, 20 mM $MgSO_4$, 1% Triton X-100 and 3% DMSO). Four units Amplitaq (Perkin-Elmer) were used per reaction. The reactions were run in a Perkin-Elmer 9600 thermal cycler. The denaturing time was 30 sec at 96°C, annealing was done at 50°C for 1 sec and the extension was carried out at 60°C for 4 min. A total of 25 cycles were performed.

A single pass of sequence reaction was done with each primer to minimize the costs. The frequency of sequencing errors in the control experiments was less than 2%. The range of sequence information varied with each clone. The sequences were added to the GenBank database.

2.4. *Sequence analysis*

Data bank search (GenBank, Release 78.0) was done with the programs BLASTN for nucleotide sequences or BLASTX for the putative amino acid sequences.[22] The search was performed using the combined sequence flanking both sides of the *Not*I sites. The simple sequence repeat in clone c19 was excluded from the sequence prior to the data bank search. DNA analysis for Sp1 binding sites, TATA and CCAAT boxes was done using the GCG V 7.3 program package.

2.5. *CpG island in the* Not*I linking clones*

To determine if the sequence contains a CpG island, a special computer program was written to perform the calculations of observed over expected (Obs/Exp) CpGs according to the formula of Gardiner-Garden and Frommer.[23] An Obs/Exp ratio of greater than 60% for a >200 bp region, combined with a GC content of >50% for the same region indicates the presence of a CpG island.

2.6. *Blot hybridization*

Mouse genomic DNA was isolated from C57BL/6Ros, DBA/2J and *Mus spretus*; human genomic DNA was isolated from a lymphoblast cell line using established protocols.[24] Five to ten micrograms genomic DNA was digested with *Eco*RI and electrophoresed on a 0.8% agarose gel. The DNA was transferred to Hybond-N^+ (Amersham) or Zetabind nylon filters (Cuno Inc.). Southern hybridization using [α-^{32}P]dCTP probes was performed in 0.5 M phosphate buffer as described earlier.[25] The blots were washed first under non-stringent conditions at 68°C. After the first exposure, the filters were washed under stringent conditions at 68°C twice for 30 min and re-exposed.

Total RNA from different mouse tissues and human HeLa cells were extracted according to the method by Chomczynski and Sacchi.[26] Poly(A)$^+$ RNA was purified from total RNA using OligoTex-dT30<Super> (Nippon Roche). The RNA was electrophoretically separated in 1% denaturing agarose gels and transferred to nylon filters as described.[24] The β-actin cDNA probe (obtained from Clontech) was used as a control.

3. Results

3.1. *Correspondence of* Not*I linking clones to spots in the RLGS profile*

In a recent study, Kawai et al.[21] used the RLGS-M method to study the methylation status of more than 2000 spots in mouse brain DNA and compared this with three different mouse cell lines. Two types of spots were identified. The first group of spots comprised DNA sequences containing consistently unmethylated *Not*I sites and the second group of spots containing sequences with methylated *Not*I sites in the cell lines. Individual *Not*I linking clones were mixed with the equivalent of 10 copies of haploid genomic DNA from the cerebrums of 8-week-old C3H/HeN mice to identify clones which correspond to specific spots of both classes. In the resulting RLGS profile, the intensity of the corresponding spots are enhanced. Seven *Not*I linking clones were identified that corresponded to specific RLGS spots in their analyses (c25, c116, c171, c212, c233, c255, and c265).[34] We have included six of these clones in a larger sampling of 22 *Not*I linking clones for our analyses. The methylation status of 15 of these clones which we used for the further analysis is summarized in Table 1.

3.2. *Sequence analysis of* Not*I linking clones*

Nucleotide sequences were identified for 16 of 22 *Not*I linking clones with both the M13 and the M13rev primers. No sequence information was obtained from clones c156, c235 and c265 using the primers M13, M13rev, T3 or T7. The sequencing data from the M13 and M13rev experiment were combined to represent the actual genomic sequences containing a *Not*I site within the sequence. We observed informative sequences from the different clones that ranged from 332 to 736 bp (Table 1). The GC content in these sequences was higher than 50%. The only exception was the sequences of clone c7 with a GC content of 48%, and the highest was found in clone c116 (73.8%). A $(CA)_n$-repeat with a repeat number of n=22 was located close to the *Not*I site in clone c19. A $(CAG)_n$ repeat (n=10) units was identified in clone c103.

