Comparative Analysis of Mouse NotI 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 Chapman¹∗

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, Japan³

(Received 9 February 1995)

Abstract

NotI 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 NotI linking clones prepared from mouse brain DNA, to determine whether these mouse NotI 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 NotI 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 NotI linking clones can be used to efficiently cross reference a comparative analysis of the mouse and human genomic maps.

Key words: NotI linking clone; mouse; comparative mouse/human mapping

1. Introduction

NotI 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 NotI sites in the mouse genome has been estimated using whole-range RLGS analysis of NotI restriction landmarks that are identified by end labeling.⁵ These authors estimated that there are 2350–2500 NotI sites in the mouse genome that were identified by NotI 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 NotI site which is slightly lower than that suggested by others.⁶

The NotI 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 lineages⁸⁻⁻¹⁰ and in the case of genomic imprinting¹⁰⁻⁻¹² or X-chromosome inactivation.¹³

NotI linking clones, containing the flanking sequences of a NotI site, are a useful tool in genetic and phys-
ical mapping because of their preferential location in CpG islands and their relatively infrequent occurrence in the genome. Several recent studies were performed using NotI linking clones derived from libraries specific for human chromosome 3,6 human chromosome 21,14 human chromosome 1115 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 NotI linking clones a method has been recently reported that uses a NotI restriction trapper to purify the DNA fragments with a NotI site for the construction of NotI linking libraries.17 This new method is based on the ligation of DNA fragments with NotI restriction sites to a oligonucleotide linker which is covalently bound to the surface of latex beads (NotI restriction trapper). The selectivity of this procedure is based on the specific NotI 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 NotI 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 NotI 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, NotI linking clones provide a means for recovering physical clones of RLGS loci. We have selected a sample of these NotI 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 NotI site to estimate the proportion of NotI 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. NotI linking clones

The NotI 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 BamHI and circularized with T4 DNA ligase. The resulting DNA was then digested with NotI and fragments containing NotI overhang were purified with the NotI 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 NotI. The resulting restriction sites were end-labeled in a fill-in reaction using Sequenase Ver. 2.0 in the presence of [α-32P]dCTP (6000 Ci/mmol) and [α-32P]dGTP (3000 Ci/mmol). The end-labeled DNA was digested with BamHI and separated in 1% agarose thin layer gels (first dimension). After electrophoresis, a HindIII 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 NotI linking clone with a spot in the RLGS profile, DNA of the NotI 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 NotI linking clones. A sample of 22 NotI 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 (NH4)2SO4,
from total RNA using Oligo-TdT30 (Nippon earlier). The blots were washed first under non-stringent conditions at 68°C twice and re-exposed. After the first exposure, the niters were washed under stringent conditions at 68°C. After the second wash, the niters were transferred to Hybond-ECL (Amersham) or Zetabind nylon filters (Cuno Inc.).

Southern hybridization using [α-32P]dCTP probes was performed in 0.5 M phosphate buffer as described earlier. The search was performed using the combined sequence flanking both sides of the NotI sites. The simple sequence repeat in clone c19 was excluded from the sequence prior to the data bank search. DNA analysis for Spl binding sites, TATA and CCAAT boxes was done using the GCG V 7.3 program package.

2.5. CpG island in the NotI linking clones

To determine if the sequence contains a CpG island, we calculated the ratio of observed over expected CpGs (Obs/Exp) according to the formula of Gardiner-Garden and Frommer. 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. Five to ten micrograms genomic DNA was digested with EcoRI 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-32P]dCTP probes was performed in 0.5 M phosphate buffer as described earlier. The blots were washed first under non-stringent conditions at 68°C. After the first wash, the blots 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. Poly(A)+ RNA was purified from total RNA using Oligo-TdT30<Super> (Nippon Roche). The RNA was electrophoretically separated in 1% denaturing agarose gels and transferred to nylon filters as described. The β-actin cDNA probe (obtained from Clontech) was used as a control.

3. Results

3.1. Correspondence of NotI linking clones to spots in the RLGS profile

In a recent study, Kawai et al. 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 NotI sites and the second group of spots containing sequences with methylated NotI sites in the cell lines. Individual NotI linking clones were mixed with the equivalent of 10 copies of haploid genomic DNA from the cerebriums 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 NotI linking clones were identified that corresponded to specific RLGS spots in their analyses (c25, c116, c171, c212, c233, c255, and c265). We have included six of these clones in a larger sampling of 22 NotI 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 NotI linking clones

Nucleotide sequences were identified for 16 of 22 NotI 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 NotI 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 NotI site in clone c19. A (CAG)n repeat (n=10) units was identified in clone c103.

3.3. CpG islands in the NotI linking clones

To determine if the sequence contains a CpG island, we calculated the ratio of observed over expected (Obs/Exp) CpGs according to the formula of Gardiner-Garden and Frommer with a moving window of 100 bp. 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 NotI 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. 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.).

