Comparative Genomic Approach Toward Species-Specific Imprinting

by

Kohji OKAMURA

A Dissertation

Presented to Department of Molecular Cell Biology,

Graduate School of Medicine, University of Tokyo

in Partial Fulfillment of the Requirements

for the Degree of Ph. D.

2003

Table of Contents

Abbreviations

AMD	allelic message display
B6	C57BL/6J
BAC	bacterial artificial chromosome
bp	base pair
BLCAP	bladder cancer-associated protein
cDNA	complementary deoxyribonucleic acid
DMGDH	dimethylglycine dehydrogenase
DMR	differentially methylated region
DMSO	dimethylsulfoxide
DNase	deoxyribonuclease
dNTP	deoxynucleoside triphosphate
DT-A	A domain of diphtheria toxin
EDTA	ethylenediaminetetraacetic acid
EST	expressed sequence tag
FISH	fluorescence *in situ* hybridization
%GC	percentage of G+C content
GPCR	G-protein-coupled receptor
HRH4	histamine H_4 receptor
I.M.A.G.E.	Integrated Molecular Analysis of Genomes and their Expression
JF	JF1/Msf
kb	kilobase (1000 DNA bases)
LINE	long interspersed transposable element
mRNA	messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
Obs/Exp CpG	ratio of observed versus expected CpG dinucleotides
ORF	open reading frame
OSBP	oxysterol binding protein
OSBPL1	oxysterol binding protein-like 1
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PH	pleckstrin homology
RFLP	restriction fragment length polymorphism
RNase	ribonuclease
RT-PCR	reverse transcription polymerase chain reaction
SDS	sodium dodecyl sulfate
SINE	short interspersed transposable element
SNP	single nucleotide polymorphism
STS	sequence-tagged site
Tris-HCl	tris(hydroxymethyl)aminomethane-hydrochloride
UTR	untranslated region

Abstract

Genomic imprinting is an epigenetic modification of the gamete or zygote that leads to exclusive or preferential expression of a specific parental allele in somatic cells of the offspring. Mouse *Impact* is a maternally imprinted or a paternally expressed gene encoding an evolutionarily conserved protein of unknown function. While this gene is expressed exclusively from the paternal allele, its human homologue *IMPACT* is not imprinted or expressed biallelically. For almost all imprinted genes, parent-of-origin-specific expression is conserved between mouse and human. However, this is not the case but serves as a good model to understand the molecular mechanism of genomic imprinting. To reveal the structural basis for the difference in allelic expression between the two species, I elucidated complete nucleotide sequences for both mouse *Impact* (38 kb) and human *IMPACT* (30 kb) by a unique nested deletion strategy. Sequence comparison revealed that the two genes share a well-conserved exon-intron organization but bear significantly different CpG islands. The mouse island lies in the first intron and contains characteristic tandem repeats. To examine the methylation status, a methylation-specific PCR assay was developed and it clearly demonstrated that this island serves as a differentially methylated region (DMR) consisting of a hypermethylated maternal allele and an unmethylated paternal allele. Intriguingly, this intronic island is missing from nonimprinted human *IMPACT*, whose sole CpG island spans the first exon, lacks any apparent repeats, and escapes methylation on both chromosomes. Subsequently, taking advantage of results of the human genome project, I found neighboring genes lying next to human *IMPACT*. One is the *HRH4* gene residing downstream of *IMPACT*, and the other is the *OSBPL1* gene upstream. Allelic expression analysis using mRNAs from peripheral blood leukocytes revealed that these human genes are also expressed biallelically like *IMPACT* is. Then I successfully found their mouse homologues, the *Hrh4* and *Osbpl1* genes, at the corresponding loci in the mouse genome. As for *Osbpl1*, I identified two transcripts, short *Osbpl1a* and long *Osbpl1b*, whose human homologues had already been reported. The two have their own promoters and have different tissue distribution and expressional quantities. Although *Osbpl1a* is expressed biallelically, the longer transcript, *Osbpl1b*, whose promoter is located near that of *Impact*, is a little preferentially expressed from the paternal allele. Because DNA methylation analyses gave no more DMRs in these mouse and human regions except for the CpG island and the promoter region of mouse *Impact*, the expression of mouse *Osbpl1b* may be due to the allele-specific chromatin structure of the promoter region of *Impact*. Apart from

this biased transcription, the *Impact* locus fails to show any evidence for imprinted cluster even though clustering into chromosomal domain is one of the characteristic features of such genes. Although two genes were also documented as solitary imprinted genes, there are several differences between *Impact* and the two. Since the two imprinted genes occur in introns of nonimprinted genes and have few introns, they were conceivably generated by as retrotransposons, which are often subject to imprinting. In contrast, *Impact* consists of 11 exons and is not embedded in an intron of another gene but having nearby genes that are not imprinted. It is unlikely that *Impact* was generated by a retrotranspositional event. Hence, this gene appears to have a unique structure as an isolated imprinted gene. Fortunately, the human orthologue *IMPACT* does not exhibit allele-specific expression. Therefore, the structural element unique to mouse *Impact*, such as the differentially-methylated intronic CpG island containing characteristic tandem repeats, may represent the basis for the species-specific imprinting. Sizes of introns of mouse *Impact* are relatively shorter than those of *IMPACT* throughout these genes. However, the first intron harboring the island is exceptional and the human *IMPACT* lacks a corresponding sequence to the mouse CpG island. These results suggest that the intronic DMR plays a crucial role in the imprinting of this gene. Thus, *Impact* would serve as an ideal model for the investigation of imprinted expression achieved by a unique mechanism independent of the regulation as a chromosomal domain.

Introduction

Mammals have a pair of genomes, one is maternal and the other is paternal. But not all the genes are expressed equally from both genomes. For instance, one of the X chromosomes in female cells is almost totally inactivated or silenced (Lyon 1961; Boumil and Lee 2001). It is also known that a small number of mammalian genes are expressed predominantly or exclusively from one of the two chromosomes. Selection of the expressed allele in such monoallelic expression can be either random or parent-of-origin dependent. The former type of monoallelic expression is observed in particular combinations of genes and cells, including odorant receptor genes in olfactory neurons (Chess *et al*. 1994) and genes for immunoglobulin (Alt *et al*. 1984) and several cytokines (Hollander *et al*. 1998) in lymphocytes. In contrast, the latter type of monoallelic expression occurs in almost all the mammalian cells. This parent-of-origin dependent gene expression is termed parental imprinting or genomic imprinting (Barlow 1997; Constancia *et al*. 1998; Tilghman 1999), which shows non-Mendelian modes of inheritance. The concept of "imprinting" was fist introduced by Crouse who studied chromosome behavior in a certain insect (Crouse 1960). The imprint is a mark established during gametogenesis for the cell to distinguish between maternal and paternal alleles. While genomic imprinting is observed not only in mammals but also in insects (Lloyd 2000) and plants (Baroux *et al*. 2002), it is most well characterized in mammals to provide evidence for a role of DNA methylation and chromatin structure in epigenetic marking of the chromosome. Nearly 50 mammalian imprinted genes have been discovered since 1991 (Barlow *et al*. 1991), nevertheless, why and how some mammalian genes are imprinted still remains largely elusive.

It is well known that both maternal and paternal genomes are required for mammalian development (Solter 1988), suggesting that they are functionally nonequivalent. This is due, in part, to the existence of imprinted genes playing essential roles in the development. In fact, imprinted genes identified so far include those regulating proliferation and differentiation of the cell and playing pivotal roles in early development, postnatal growth, and behavior of the animal. Accordingly, defects in imprinted genes lead to a variety of pathological states. Besides the characteristic congenital defects such as Prader-Willi syndrome, Angelman syndrome and

Beckwith-Wiedemann syndrome, common diseases including atopic hypersensitivity, diabetes mellitus, bipolar affective disorder, and various malignant tumors may also involve genomic imprinting in their pathogenesis (Deal 1995). Recently the importance of genomic imprinting has been realized again after several mammals cloned from somatic cells were reported (Wilmut *et al*. 1997; Wakayama *et al*. 1998), because phenotypic anomalies observed among them may be caused by epigenetic alterations, especially those of imprinted genes (Rideout *et al*. 2001; Inoue *et al*. 2002). Thus, the identification of novel imprinted genes is of particular importance not only for the elucidation of the molecular mechanisms for imprinting but for the identification of genes potentially involved in these and other pathological states.