3.3. *CpG islands in the* Not*I linking clones*

To determine if the sequence contains a CpG island, we calculated the ratio of observed over expected (Obs/Exp) CpG's according to the formula of Gardiner-Garden and Frommer[23] with a moving window of 100. A ratio greater than 0.6 over a sequence of 200 bp and a GC content over 0.5 indicates the presence of a CpG island. According to these criteria, 12 clones contain a CpG island (Table 1).

Table 1. Sequence analysis of *Not*I linking clones. The clone numbers of the individual clones and their methylation status in the cell lines is given (+ methylated; − unmethylated). The length of the sequence obtained is given in base pairs (bp). The GC content is calculated for the total sequence. The presence of a CpG island and the ratio of observed over expected (Obs/Exp) CpG's is calculated according to the criteria of Gardiner-Garden and Frommer.[23] A (+) indicates the presence of a CpG island within the sequenced region and (+/−) indicates that the size of the putative CpG island is smaller than 200 bp but might extend into the flanking region which was not sequenced.The number of potential Sp1 binding sites and the position of the CCAAT and TATA box in clone c211 is given. No CCAAT and TATA-box in tandem was found in the other clones. Clones c159 and c195 were not sequenced (n.s.).

Clone	Methylation status	bp	GC content	Obs/Exp	CpG island	Size of CpG island	No. of potential Sp1 binding sites	Position of CCAAT-TATA box
c5		598	72.5	1.02	+	>480	5	-
c7	+	673	48.8	0.65	+/−	180	2	-
c19	−	691	58.8	0.31	+/−	>80	1	-
c21		360	73.1	0.98	+	>260	1	-
c25	−	589	61.3	0.76	+/−	125	0	-
c30		504	56.0	0.58	+/−	120	0	-
c59		460	73.7	0.82	+	>350	5	-
c94	−	n.s						
c103		616	71.9	0.73	+	>375	3	-
c116	−	332	73.8	0.81	+	>225	2	-
c156		-						
c159	+	n.s.						
c171	−	710	60.3	0.93	+	>400	0	-
c195	−	n.s.						
c198	+	602	61.1	0.95	+	>340	0	-
c201	−	630	59.0	0.69	+	>260	1	-
c211	+	676	56.6	0.65	+	210	1	542-574
c223	+	531	71.1	0.74	+	>370	0	-
c233	−	594	60.2	0.74	+	250	2	-
c235		-						
c255	−	736	67.4	0.91	+	>510	4	-
c265	−	-						

In clones c7, c19, c25 and c30, the size of the CpG-island would be smaller than 200 bp (180, 80, 125 and 120 bp, respectively). In clones c19 and c 25, the CpG island might not be recognized due to limited sequence information. An Obs/Exp ratio over 0.6 and a GC content over 0.5 at either end of these sequences leaves the possibility of a CpG island. With the exception of clone c19, the *Not*I site is located in a CpG island or in a region with an Obs/Exp greater than 0.6 and a GC content over 0.5.

3.4. Sequence comparisons with GenBank

A comparison of *Not*I linking clones sequences with the GenBank database was performed using the BLAST program.[22] Three sequences were identified as identical to known gene sequences in the database (Table 2). Clone c19 contains exon1 of the mouse cathepsin gene (accession No: MMCATHD1). Clone c21 was identified as a member of the rRNA gene cluster (MUSRGEB3). The sequence comparison showed that in clone c21 a 277-bp *Not*I fragment is missing. Internal *Not*I fragments are eliminated from genomic sequences with more than one *Not*I site in a *Bam*HI restriction fragment during the establishment of *Not*I linking clones (see Hayashizaki et al.[17] for construction of *Not*I linking clone libraries). The sequence of clone c59 is derived from the mouse scip POU-domain protein (MUSPOUDOMD). The *Not*I site is located within the cDNA of the POU-domain protein.

Two sequences showed significant similarity to human nucleotide sequences and the corresponding amino acid sequence. Clone c103 is similar (189 positive nt/255 total nt) to the human transforming growth factor-beta (TGF-β) cDNA (HUMTGFBC). A high similarity was also observed (86%, 152/167) between the sequences of c103 and the mouse Vgr-1 cDNA (MUSVG1A). If we allow gaps in the alignment, the similarity to the Vgr-1 is even higher (92%), but the homology does not clearly identify the c103 sequence as the mouse Vgr-1. Restriction length variation was observed between *Mus spretus* and B6 using c103 as a probe to analyze Southern blots of genomic DNA cleaved with *Eco*RI (data not shown). We localized this variation to mouse chromosome 13 using

Table 2. Matches of *NotI* linking clone sequences to sequences in the GenBank. The similarity is given as identical matches per total aligned nucleotides.