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

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 c25, 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 NotI 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 NotI linking clones sequences with the GenBank database was performed using the BLAST program. 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 NotI fragment is missing. Internal NotI fragments are eliminated from genomic sequences with more than one NotI site in a BamHI restriction fragment during the establishment of NotI linking clones (see Hayashizaki et al.17 for construction of NotI linking clone libraries). The sequence of clone c59 is derived from the mouse scip POU-domain protein (MUSPOUDOMD). The NotI 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 EcoRI (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.

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

the BSS backcross panel previously described. 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 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.
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: GGCGGG or its reverse complement CCGCCC) are found upstream of the transcription start point instead of a TATA-box. 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 NotI linking clones as reagents for comparative genetic analysis of the mouse with the human genome was a central focus of this work. NotI 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, NotI 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 EcoRI-digested DNAs for clones c116, c198 and c255. All three clones hybridized with human DNA. However, the in-
tensity of the hybridization signal in human DNA was weaker than in mouse DNA, indicating a lower sequence identity. Clone c198 shows EcoRI site polymorphism between \textit{M. spretus} and the laboratory strains C57BL/6 and DBA/2. Out of 22 tested \textit{NotI} 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 \textit{NotI} linking clones (70\% of the clones) contain conserved sequences which cross-hybridize with human DNA.

3.7. \textit{Northern hybridization of mouse and human RNA}

A subset of 11 \textit{NotI} linking clones were previously used as hybridization probes for Northern analysis 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, c116, and c255) 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 \textit{NotI} linking clones detected transcripts in at least one of the tested mouse RNAs and four out of 11 \textit{NotI} linking clones detected transcripts in HeLa cell RNA.

4. Discussion

Our analysis of \textit{NotI} 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 \textit{NotI} 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 \textit{NotI} linking clones which were identified and characterized with the RLGS-M method. Previously Kawai et al.\textsuperscript{21} analyzed the methylation changes in the \textit{NotI} sites during the establishment of cell lines and during neural development.\textsuperscript{8} Several features suggest that the \textit{NotI} clones can identify conserved sequences between the mouse and human genomes, making them highly useful for comparative analysis. First, five out of 16 \textit{NotI} 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.\textsuperscript{29} who analyzed cDNA sequences derived from human brain RNA and found that 37\% of the clones

\begin{table}[h]
\centering
\caption{Southern hybridization data of the \textit{NotI} linking clones with human genomic DNA}
\begin{tabular}{|l|c|c|}
\hline
Clone & Cross-hybridization with human DNA & Comments \\
\hline
c5 & + & \\
c7 & + & \\
c19 & - & cathepsin \textit{rRNA gene family} \\
c21 & + & \\
c25 & + & \\
c30 & - & \\
c59 & + & scip POU-domain \\
c94 & + & not sequenced \\
c103 & - & Vgr -1 \\
c116 & + & see Fig. 1 \\
c156 & + & \\
c159 & - & not sequenced \\
c171 & - & \\
c198 & + & not sequenced \\
c201 & repetitive +after 72°C & \\
c211 & - & \\
c223 & + & repetitive \\
c233 & + & \\
c255 & repetitive & see Fig. 1 \\
c265 & repetitive & \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{Northern hybridization of \textit{NotI} 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.\textsuperscript{34}}
\begin{tabular}{|l|c|c|c|c|c|c|}
\hline
Clone & Mouse liver RNA & Mouse brain RNA & Mouse muscle RNA & Mouse skin RNA & Human (HeLa cell) RNA \\
\hline
c7 & - & - & - & - & \\
c25 & + & + & + & + & \\
c116 & + & + & + & + & \\
c159 & - & - & - & - & \\
c171 & + & + & - & + & + \\
c198 & - & + & + & + & \\
c211 & - & + & - & - & - \\
c223 & - & - & - & - & - \\
c233 & + & - & - & - & \\
c255 & - & + & - & - & \\
c265 & - & + & + & + & \\
\hline
\end{tabular}
\end{table}
could be identified based on matches with genes published in the databases. In sets of NotI linking clones from chromosome-specific libraries, no matches or only 13% of known sequences were found. 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 NotI linking clones that detect linking clones in at least one analyzed mouse tissue (73%) make these clones excellent tools for comparative analysis. Sanford et al. found that 17 out of 30 tested NotI boundary clones detected transcribed sequences. Allikmets et al. analyzed 10 NotI 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. 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.), in a tissue-specific manner, 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. The NotI landmarks can be physically analyzed by direct methods such as spot cloning with a restriction trapper enrichment of genomic DNA, spot amplification with PCR methods, or indirectly with the mixing of NotI linking clones with genomic DNA followed by RLGS analysis to identify enhanced labeling of spots.

Assuming 2500 NotI sites per genome, this would be the number of NotI 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 NotI 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

15. Sanford, S., Kim, B.-W., Deaven, L. L. et al. 1993, A human chromosome 11 NotI end clone library, Genomics,
15. 653–658.