Our laboratory previously developed a novel screening for imprinted genes by allelic message display (AMD) and this *tour de force* identified a paternally expressed gene *Impact* (Hagiwara *et al*. 1997), which encodes a protein of unknown function but of remarkable evolutionary conservation (Kubota *et al*. 2000). It is the first imprinted gene mapped to mouse chromosome 18, which was suggested to bear at least one imprinted gene by a genetic study using mice with the Robertsonian translocation chromosomes (Oakey *et al*. 1995). Recently, our team and other researchers have isolated the cDNA for its human homologue *IMPACT* and mapped it to human chromosome 18q12.1 (Okamura *et al*. 2000; Kosaki *et al*. 2001), a region syntenic to the mouse *Impact* locus (O'Brien *et al*. 1993; Gregory *et al*. 2002). Unexpectedly, human *IMPACT* was found to be expressed biallelically in various human tissues, in contrast to the mouse *Impact* that is expressed exclusively from paternal genome in all the tissues examined (Hagiwara *et al*. 1997).

I assume that the monoallelically expressed mouse *Impact* and the biallelically expressed human *IMPACT* could serve as an ideal case to reveal structural elements laying the basis for imprinted gene expression through comparative structural analysis, because they encode highly conserved proteins. Henceforth, I thus elucidated the genomic structures of these genes by means of bacterial artificial chromosome (BAC) cloning and a unique nested deletion strategy. This comparative genome analysis revealed a characteristic structure that may explain the difference in allelic expression between the two species.

Because most imprinted genes showed close physical clustering, it is supposable that mouse *Impact* provides a lead to the putative imprinted region on mouse

chromosome 18A1. The identification and characterization of genes adjacent to *Impact* excite my curiosity. As for humans, a parent-of-origin effect was seen in bipolar affective disorder (McMahon *et al*. 1995) and several groups have reported evidence suggesting possible linkage of the trait to the pericentric region of the human chromosome 18 (McMahon *et al*. 1997). Since human *IMPACT* is a nonimprinted gene, there may be some imprinted genes around this region which is related to the disease. Therefore, the identification of human genes adjacent to *IMPACT* is also very attractive. Nevertheless adjacent genes were not found to be imprinted in either humans or mice. A differentially methylated region (DMR) was found only in the mouse *Impact* locus. This study provides the first view of the architecture of a species-specific and solitary imprinted gene and gives a candidate sequence element for subsequent functional analyses.

Materials and Methods

PCR-based genome walking of *Impact* and *IMPACT*

In order to walk in genomic DNAs from the 5' ends of the mouse and human cDNAs, I utilized GenomeWalker Kits (Clontech). The gene-specific primers used were MIM051 as the mouse first primer, MIM052 as the mouse nested primer, HIM093 as the human first primer, and HIM094 as the human nested primer. Amplified PCR products were cloned into pT7Blue(R) T-Vector (Novagen) and sequenced using a primer-walk method.

Screening of mouse and human BAC libraries

The BAC clones were isolated from a genomic libraries in pBeloBAC11 vector (Kim *et al*. 1996). The BACs' host was DH10B/r and the mouse genomic DNA was derived from male 129/Sv strain. These libraries were generated by partial digestion of the genomic DNAs with *Hin*d III, followed by pulsed field gel electrophoresis. Only large fragments (100-150 kb) were excised from the gel and cloned. The PCR screening of pooled mouse and human BAC libraries was performed by Research Genetics. The primers for the mouse *Impact* 5' STS were MIM064 (forward) and MIM061 (reverse), and those for the *Impact* 3' STS were MIM029 (forward) and MIM049 (reverse). The primers for the human *IMPACT* 5' and 3' STS were HIM093 (forward) and HIM097 (reverse), and HIM049 (forward) and HIM064 (reverse), respectively. Two more mouse STSs were developed from the end sequences of 365M4 which was screened by the *Impact* 3' STS. One was later mapped to an intron of *Osbpl1b*, and the other to downstream region of *Hrh4*. Primers and product sizes for the two STSs are MIG068 and MIG069 (319 bp); MIG072 and MIG073 (230 bp), respectively.

Subcloning of restriction fragments from BAC clones

The mouse and human BAC clones were cultured in L-broth in the presence of chloramphenicol (50 μg/ml), and the DNAs were isolated by an automated plasmid isolator PI100 (Kurabo). The crude DNAs were treated with RNase A (Roche) followed by purification with salt and polyethylene glycol (Sigma) precipitation. The precipitates were rinsed with 75% ethanol, dried, and then dissolved in TE (10 mM Tris-HCl [pH 8.0], 1 mM EDTA). For the sequencing of BAC ends, plasmid DNAs were prepared using Qiagen-tips according to the manufacturer instructions.

The BAC DNAs were digested with *Avr* II, *Bam*H I, *Nco* I, or *Xba* I (New England

Biolabs) overnight, followed by treatment with Klenow fragments of DNA polymerase I (TaKaRa) to partially fill the cohesive ends. For instance, *Avr* II digests were treated with Klenow fragment and 0.2 mM dCTP/dTTP and ligated to the partially filled *Hin*d III site of pSFI-CV2, a cloning vector developed for large-scale genome sequencing (Hattori *et al*. 1997). Similarly, *Bam*H I, *Nco* I, and *Xba* I digests were appropriately filled in and ligated to the partially filled *Sal* I, *Bam*H I, and *Hin*d III sites of pSFI-CV2, respectively. Each ligation mixture was transformed into *E. coli* DH5α (TaKaRa). For each digest, I randomly selected 100 colonies, and the plasmid DNAs were prepared and digested with *Hae* III (Nippon Gene) for fingerprinting. For the sizing of the inserts of these subclones, I cut them with *Sfi* I (New England Biolabs) and electrophoresed the digests, because each restriction fragment was cloned between the two *Sfi* I sites of pSFI-CV2.

Construction of the nested deletion libraries

The subclone plasmids prepared as above were digested with *Sfi* I and ligated at a high concentration to generate concatenated DNAs, which were subsequently sonicated to pieces. The variously sized fragments were directionally cloned into pSFI-SV1 and pSFI-SV10 vectors, which had been designed for the construction of nested deletion libraries (Hattori *et al*. 1997). The inserts of these clones were amplified by colony PCR and sized by gel electrophoresis. I then chose minimum clones to fully cover the insert and subjected them to DNA sequencing.

DNA sequencing and data assembly

Plasmids and PCR products were sequenced using ABI PRISM BigDye Terminator Ready Reaction Kit (PE Applied Biosystems) by ABI PRISM 377, 3100, or 3700 DNA sequencers. For the sequencing of purine-rich regions, I used ABI PRISM dGTP BigDye Terminator Kit instead of the standard one. PCR products were directly sequenced, after treating with shrimp alkaline phosphatase and *E. coli* exonuclease I (Amersham) at 37ºC for 20 min followed by heating at 80ºC for 20 min. Sequence data were analyzed and assembled using the sequence analysis software package SEQUENCHER (Gene Codes). The sequence data for mouse *Impact* and human *IMPACT* have been submitted to GenBank under the accession nos. AF232228 and AF232229, respectively.

Search of the genomes for CpG islands

The presence of CpG islands was first predicted with Grail 1.3 at http://compbio.ornl.gov/Grail-1.3/ or an open source software, newcpgreport, at

http://www.hgmp.mrc.ac.uk/Software/EMBOSS/ . The values of CpG islands were calculated by software developed by myself. The program is provided as ANSI C code. It should be compiled on any system with an ANSI C compiler, although it has been tested on only one common Unix platform, SunOS 5.6.

Mouse resource

The JF1/Msf (JF) mice were obtained from the Genetic Stock Research Center, National Institute of Genetics (Mishima). JF, B6, ICR, and their hybrid mice were bred in the Animal Center, Institute of Medical Science, University of Tokyo.

Methylation-specific PCR assay

The four kinds of mouse genomic DNAs, each from JF, B6, and their reciprocal F_1 hybrids, were prepared from tails and spleens. Human genomic DNAs were extracted from peripheral blood. These native genomic DNAs were digested with *Hha* I, *Hpa* II, *Msp* I (TaKaRa), or McrBC (New England Biolabs) overnight in the recommended buffer solution for each enzyme. Following the heat inactivation of the enzymes at 80ºC for 20 min, DNAs were extracted by phenol/chloroform, precipitated by ethanol and used as templates for PCR. As a matter of fact, I actually used *Hap* II (TaKaRa) instead of *Hpa* II in this study. Because *Hap* II is less familiar than well-known *Hpa* II which is often used in DNA methylation study, I call the TaKaRa's enzyme *Hpa* II throughout this dissertation.