Clone	Similarity on the nucleotide-level	Identities	Similarity on the amino acid level	Chromosomal location
c5	human protein phosphatase 2A β-subunit	87% (36/41)	human protein phosphatase 2A β-subunit	human chr#4
	Oryctolagus cuniculus protein phosphatase 2A1 B gamma subunit	91% (202/220)	Oryctolagus cuniculus protein phosphatase 2A1 B gamma subunit	-
c19	mouse cathepsin	97% (176/182)	mouse cathepsin	mouse chr#4
c21	rRNA gene cluster	97% (486/499)	-	-
c59	mouse scip POU-domain	99% (420/426)	mouse scip POU-domain	not mapped
c103	mouse Vgr-1	86% (152/176)	mouse Vgr-1	mouse chr #13
	human transforming growth factor-β (TGF-β)	74% (189/255)	human TGF-β	
c116	hamster nuclear factor-1-like protein	67% (43/64)	-	-
c171	rabbit alpha globin gene	80% (36/45)	-	-
c223	human protein kinase C	74% (41/55)	-	-
c255	mouse brain-specific inward rectifier potassium channel	100% (47/47)		-

the BSS backcross panel previously described.[20] In this instance, we compared the restriction patterns of a sample of six of these backcross mice that were either S/S or B/S for all genes on chromosome 13. We observed complete concordance in this sample between the variation between B6 and *M. spretus* and chromosome 13 (P for nonlinkage = 0.015). Several nonvariant fragments also hybridized with the c103 clone and it is possible that the variant fragments represent only a portion of the genomic sequences that cross-react with this probe. The localization of Vgr-1 is on chromosome 13 and the differences in the sequence may reflect the variability in the different mouse strain sources; the ICR mouse strain was used for the Vgr-1 sequence in the database, while clone c103 comes from C3H. On the other hand, we cannot exclude the possibility that c103 represents a new member in the bone morphogenetic (BMP) gene family. BMPs are a family of secreted signaling molecules which are structurally related to the TGF-β (see Kingsley[27] for a review). Clone c5 showed 87% (36/41) similarity to the human and rabbit cDNA sequences of the protein phosphatase 2A β-subunit (M78063 and U09355 respectively). No significant similarity could be identified with known mouse sequences.

The sequences of clones c255, c116, c171, and c223 showed similarity only on the nucleotide level to other sequences in the database (see Table 2). This similarity was most probably obtained randomly because of the high CG content of the sequences. No other sequence matches were identified between the GenBank database and the derived sequences for the rest of the clones.

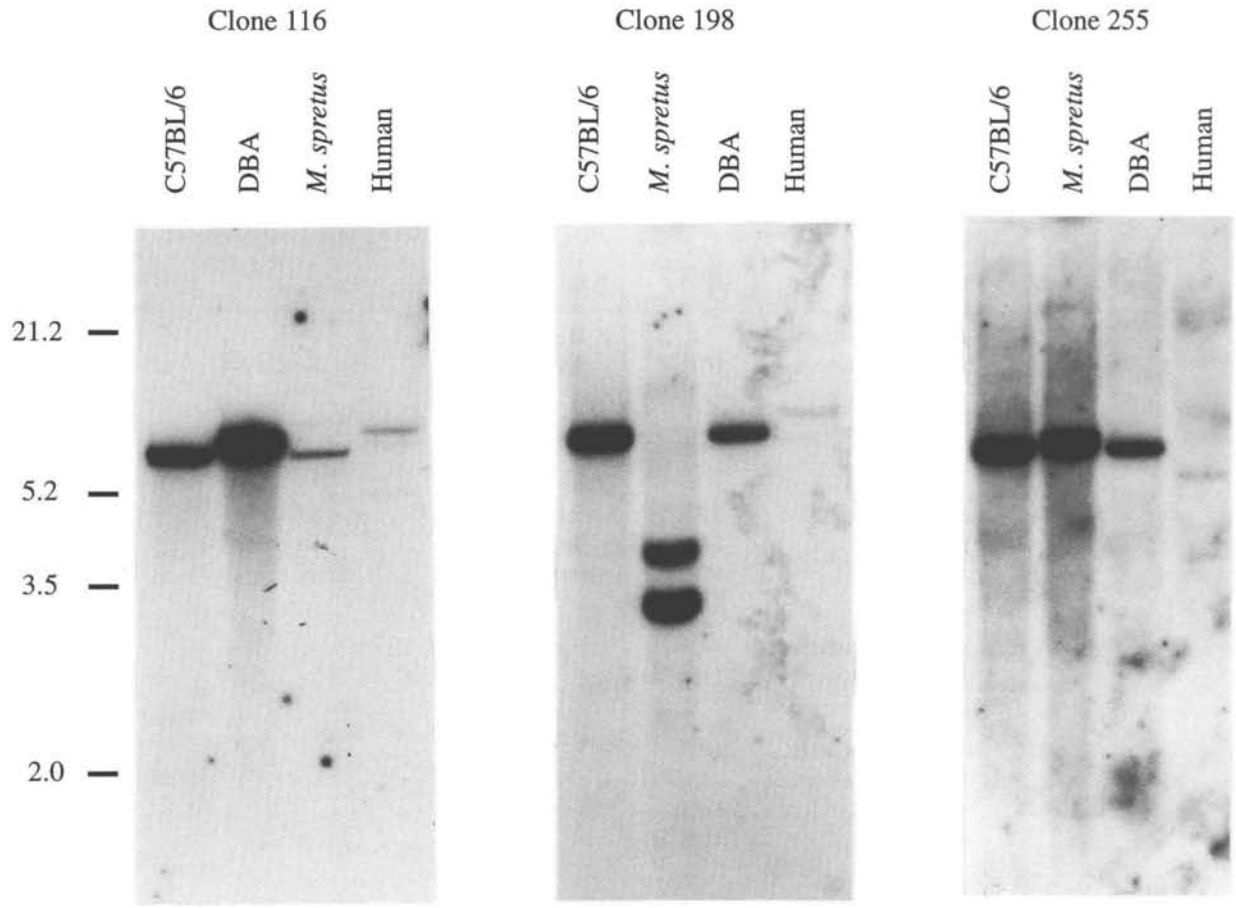

Figure 1. Southern blots of genomic mouse DNAs (C57BL/6, DBA and *M. spretus*) and human lymphoblast cell line genomic DNA digested with *Eco*RI and hybridized with clones c116, c198 and c255. The molecular size is given in kbp.

3.5. Sequence analysis for promoter elements

The promoter regions of housekeeping genes often lack the TATA box sequence normally found in genes transcribed by RNA polymerase II and thus appear to be related to the regulation of widely expressed genes. The Sp1 binding sites (G/C-box: GGGCGG or its reverse complement CCGCCC) are found upstream of the transcription start point instead of a TATA-box.[28] Only clone c211 has a CCAAT and TATA box in tandem separated by about 33bp. This region would be a good candidate for a promoter region of a tissue-specific gene. The Northern blot experiments (see below) revealed that this clone represents a gene which has tissue-specific (brain) expression. Five sequences (c25, c30, c171, c198 and c223) did not contain G/C-boxes. In all other sequences G/C-boxes could be identified. In three sequences (c5 and c59) five G/C-boxes were identified. The GC boxes are a common feature of CpG islands regardless of whether they are located 5′ or 3′ to the transcription start site.

3.6. Comparative analysis of mouse and human DNA

The utility of *Not*I linking clones as reagents for comparative genetic analysis of the mouse with the human genome was a central focus of this work. *Not*I linking clones are potentially useful for these analyses because they contain CpG islands which are located in the 5′ region of transcribed sequences and there is some expectation that these features should be evolutionarily conserved. To test this feature, *Not*I linking clones were used as probes for Southern hybridization analysis with mouse and human digested DNAs. We used stringent hybridization and washing conditions to overcome random cross-hybridization due to a high CG content. The results are summarized in Table 3. Three clones (c201, c235 and c265) contain highly repetitive elements which cross-hybridized with human DNA. Extra stringent washing conditions resolved the repetitive hybridization smear of c201 into distinct fragments in mouse and human. Figure 1 shows the three Southern hybridizations of *Eco*RI-digested DNAs for clones c116, c198 and c255. All three clones hybridized with human DNA. However, the in-

Table 3. Southern hybridization data of the *Not*I linking clones with human genomic DNA