For examination of the methylation status, the assay for the mouse *Impact* CpG island were performed using primers MIG054 (forward) and MIG55 (reverse) with the following thermal cycling parameter: 94ºC for 180 s plus (94ºC for 30 s, 58ºC for 40 s, 72ºC for 80 s) for 35 cycles, and 72ºC for 180 s. Amplified products were resolved on gel electrophoresis followed by ethidium bromide staining. The PCR for the human *IMPACT* CpG island was performed using primers HIG013 (forward) and HIG014 (reverse). The PCR for the mouse *Impact* promoter region was performed using primers MIM064 (forward) and MIM052 (reverse), and sequenced directly with MIM066. The PCR for the upstream region of mouse *Impact* was performed using primers MIG080 (forward) and MIG081 (reverse), and sequenced directly with MIG081. The PCR for the mouse *Impact* 3' UTR was performed using primers MIG076 (forward) and MIG077 (reverse), and sequenced directly with MIM018. The PCR for the promoter region of mouse *Hrh4* was performed using primers MIG092 (forward) and MIG093 (reverse), and sequenced directly with MHG013. The PCR for the promoter region of human *HRH4* was performed using primers HHG013 (forward) and HHG10 (reverse). The PCR for the mouse *Osbpl1a* CpG island was performed using primers

MOG007 (forward) and MOG008 (reverse). The PCR for the human *OSBPL1A* CpG island was performed using primers HOG017 (forward) and HOG016 (reverse). The PCR for the mouse *Osbpl1b* CpG island was performed using primers MLG015 (forward) and MLG016 (reverse), and sequenced directly with MLG015. The B6-specific PCR for the island was performed using primers MLG019 (forward) and MLG022 (reverse). The PCR for the human *OSBPL1B* CpG island was performed using primers HOG021 (forward) and HOG022 (reverse). The PCR assay for CpG islands were performed in the presence of 5% DMSO. But I did not add it for the mouse *Impact* CpG island, whose %GC is the smallest among the CpG islands I studied.

Before the assay, the mouse *Impact* CpG island was amplified from JF and B6 genomic DNAs using the following primers: MIM075 (forward) and MIM048 (reverse). Amplified products were directly sequenced to find polymorphisms. The oligonucleotides for this primer-walk sequencing were MIG025, MIG030, MIG031, MIG053, MIG054, MIG055, and MIM047.

Bisulfite based cytosine methylation analysis

Native genomic DNAs of hybrid mice were denatured by 0.3 M sodium hydroxide at 37ºC for 15 min. The bisulfite (Sigma) solution, which was adjusted the pH to 5 with sodium hydroxide, and freshly prepared hydroquinone (Sigma) were then added to the denatured DNAs to final concentrations of 3.2 M and 0.5 mM, respectively. The reaction mixes overlaid with 100 μl of mineral oil were incubated at 55ºC overnight. The bisulfite treated DNAs were purified by Wizard DNA Clean-Up System (Promega) followed by desulfonation in 0.3 M sodium hydroxide for 15 min. Finally the DNAs were precipitated by ethanol and used as templates for PCR. The primers used for the *Impact* CpG island were MIG111 (forward) and MIG112 (reverse), and sequenced directly with MIG111. Those for the *Osbpl1b* CpG island were MLG025 (forward) and MLG026 (reverse), and sequenced directly with MLG026. These PCRs were performed with Platinum *Taq* DNA Polymerase (Invitrogen) which was complexed with proprietary antibody in order to inhibit polymerase activity until heat denaturation.

Southern blot hybridization for DNA methylation analysis

5 μg of human genomic DNA was completely digested with restriction endonucleases and electrophoresed in 1% agarose gel. Southern blot were transferred to nylon membrane and hybridized under high stringency conditions to the ^{32}P-labeled PCR fragment that had been amplified with HIG017 (forward) and HIG018 (reverse). Probe labeling was done with [α-^{32}P]dCTP using RadPrime DNA Labeling System

(Invitrogen).

RT-PCR and RFLP assay

RNAs were prepared by homogenizing frozen tissues in TRIZOL Reagent (Gibco BRL) with subsequent steps carried out according to the supplier's recommendations. As for the extraction from whole blood, I used QIAamp RNA Blood Mini Kits (Qiagen) for the preparation of total cellular RNA. After DNase treatment, reverse transcriptase reactions were performed using 1 μg of RNA and oligo(dT) primer (Gibco BRL).

The following primers were used to amplify mouse *Impact*, MIM025 (forward) and MIM016 (reverse). The amplified products were digested with *Tsp*509 I (New England BioLabs) at 65ºC for more than two hours, then subjected to gel electrophoresis. The PCR for mouse *Hrh4* was performed using primers MHG001 (forward) and MHG004 (reverse), and sequenced directly with MHG003. The amplified products were digested with *Hha* I (TaKaRa) at 37ºC overnight. The PCR for human *HRH4* was performed using primers HHG005 (forward) and HHG002 (reverse), and sequenced directly with HHG007 (forward) or HHG008 (reverse). The PCR for mouse *Osbpl1a* was performed using primers MOG001 (forward) and MOG004 (reverse), and sequenced directly with MOG001. The amplified products were digested with *Alu* I (New England Biolabs) overnight. The PCR for mouse *Osbpl1b* was performed using primers MLG007 (forward) and MLG006 (reverse), and sequenced directly with MLG007. Its genomic DNAs were amplified with MOG013 (forward) and MOG012 (reverse), and sequenced directly with MOG015. The amplified products were digested with *Alu* I overnight. The PCR for human *OSBPL1* was performed using primers HOG005 (forward) and HOG006 (reverse), and sequenced directly with HOG003. The SNP was discovered by a genomic PCR with HOG003 (forward) and HOG004 (reverse), which was subsequently sequenced with HOG003.

Northern blot hybridization

Tissue distribution analyses were examined by Northern blot hybridization using filters containing poly(A)$^+$ RNAs isolated from multiple tissues (Clontech). These analyses were also performed in order to identify splicing valiants. Probes were purified by agarose gel and labeled with [α-^{32}P]dCTP (Amersham) using RadPrime DNA Labeling System (Invitrogen). And then they were hybridized to the filters for more than two hours at 68ºC in Expresshyb solution (Clontech). The filters were subsequently washed two times according to the manufacturer's instructions. And they were exposed to Imaging-Plates (Fuji Film) to be analyzed on Fuji BioImaging

Analyzer BAS2000 system. The hybridized probes were stripped from the filters with boiling water containing 0.5% SDS for the next use.

The probe for the human *IMPACT* ORF was prepared by a nested PCR of human brain cDNA. The primers for the 1st PCR were HIM110 (forward) and HIM006 (reverse), and those for the nested amplification were HIM028 (forward) and HIM007 (reverse). The probe for the 3' UTR was amplified from 457A4 with HIM021 (forward) and HIM064 (reverse). The probe for human *OSBPL1* was amplified from human leukocyte cDNA with HOG005 (forward) and HOG006 (reverse). The probe for mouse *Osbpl1a* was amplified from mouse brain cDNA with MOG001 (forward) and MOG004 (reverse). The *Osbpl1b*-specific probe was amplified from mouse brain cDNA with MLG011 (forward) and MLG012 (reverse).

Identification of the mouse *Osbpl1b* gene

The first RT-PCR was performed using MLG011 (forward) and MOG004 (reverse). The forward and reverse primers were designed from the sequences of *LOC211564* and the mouse *Osbpl1a* cDNA, respectively. B6 brain cDNA was used as the template. Its end sequence of the PCR product had high homology to I.M.A.G.E. clone 3990991. Further RT-PCR was performed on these DNA sequences and its 3.12-kb cDNA fragment covering the whole ORF was determined using MLG001, MLG003, MLG004, MLG005, MLG007, MLG013, MLG021, MLG023, MLG024, MOG010, and MOG011 by direct sequencing.