Clone	Cross-hybridization with human DNA	Comments
c5	+	
c7	+	
c19	−	cathepsin
c21	+	rRNA gene family
c25	+	
c30	−	
c59	+	scip POU-domain
c94	+	not sequenced
c103	−	Vgr -1
c116	+	see Fig. 1
c156	+	
c159	−	not sequenced
c171	−	
c195	+	not sequenced
c198	+	see Fig. 1
c201	repetitive +after 72°C	
c211	−	
c223	+	
c233	+	
c235	repetitive	
c255	+	see Fig. 1
c265	repetitive	

Table 4. Northern hybridization of *Not*I linking clones on mouse liver, brain, muscle and skin poly(A)$^+$ -RNAs, and human HeLa cell poly(A)$^+$ RNA. A (+) indicates cross-hybridization and a (−) that no transcript was detected. The Northern hybridization data with mouse RNAs were reported in Watanabe et al.[34]

Clone	Mouse liver RNA	Mouse brain RNA	Mouse muscle RNA	Mouse skin RNA	Human (HeLa cell) RNA
c7	−	−	−	−	−
c25	+	+	+	+	+
c116	+	+	+	+	+
c159	−	−	−	−	−
c171	+	+	−	+	+
c198	−	+	+	+	−
c211	−	+	−	−	−
c223	−	−	−	−	−
c233	+	−	−	−	−
c255	−	+	−	−	−
c265	−	+	+	+	+

tensity of the hybridization signal in human DNA was weaker than in mouse DNA, indicating a lower sequence identity. Clone c198 shows *Eco*RI site polymorphism between *M. spretus* and the laboratory strains C57BL/6 and DBA/2. Out of 22 tested *Not*I linking clones, only six (c19, c30, c103, c159, c171 and c211) did not cross-hybridize with human DNA. Excluding the two mainly repetitive sequences, 14 out of 20 *Not*I linking clones (70% of the clones) contain conserved sequences which cross-hybridize with human DNA.

3.7. Northern hybridization of mouse and human RNA

A subset of 11 *Not*I linking clones were previously used as hybridization probes for Northern analysis[34] of poly(A)$^+$ RNAs isolated from four mouse tissues (liver, brain, muscle, and skin). We extended the analysis to human HeLa cell RNA. The data are summarized in Table 4. In the previous study, three clones (c7, c159 and c223) did not detect a transcript in any of the mouse RNAs. These clones do not detect a transcript in human HeLa cell RNA. Two clones (c25 and c116) detected transcripts in all four mouse tissues analyzed. These two clones detect also a transcript in HeLa cell RNA and are classified as part of housekeeping genes. The remaining clones c171, c198, c211, c233, c255 and c265) detected transcripts in the mouse in a tissue-specific way. Out of these clones, two (c171 and c265) hybridized also with human HeLa cell RNA. Altogether 8 out of 11 *Not*I linking clones detected transcripts in at least one of the tested mouse RNAs and four out of 11 *Not*I linking clones detected transcripts in HeLa cell RNA.

4. Discussion

Our analysis of *Not*I linking clones which were constructed using the restriction trapper indicates that these clones are useful for genetic analysis in both mouse and humans. Three-fourths (9/12) of the clones analyzed had *Not*I sites that were located in CpG islands and thus in potential promoter regions. The high degree of conserved sequences in these clones suggested by the cross-reaction of 14/22 clones with human genomic DNA in Southern analysis makes them a favorable tool for further comparative analysis. In this paper, we used *Not*I linking clones which were identified and characterized with the RLGS-M method. Previously Kawai et al.[21] analyzed the methylation changes in the *Not*I sites during the establishment of cell lines and during neural development.[8]

Several features suggest that the *Not*I clones can identify conserved sequences between the mouse and human genomes, making them highly useful for comparative analysis. First, five out of 16 *Not*I linking clones contained known mouse sequences or sequences that were similar to human gene sequences. This high ratio of about 30% known sequences corresponds to the findings of Adams et al.[29] who analyzed cDNA sequences derived from human brain RNA and found that 37% of the clones

could be identified based on matches with genes published in the databases. In sets of *Notl* linking clones from chromosome-specific libraries, no matches[30] or only 13% of known sequences were found.[6] The low frequency of known sequences in these studies might be due to the selection of specific genomic regions which might be less extensively analyzed by other research groups. Second, the high ratio of the *Notl* linking clones that detect transcripts in at least one analyzed mouse tissue (73%) make these clones excellent tools for comparative analysis. Sanford et al.[15] found that 17 out of 30 tested *Notl* boundary clones detected transcribed sequences. Allikmets et al.[6] analyzed 10 *Notl* linking clones in Northern blot experiments with eight different tissues and identified 9 out of these 10 clones which hybridized with RNA. Third, 72% of all clones cross-hybridized with human DNA indicating that their sequences contain conserved regions between mouse and human DNAs.

All housekeeping genes tested so far and about 40% of tissue-specific genes are associated with CpG islands.[4] In our survey, all tissue-specific genes identified by Northern blot experiments possess a CpG island. Two clones which were classified as housekeeping genes detected transcripts in human HeLa cell RNA. Four probes which hybridized in a tissue specific way in mouse RNA samples did not hybridize in HeLa cell RNA. Most probably these genes are also expressed in a tissue specific manner in humans and are repressed in HeLa cell lines.

The RLGS method analyzes restriction sites in genomic DNA as landmarks that are primarily associated with CpG islands. Thus, this method can detect a large number of landmarks that are associated with expressed genes. The use of methylation-sensitive enzymes allows the analysis of landmarks which are transmitted in a parent-specific manner (genomic imprinting, see Hayashizaki et al.[12]), in a tissue-specific manner,[8] or lineage-specific patterns such as X chromosome inactivation. The RLGS methods can be used to analyze genetic variation between *M. spretus* and C57BL/6 as either dominant or additive variation in backcrosses between these species.[20,31] The *Notl* landmarks can be physically analyzed by direct methods such as spot cloning with a restriction trapper enrichment of genomic DNA,[32] spot amplification with PCR methods,[33] or indirectly with the mixing of *Notl* linking clones with genomic DNA followed by RLGS analysis to identify enhanced labeling of spots.[17]

Assuming 2500 *Notl* sites per genome, this would be the number of *Notl* linking clones which cover one genome. This small number of clones is ideal to order in plates, rows and columns as an arrayed library. Analysis of the DNA from pools using the RLGS method should enhance the intensity of certain spots. If clones are presented only once in a pool it should be possible to deduce the plate location of a specific clone. By this means several hundred *Notl* linking clones can be assigned to specific spots which are identified in the RLGS analysis.

Acknowledgments: We would like to thank Alfred Cairo, Roswell Park Cancer Institute Biopolymer Facility for his help with the automated sequencing. We also thank Gregory Kohler for his assistance in analyzing the sequences. This work was supported in part by U.S. Public Health Service Grant GM33160 to VMC.

References

1. Bird, A. 1986, CpG-rich islands and the function of DNA methylation, *Nature,* **321,** 209-213
2. Bird, A. 1987, CpG-islands as gene markers in the vertebrate nucleus, *Trends Genet,* 3, 342-347.
3. Antequera, F. and Bird A. 1993, Number of CpG islands and genes in human and mouse, *Proc. Natl. Acad. Sci. USA,* 90, 11995-11999.
4. Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. 1992, CpG islands as gene markers in the human genome, *Genomics,* **13,** 1095-1107.
5. Imoto, H., Hirotsune, S., Muramatsu, M. et al. 1994, Direct determination of *Notl* cleavage sites in the genomic DNA of adult mouse kidney and human trophoblast using whole-range restriction landmark genomic scanning, *DNA Res.,* 1, 239-243.
6. Allikmets, R. L., Kaschuba, V. I., Pettersson, B. et al. 1994, *Notl* linking clones as a tool for joining physical and genetic maps of the human genome, *Genomics,* 19, 303-309.
7. Vanyushin, B. F., Tkacheva, S. G., and Belozersky, A. N. 1970, Rare bases in animal DNA, *Nature,* **225,** 948-949.
8. Kawai, J., Hirotsune, S., Hirose, K., Fushiki, S., Watanabe, S., and Hayashisaki, Y. 1993, Methylation profiles of genomic DNA of mouse developmental brain detected by restriction landmark genomic scanning (RLGS) method, *Nucleic Acids Res.,* **21,** 5604-5608.
9. Antequera, F., Boyes, J., and Bird, A. 1990, High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines, *Cell,* **62,** 503-514.
10. Stoeger, R., Kubicka, P., Liu, C. G. et al. 1993, Maternal-specific methylation of the imprinted mouse Igf2r locus identifies the expressed locus as carrying the imprinting signal, *Cell,* 73, 61-71.
11. Li, E., Bread, C, and Jaenisch, R. 1993, Role for DNA methylation in genomic imprinting, *Nature,* **366,** 362-365.
12. 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, 33-39.
13. Grant, S. and Chapman, V. 1988, Mechanism of X-chromosome regulation, *Ann. Rev. Genetics,* **22,** 199-233.
14. Ichikawa, H., Hosoda, F., Arai, Y. et al. 1993, A *Notl* restriction map of the entire long arm of human chromosome 21, *Nature Genetics,* 4, 361-366.
15. Sanford, S., Kim, B.-W., Deaven, L. L. et al. 1993, A human chromosome 11 *Notl* end clone library, *Genomics,*