Allele-specific quantitative PCR analysis of the mouse *Osbpl1b* gene

Allele-specific transcription of the mouse *Osbpl1b* was quantitatively confirmed by ABI PRISM 7900 using SYBR Green PCR Master Mix (PE Applied Biosystems). The primers used for B6 allele were MLG011 (forward) and MLG028 (reverse). Those for JF were MLG011 (forward) and MLG030 (reverse).

Oligonucleotides used in this study

Primers used in this study are as follows.

HHG002 : 5'- GGG CAG ACC TGA TTC ATT TAG -3'
HHG005 : 5'- TAC CTG TCA GTC TCA AAT GCT -3'
HHG007 : 5'- ACT CAA CAT ACT GGG GTC TT -3'
HHG008 : 5'- CTT GGT TCT TGA GGA AAA CA -3'
HHG010 : 5'- CCA GCC AGA CAA TTC TGA CA -3'
HHG013 : 5'- AGA CCT CAT CCG TTC TCT CA -3'
HIG013 : 5'- CCC TAG GAA TGT AAA GAC GAG -3'

HIG014 : 5'- CCA GAA GGA GTG AGA TTC GG -3'
HIG017 : 5'- CTC TAC AGG TGT TGA ACT CC -3'
HIG018 : 5'- GGA AGG GAC TCA AAC CTC AG -3'
HIM006 : 5'- GTT GGT ATC CAT GCT GAA CTG G -3'
HIM007 : 5'- CCT TAG ATG ACT CCT CAG GTG AA -3'
HIM021 : 5'- CAA GGT AAC AGT TGC CCA GG -3'
HIM028 : 5'- GGC GAG GAG TGG TGT GTC A -3'
HIM049 : 5'- CGT AGA GTG GGA TAG AGG TGG CAG AAT G -3'
HIM064 : 5'- CTG GAA GAT GAA AGA TAC AT -3'
HIM093 : 5'- GGA CGG TGT CCT CGT CAA CCA TTA ACA -3'
HIM094 : 5'- TGG GCC GAC GAA AAA CCG GGG TTT CGA -3'
HIM097 : 5'- ACC TGC AGG GTC TGG GCT ATT GCC ATT -3'
HIM110 : 5'- TGA GGA AAT TGA AGC AAT GG -3'
HOG003 : 5'- CAG TTT TCT GCA GTC AGT ATC -3'
HOG004 : 5'- TGG GGA ACA TTC TAA AGC CT -3'
HOG005 : 5'- GAA GAG GAC TGG AAG ACG AG -3'
HOG006 : 5'- TGA TAC TTA CAT GAG TGC AAC -3'
HOG016 : 5'- GCG GCC TCT GAA GAG CGG AT -3'
HOG017 : 5'- CAG GCT GCG CAA AGG TGA CT -3'
HOG021 : 5'- GGG AGT GCC AGC CAG AGT T -3'
HOG022 : 5'- GGC ACG CAG CTG AAG ATC TG -3'
MHG001 : 5'- AAT ATT GTC CTC ATT AGC TAC GAT CG -3'
MHG003 : 5'- TGT CTG TTC ACA ATT GTC CTT TCA AC -3'
MHG004 : 5'- TGG TTG CTT TGT CAC ACA AAG TAT CT -3'
MHG013 : 5'- GCA TGG TGG ACT GCA GGT -3'
MIG025 : 5'- AAG TCA CAA TGC CGA GCT GA -3'
MIG030 : 5'- CAG CGT TGT CAC ACA AGC AA -3'
MIG031 : 5'- TGA GCA GGG ATT GCA CAC GT -3'
MIG053 : 5'- GCT CAT ATG ATG CAG TAG ATG AA -3'
MIG054 : 5'- CCG TAG CAT CAC ACT ACG TA -3'
MIG055 : 5'- TCG AAC ACA CAC TCG AGG TA -3'
MIG065 : 5'- GGC GCG GCC AAC TCT GCT -3'
MIG068 : 5'- GCT TGT TAA CAT GTC AAC TTT C -3'
MIG069 : 5'- GCA GTT GAG ACA TTG CAT TAG T -3'
MIG072 : 5'- CAG ATG ACT AAC CCT GTT CA -3'
MIG073 : 5'- CCT AGG GTA TAA GCA ACT AC -3'
MIG076 : 5'- GCC ATG CTG TTA AGT GAG CAT TG -3'

MIG077 : 5'- GGC CAG AGT TCA AAT GGC AAA TG -3'
MIG080 : 5'- CTC CCA TTC TCA TTC CAG AT -3'
MIG081 : 5'- GTC CTG TCG GTG TTA CTC TT -3'
MIG092 : 5'- TGG GGC TAG CTA AAC ATA GT -3'
MIG093 : 5'- GAC GCT CGC AGG AGG CTT CA -3'
MIG111 : 5'- TAT GAG TTA GTT TTG GGT TTA -3'
MIG112 : 5'- TAA CCC TAC CCA TCT AAA CCA C -3'
MIM016 : 5'- TAC GTG GAT GCC ATT TGC T-3'
MIM018 : 5'- AAA CTG GTG CAT GTG GGC -3'
MIM025 : 5'- CCT TCT TCA CCT CAT GGA G -3
MIM029 : 5'- ACG TTT CCC CAT TTT ACA AG -3'
MIM047 : 5'- CAC CAC TCC TCG CCA TAA ATG GCT GCC A -3'
MIM048 : 5' -CCA TTT GGG GTC ATC CAT GAA GTC AGT G -3'
MIM049 : 5'- AGT ATC ACT CAC CTG CCC TG -3'
MIM051 : 5'- TTC CTC TTC AGC CAT GGT GCT CAG GAT C -3'
MIM052 : 5'- TGG CAA GCA GCA AAT GAA TGC AAC TGC G- 3'
MIM061 : 5'- TAG TGT AGA CTG GGC TCA -3'
MIM064 : 5'- GTG GGG TAC AGT AAG AGT -3'
MIM066 : 5'- TCT CCA GCT CTC GTT CAT -3'
MIM075 : 5'- CGG AAG CAA TTC AGG AAG TGG GTG GTG T -3'
MLG001 : 5'- CTG GCA GCC ATG GCA CGG AT -3'
MLG003 : 5'- CAA TGC TGA AGA GGT GCG AA -3'
MLG004 : 5'- CCC CCA TGT CGT TTA ACA TG -3'
MLG005 : 5'- GTT AAA CGA CAT GGG GGA CA -3'
MLG006 : 5'- CAG TTC ACA TCA GGA GGA TT -3'
MLG007 : 5'- GTG GTC GAG GAT CTG TTG AA -3'
MLG011 : 5'- GGT CGC TGA CAT CGA CTG TA -3'
MLG012 : 5'- GCT TGA GTC AGG TGT TTG CAT 3'
MLG013 : 5'- AGT ATC AGC TCT GCT CAG CA -3'
MLG015 : 5'- TGC TGC CGC CCC TCT TTC AC -3'
MLG016 : 5'- ACC CAC GCG GGC CGC TCT CA -3'
MLG021 : 5'- GGA AAT TTA CAA CTT TCT CA -3'
MLG023 : 5'- TAC CCG GGG CGT CCT CGT GA -3'
MLG024 : 5'- GGT TTG TCC ATG TGT ATG CT -3'
MLG025 : 5'- TCA CCC CCA CCT CCC CCA CCC AAA TCT -3'
MLG026 : 5'- GGG TTA GGA GTA AGG AGG GTT TYG GGA A -3'
MLG028 : 5'- CCT TTC GTC CTG TGA AGG CG -3'

MLG030 : 5'- CCT TTC GTC CTG TGA AGG CA -3'
MOG001 : 5'- TGC AGA AGG GCT CAA CAA TG -3'
MOG004 : 5'- AGT CCT CTT CCG ACT TGG AC -3'
MOG007 : 5'- GCT ACA GCC AGG ATC CCT TA -3'
MOG008 : 5'- CCT GGG CTG GGT CTT GAA GA -3'
MOG010 : 5'- GGC CAT GAG CAG GGC TCC AA -3'
MOG011 : 5'- AGA ACG GAG AGA TAG ATC TA -3'
MOG012 : 5'- AAG CTG AGC ATG TTC AAT CA -3'
MOG013 : 5'- CGT GGC AAT AAG CAG TGT TA -3'
MOG015 : 5'- GTG TGT GTA TTT ACT AAT G -3'

Results

1. Screening of mouse and human BAC libraries

To obtain genomic clones containing the mouse *Impact* gene and the human *IMPACT* gene, I adopted the BAC system based on a single-copy plasmid F factor of *E. coli*, which is suitable for a long-range genome analysis because of its high cloning capacity, easy manipulation of the cloned DNA, and stable maintenance of inserted DNA (Shizuya *et al*. 1992). I used the BAC system not only to reveal entire genomic structures of *Impact* and *IMPACT* but also to gain access to their neighbors. Since imprinted genes often cluster in the genome, allelic expression status of genes adjacent to mouse *Impact* and human *IMPACT* is of particular interest.