15, 653–658.
16. ten Hoeve, J., Morris, C., Poustka, A., Groffen, J., and Heisterkamp, N. 1993, Isolation of *Not*I sites from chromosome 22q11, *Genomics*, **18**, 588–597.
17. Hayashizaki, Y., Hirotsune, S., Hatada, I. et al. 1992, A new method for constructing *Not*I linking and boundary libraries using a restriction trapper, *Genomics*, **14**, 733–739.
18. 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.
19. Hayashizaki, Y., Hirotsune, S., Okazaki, Y. et al. 1993, Restriction landmark genomic scanning method and its various applications, *Electrophoresis*, **14**, 251–258.
20. Hayashizaki, Y., Hirotsune, S., Okazaki, Y. et al. 1994, A linkage map of the mouse using restriction landmark genomic scanning (RLGS mapping), *Genetics*, **138**, 1207–1238.
21. Kawai, J., Hirose, K., Fushiki, S. et al. 1994, Systematic comparison of genomic DNA methylation patterns among mouse cell lines by restriction landmark genomic scanning (RLGS) method, *Mol. Cell. Biol.*, **14**, 7421–7427.
22. Altschul, S. F., Gish, W., Miller, W., Meyers, E. W., and Lipman, D. J. 1990, Basic local alignment search tool, *J. Mol. Biol.*, **215**, 403–410.
23. Gardiner-Garden, M. and Frommer, M. 1987, CpG islands in vertebrate genomes, *J. Mol. Biol.*, **196**, 261–282.
24. Sambrook, J., Fritsch, E. F., and Maniatis, T. 1989, Molecular cloning: A Laboratory Manual, 2nd Ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
25. Eckert, W. A., Plass, C., Weith, A., Traut, W., and Winking, H. 1991, Transcripts from amplified sequences of an inherited homogeneously staining region in chromosome 1 of the house mouse (Mus musculus), *Mol. Cell. Biol.*, **11**, 2229–2235.
26. Chomczynski, P. and Sacchi, N. 1987, Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction, *Analytical Biochem*, **162**, 156–159.
27. Kingsley, D. M. 1994, What do BMPs do in mammals? Clues from the mouse short-ear mutation, *Trends in Genetics*, **10**, 16–21.
28. Dynan, W. S. 1986, Promoters for houskeeping genes, *Trends Genet*, **2**, 196–197.
29. Adams, M. D., Soares, M. B., Kerlavage, A. R., Fields, C., and Venter, J. C. 1993, Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library, *Nature Genetics*, **4**, 373–380.
30. Zhu, Y., Cantor, C. R., and Smith, C. L. 1993, DNA sequence analysis of human chromosome 21 *Not*I linking clones, *Genomics*, **18**, 199–205.
31. Okuizumi, H., Okazaki, Y., Ohsumi ,T., Hayashizaki, Y., Plass, C., and Chapman, V. 1995, Genetic mapping of restriction landmark genomic scanning (RLGS) loci in the mouse, *Electrophoresis*, **16**, 233–240.
32. Hirotsune, S., Shibata, H., Okasaki, 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.
33. Suzuki, H., Kawai, J., Taga C., Ozawa, N., and Watanabe, S. 1994, A PCR mediated method for cloning spot DNA on restriction landmark genomic scanning (RLGS) gel, *DNA Res.*, **1**, 245–250.
34. Watanabe, S., Kawai, J., Hirotsune, S., Susuki, H., Hirose, K., Taga, C., Ozawa, N., Fushiki, S., Hayashizaki, Y. 1995. Accessibility to tissue-specific genes from methylation profiles of mouse brain genomic DNA, *Electrophoresis*, **16**, 218–226.