For the PCR screening of BAC libraries, I developed two STSs for each gene (Fig. 1.1). One STS is derived from the 5'-end portion of the gene, which was obtained by genome walking (Siebert *et al*. 1995) from the 5' end of cDNA clone, and the other is derived from the 3'-end portion of the cDNA. Following the confirmation that these STSs can be readily amplified from genomic DNAs, they are used for PCR screening of mouse and human BAC libraries. Consequently, I obtained two mouse BAC clones (200P19 and 365M4) and two human BAC clones (457A4 and 558E15). Although STS content mapping showed that 200P19 lacks the 5'-end portion of *Impact*, the other clone, 365M4, includes both ends. Both the human clones cover the *IMPACT* from its 5'-end to 3'-end, and restriction digestion analysis indicated that 457A4 has a longer insert than 558E15. These BAC clones were used as probes for fluorescence *in situ* hybridization (FISH). The mouse clones were both mapped to chromosome 18A2-B1, to which *Impact* was mapped (Hagiwara *et al*. 1997). Similarly, both the two human clones were mapped to chromosome 18q11.2-12.1, as was the *IMPACT* cDNA (Okamura *et al*. 2000). No signs of chimerism were observed for these clones in these FISH analyses. I thus decided to use 365M and 457A4 for the sequencing analysis of the mouse and the human genes, respectively.

For further analysis, two other STSs were also developed from both ends of 365M4 to cover adjacent genes. Some clones were screened, which is described in Figure 5.1.

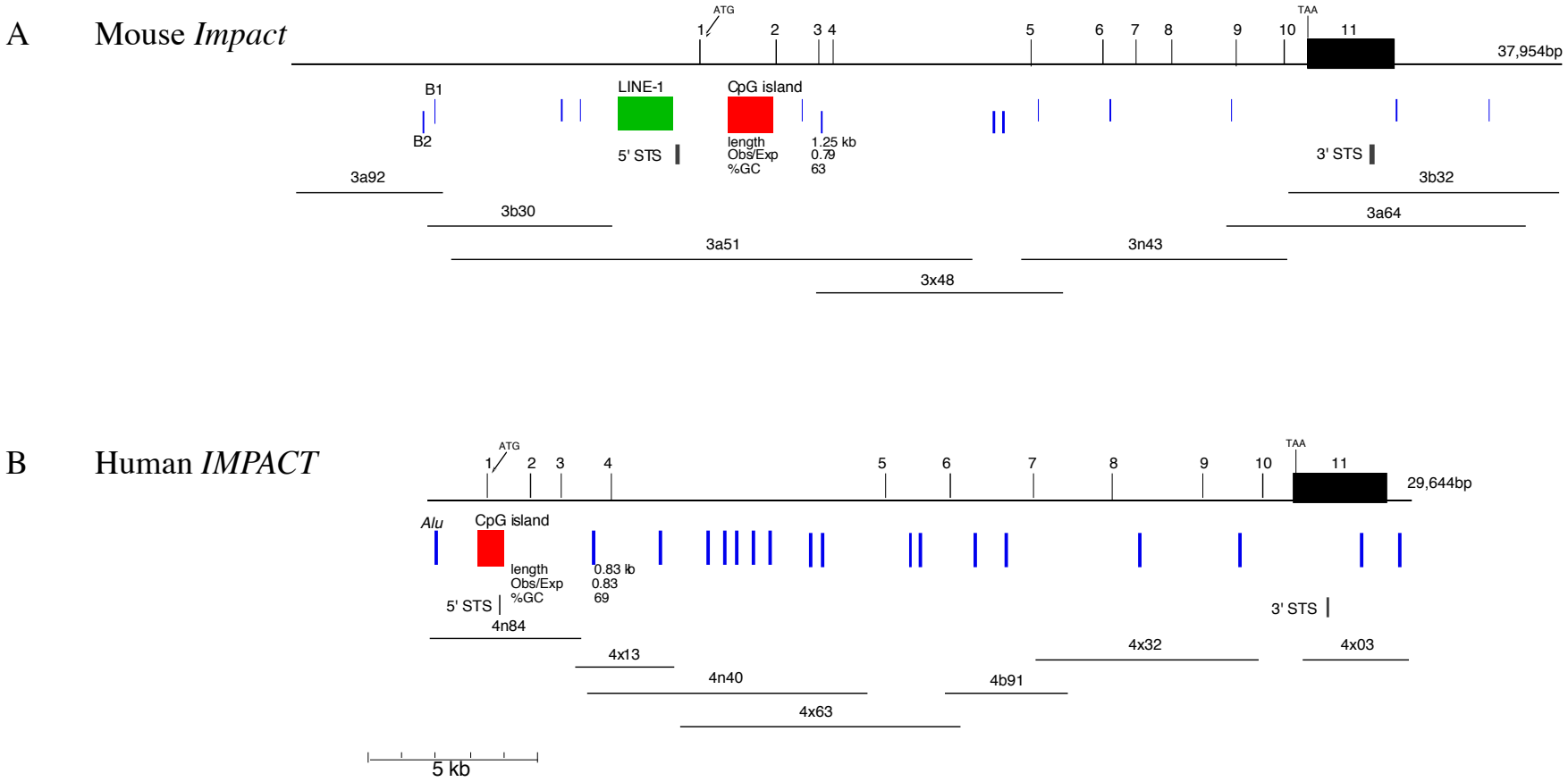


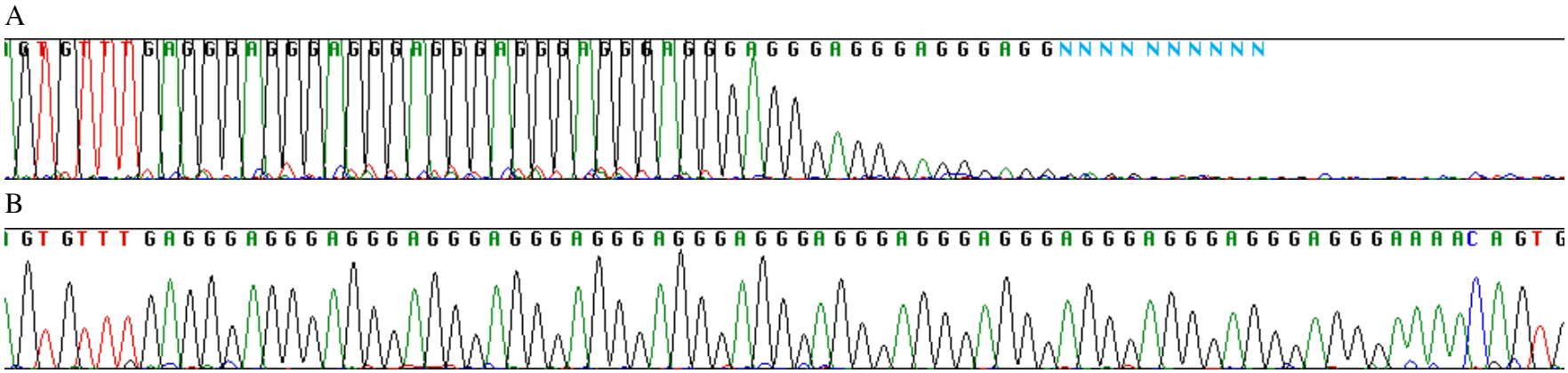
Figure 1.1 Genomic organizations of mouse *Impact* and human *IMPACT*. Physical maps of mouse (A) and human (B) genes are shown. Minimum contig of subclones to cover each gene is also shown. PCR and direct sequencing closed a gap between human subclones 4x32 and 4x03. Exons are shown as solid boxes and numbered 1 to 11. Arrows indicate the initiation and termination codons. The positions of SINEs, LINE-1, CpG islands, and STSs used for the library screening are also illustrated.

2. Sequencing of BAC clones by the nested deletion method

I used a unique nested deletion strategy for the sequencing of the BAC clones (Hattori *et al*. 1997). The mouse 365M4 and human 457A4 were first digested with *Avr* II, *Bam*H I, *Nco* I, or *Xba* I. After partially filling the restriction ends in, the restriction fragments were cloned into pSFI-CV2, a cloning vector developed for this strategy. The half-fill ligation approach were designed to minimize the possibility of chimeric clones. I then randomly picked 100 colonies for each digest to prepare 400 subclones for each BAC. Plasmid DNAs were prepared from these clones for the sizing of the inserts and restriction fingerprinting. Through these analyses, I selected about 150 apparently independent clones and subjected them to sequence tagging from both ends. These sequence data helped us identify a minimum set of independent subclones covering each BAC. It also allowed us to eliminate clones derived from pBeloBAC11 vector and *E. coli* genomic DNA, which had been contaminated during the preparation of BAC DNAs. Finally, I obtained about 100 independent subclones for each BAC. From these subclones, I first chose those containing exons of *Impact* or *IMPACT* in either end and subjected them to the nested deletion strategy.

The insertion of each deletion clone was amplified by colony PCR and directly sequenced as described (Hattori *et al*. 1997). However, I encountered several difficulties in the direct sequencing of colony PCR products. For instance, a few subclones including the first exon of *IMPACT*, which later turned out to contain a CpG island whose G+C content was 69%, were refractory to colony PCR. For this region, I used the plasmid as the template. Also, the amplification of DNA fragments containing poly(A) or poly(T) stretches resulted in DNAs heterogeneous in the length of the stretch, presumably due to the slippage during the PCR. It is thus difficult to sequence beyond these repeats and is impossible to determine the exact length of the stretch. For these regions, I took the primer-walking strategy on the plasmid templates. Another obstacle was a purine-rich repeat $(AGGG)_{15}$ in the fourth intron of mouse *Impact*, which could not be sequenced unless I used dGTP in the deoxynucleoside triphosphate mix in the presence of 5% DMSO (Fig 2.1).

Sequencing of one subclone often indicated the next to be sequenced, and hence the contig of sequences gradually expanded. Following the gap closure by long PCR from the BAC clones and resequencing of ambiguities by primer-walk, I finally elucidated 37,954 bp of contiguous finished sequence for mouse *Impact* and 29,644 bp for human *IMPACT*.



18 **Figure 2.1** Sequencing with dITP (A) and dGTP (B). Standard dye terminator kits for DNA sequencing contain dITP instead of dGTP. Replacement of dITP with dGTP allows successful extension through difficult-to-sequence regions, *e.g.* a purine-rich region in *Impact*.

3. Genomic organization of mouse *Impact* and human *IMPACT*

The genome structures of mouse *Impact* (GenBank acc. no. AF232228) and human *IMPACT* (GenBank acc. no. AF232229) are depicted in Figure 1.1 with the minimum contigs of the subclones. Alignments of the genome sequences with those of cDNAs revealed that both genes have 11 exons. The average size of the exons is about 100 bp except for the last ones, which contain the termination codon and are longer than 2 kb. All the splice junctions shown in Table 1 follow the Chambon rule (Breathnach and Chambon 1981). And most of them split the open reading frames (ORFs) at the identical positions between the two species, although there is one exception (Table 1). Thus, the overall genome organizations of these genes are well conserved, thereby providing further evidence for their orthology.

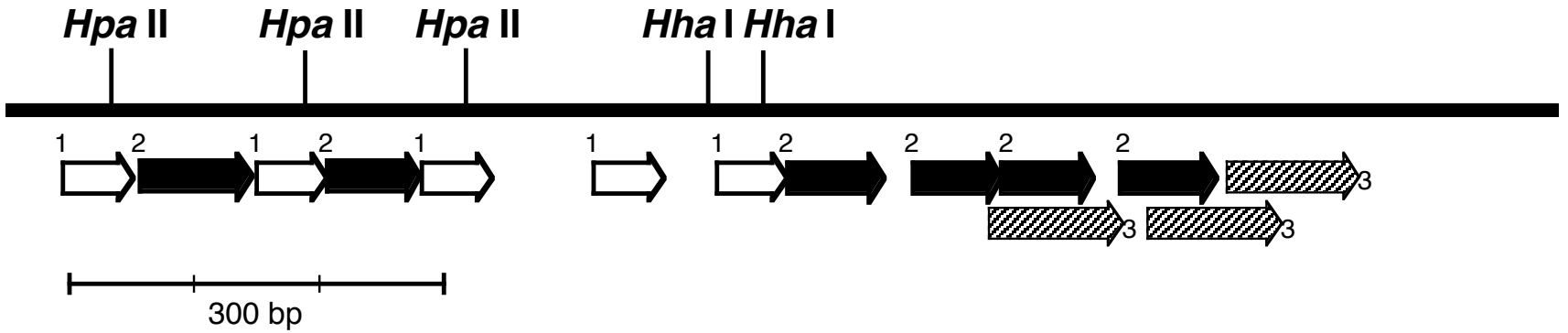
A remarkable difference between the two genes was found in their upstream promoter regions. The promoter region and the first exon of human *IMPACT* constitute the sole CpG island of this gene. In contrast, the corresponding region of the mouse gene is rather AT rich; the %GC is 43. Although the ratio of observed versus expected CpG dinucleotides (Obs/Exp CpG) of this region is 0.35, which is significantly higher than the average for the 38-kb region (0.25), it does not meet the criteria for a CpG island.

Instead, mouse *Impact* has a CpG island in its first intron. The intronic island has many TCGGC sequences and a characteristic tandemly reiterated structure (Fig. 3.1). It is known that such tandem repeats often associate with imprinted genes (Constancia *et al*. 1998). Notably, I failed to find any such tandemly repeated structures either in the CpG island or elsewhere in nonimprinted human *IMPACT*.

Table 1. Aligned Exon-Intron Organizations of Mouse *Impact* and Human *IMPACT*

Gene	Exon	Size (bp)	5' Splice Donor	Intron	Size (bp)	3' Splice Acceptor
Impact	1	121	AGC CAG AGG CAG gtaaggtccc...	1	2147	...caactttcag AGT GAA GAA ATC
IMPACT	1	75	GAC CAG AGG CAG gtgaggcccc...	1	1126	...tttctttcag AAT GAG GAA ATT
Impact	2	129	CTT TGT TTA CAG gtaacttata...	2	1121	...ttctttgtag GTG ATG TTG CCA
IMPACT	2	129	CTT TGC TTG CAG gtacttttc...	2	821	...tgcattgtag GTG ATG CTG CCG
Impact	3	53	TAT CAG CTG AA gtaagctttg...	3	434	...tcaatcttag C GCT CCC TGG
IMPACT	3	53	TAC CAG TTG AA gtaagctgta...	3	1400	...ttaatcttag T GCT CCT TGG
Impact	4	63	GAG ATA TAT GT gtaagtgaca...	4	5802	...ttttgtttag C CAC AAC ATG
IMPACT	4	63	GAA ATA TAT AT gtaagtgaca...	4	7565	...ttttgtttag T CAG AAT ATC
Impact	5	86	ACC GAG CCA G gtgagagtga...	5	2060	...ttttacatag AC CCA GAT GTC
IMPACT	5	86	ACA GAA CCA G gtaggattga...	5	2454	...attcacatag GC CCA GAT GTA
Impact	6	126	ACT CAA CCA G gtattgtgtt...	6	834	...gactttatag AA ACT
IMPACT	6	123	ACA G gtataatgtt...	6	2429	...acatttctag AA GTA GAA GTA
Impact	7	98	TGT CCT GAG CAG gtgcgtgacc...	7	1006	...gtttttgaag GTG AAA CTG GTT
IMPACT	7	104	TGT CCC AAA CAG gtaaagttcc...	7	2240	...tgtttttaag GTG AAA ATG GTT
Impact	8	74	TAT GCC TAC AG gtgagtgact...	8	1874	...gtgtttccag G ATA TTC TGT
IMPACT	8	74	TAT GCC TAC AG gtgagtaatc...	8	2626	...gtatttgcag A ATA TAT TGT
Impact	9	91	CAC CTC ATG GAG gtgagtgtag...	9	1349	...aaaataatag ATT TTG AAT GTG
IMPACT	9	91	CAT CTC ATG GAG gtaggtgtaa...	9	1635	...aaaataacag ATT TTG AAT GTG
Impact	10	135	ACG AAC ACT CCT gtgagtcttg...	10	613	...tctcttccag GAC GAG TCA ACT
IMPACT	10	135	ACA AAT TCA CCT gtaagtggct...	10	825	...tctttttcag GAG GAG TCA TCT
Impact	11	2462				
IMPACT	11	2749				

A



B

21

```
 46  GCACTAGCTTTGCCGCATTGTCACATGAGCAGGCCCGGCCCACTTGGCTGGGCTCGG  102
199  GCACTAGCTTTGCCGCATTGTCACATGAGCAGGCCCGGCCCACTCGGCTTGGCTCGG  255
332  GAGCAGGCTTTGCTGCATTGTCACATGAGCAGGCCCGGCCCACTCGGCTTGGCTCGG  388
468  GCACTAGCTTTGCCGCATTGTTACACGAGCAGGCCTGGCCCATTCGGCCGCGTTATT  524
568  CCACTAGCTTTGCCGCATTGTCACATAAACAGGCGCGGCCAACTCTGCTCGGCTTGG  624
```

C

```
105  TGGCTCGGCTGCTGCGTCACTGGACTCGGCTGCTCGGCTGCGTTGTCACATGTTAGCAAGGCGGACTAGGCTGCTGCGTCACACGAGC-AGGCTC  198
248  TGGCTCGGCTGCTGCGTCACTGGACTCGGCTGCTCGGCTGCGTTGTCACATGTTAGCAAGGCCGACTAGGCTGCTGCGTCACACGAGC-AGGCTT  341
617  CGGCTTGGCTGCTGAGTCACTGGGCTCTGC-------C-GAGTTGTCACGTGT--GCAATCCCTGCTCAGCTGCTGCGTCACAAGAGCTCGGCTC  701
725  TGGCTCGGCTCGG------CTTGGCTCGGCT--T-GGCTGAGTTGTCACATGG--GCAAGCCCTGCTC-----------ATTCGAGC-AGGCTC  795
790  AGGCTCGGCTCGG------CTCGGCTCGGC--------TGAGTTGTCACATGG--GCAAGCCCTGCTCGGCTGCTGCGTCATACGAGC-AGTCTC  867
877  CGGCTCGGCTCGG------CTCGGCTCGGCT---CGGCTGAGTTGTCACATGA--GCAAGCCCTGTTCATCTACTGCATCATATGAGC-AGGCTC  959
```

D

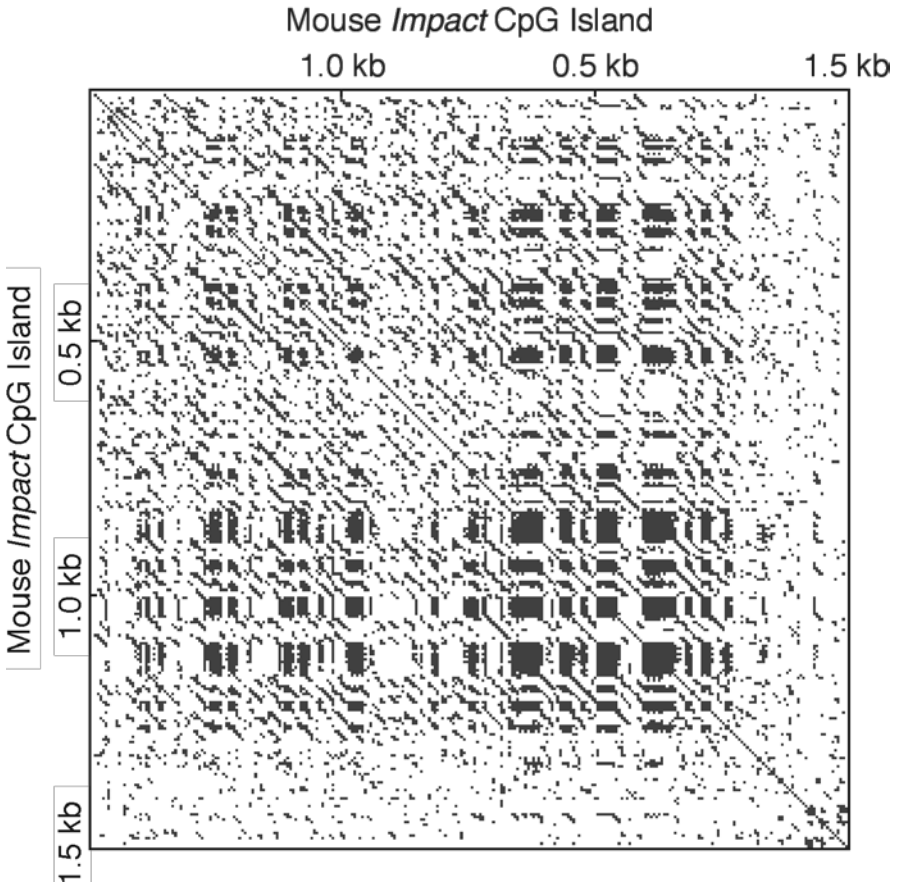


Figure 3.1 Structure of the CpG island in the first intron of *Impact*. Tandem repeated structures are schematized by broad arrows (A). The island contains two units designated as 1 (B) and 2 (C), the nucleotide sequences of which are shown respectively. Although many polymorphic sites are found between the two mouse strains, two *Hha* I sites and three *Hpa* II sites are conserved and indicated in the figure. A sequence numbered as 3 is inserted only in B6 genome. These sites were thus used for the methylation-specific PCR assay. The reiterated structure was first recognized by means of Harr plot analysis (Harr *et al*. 1982). The plot of a 1.5-kb region containing the island shows a distinctive pattern (D).

4. Parent-of-origin specific methylation of the mouse *Impact* CpG island

Because tandem repeats often occur in imprinted genes and are implicated in the establishment of genomic imprinting (Constancia *et al*. 1998; Feil and Khosla 1999), I examined the methylation status of *Impact*. For this, I prepared two parental mouse strains, *M. musculus domesticus* C57BL/6J (B6) and *M. musculus molossinus* JF1/Msf (JF), and reciprocal F_1 hybrids between the two as described (Hagiwara *et al*. 1997). Then I amplified and sequenced the CpG islands from B6 and JF to search for polymorphisms between the two mouse strains. Fortunately, the island of B6 is 181 bp longer than that of JF, owing to the difference in repeat organization. This allows me to discriminate between the B6 and JF alleles simply by their lengths.

To examine the methylation status, I developed a methylation-specific PCR assay. In this assay, genomic DNAs are first digested with methylation-sensitive restriction endonucleases such as *Hha* I or *Hpa* II, and then used for PCR to amplify the locus of interest. Although unmethylated targets are cut by the enzymes and will not be amplified, a methylated target survives the digestion to serve as the template for subsequent PCR. In other words, the methylated allele is amplified. Because both B6 and JF alleles for the CpG island of *Impact* share the same five methylation-sensitive restriction sites, namely, two *Hha* I sites and three *Hpa* II sites, I can apply the methylation-specific PCR assay to this island.

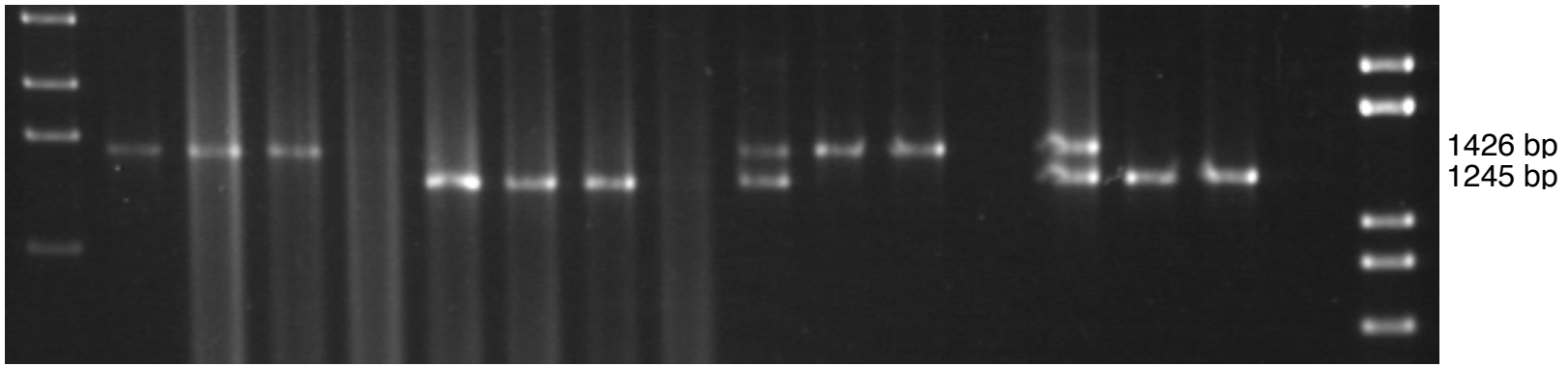
I digested the genomic DNAs from B6, JF, (B6 x JF) F_1, and (JF x B6) F_1 with *Hha* I, *Hpa* II, or *Msp* I and used them as the templates for PCR spanning the intronic island (Fig. 4.1). When native undigested genomic DNAs of the F_1 hybrid mice were used as the template, I readily obtained two bands derived from B6 and JF alleles that can be clearly separated by gel electrophoresis. When the DNAs treated with *Hha* I or *Hpa* II were used for the PCR, only one of the two bands was obtained. The B6 allele was amplified from (B6 x JF) F_1, whereas only the JF allele was detected from (JF x B6) F_1. When I used *Msp* I, a methylation-insensitive isoschizomer of *Hpa* II, as a control, I could not amplify any bands at all. These results manifestly demonstrated that the island is methylated in a parent-of-origin-dependent manner - the silenced maternal allele is hypermethylated, and the active paternal one is undermethylated. Thus, the island serves as a differentially methylated region (DMR) for this gene. This monoallelic methylation was confirmed by bisulphite based cytosine methylation analysis. Native genomic DNAs of hybrid mice were denatured and treated with bisulphite to convert unmethylated cytosines to thymines (Clark *et al*. 1994). Only one strand was amplified and analyzed by direct sequencing, because at some loci the two modified alleles clone with different efficiencies.

An experiment using embryonic genomic DNAs were undertaken to investigate when the methylation pattern of the island is established (Fig. 4.2). The result suggests that the pattern has already been fixed by 8.5 d.p.c.

I next examined the methylation status of the promoter region, which bears two *Hpa* II sites at positions -55 and -14. I readily amplified the expected DNA fragments from the *Hpa* II-digested genomic DNAs derived from the parental strains and reciprocal F_1 hybrids (Fig. 4.3). This indicated that the region is at least partially methylated. I thus sequenced the amplified fragments or the methylated allele to know their parental origins by SNP at the position -355. When the amplified fragments from undigested genomic DNAs of F_1 hybrid mice were used as the templates, I detected both T and C at this position, representing B6 and JF alleles, respectively. In contrast, I can detect only T from the *Hpa* II-digested DNA from (B6 x JF) F_1 and only C from (JF x B6) F_1. These results indicate that the imprinted maternal allele is methylated, and the expressed paternal one is not methylated, like the CpG island. I also applied a similar assay to the regions flanking this gene and found that these sites are methylated on both alleles (Fig. 4.4).

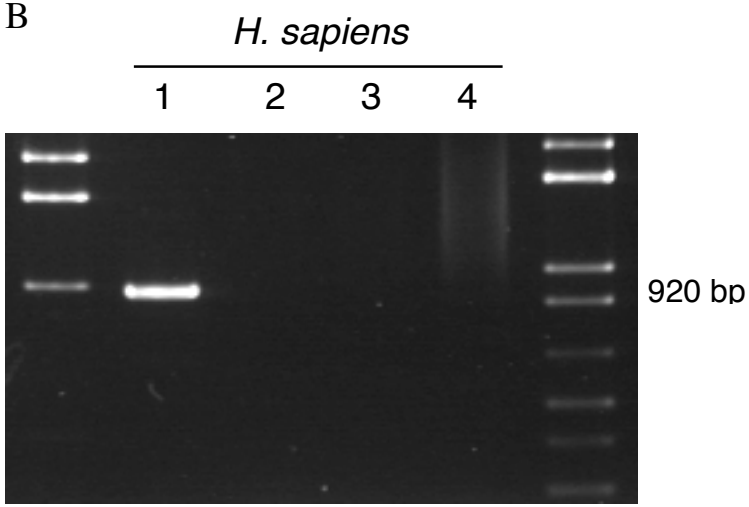
I applied a similar methylation-specific PCR assay to the human CpG island, which spans the promoter region, the first exon, and the first intron. In contrast to the mouse intronic island and promoter, not only *Msp* I digestion but also *Hha* I or *Hpa* II treatment completely abolished the amplification of the human island (Fig. 4.1), thereby suggesting its undermethylation. Furthermore, Southern blot hybridization analysis revealed that *Hpa* II digests the island to tiny fragments as efficiently as *Msp* I does (Fig. 4.5). These results indicate that the whole island is unmethylated on both chromosomes as are conventional CpG islands (Gardiner-Garden and Frommer 1987; Takai and Jones 2002).

A

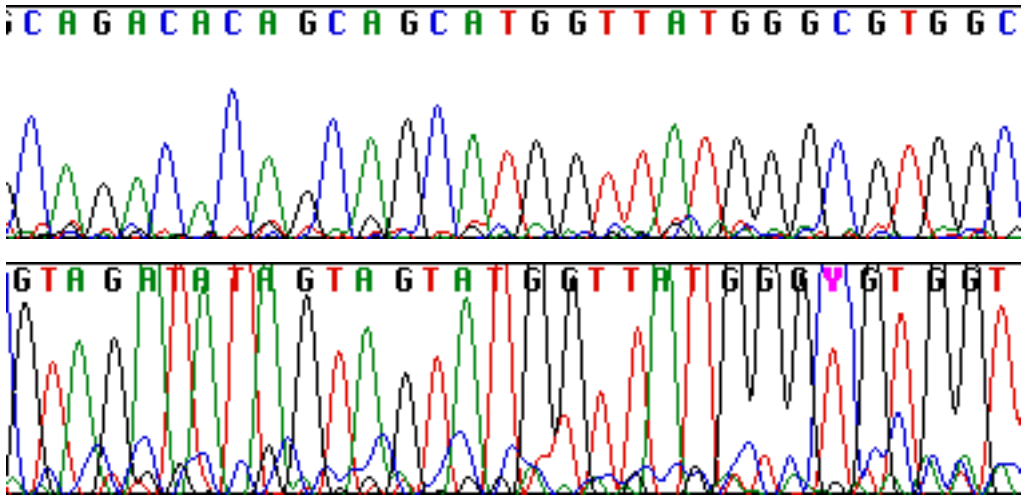

B6
JF
(B6 x JF) F1
(JF x B6) F1
1 2 3 4 1 2 3 4 1 2 3 4 1 2 3 4
1426 bp
1245 bp

25

B


H. sapiens
1 2 3 4
920 bp

C



bisulphite modified genomic DNA

26

Figure 4.1 Parent-of-origin-specific methylation of *Impact*. Methylation-specific PCR assays for the CpG island of mouse *Impact* (A). PCR products from native genomic DNA and those digested with *Hha* I, *Hpa* II, or *Msp* I (lanes 1 to 4, respectively) were subjected to 1.5% agarose gel electrophoresis and stained with ethidium bromide. *Sty* I-digest λ DNA and 1 kb PLUS DNA LADDER (GIBCO BRL) were used as size standards in the left- and rightmost lanes, respectively. The mouse intronic CpG island was analyzed in JF, B6, (B6 x JF) F_1, and (JF x B6) F_1. Methylation-specific PCR assays for the CpG island of human *IMPACT* (B). Human CpG island, which overlaps the promoter region and lacks length polymorphisms, was also analyzed as described in A. Monoallelic DNA methylation of the mouse *Impact* CpG island was confirmed by bisulphite based cytosine methylation analysis (C). The upper and lower chromatograms were obtained by direct sequencing of non-treated and bisulphite-treated genomic DNAs. Unmethylated cytosines converted to thymies. Monoallelically methylated cytosines were detected as mixtures of the two pyrimidines.

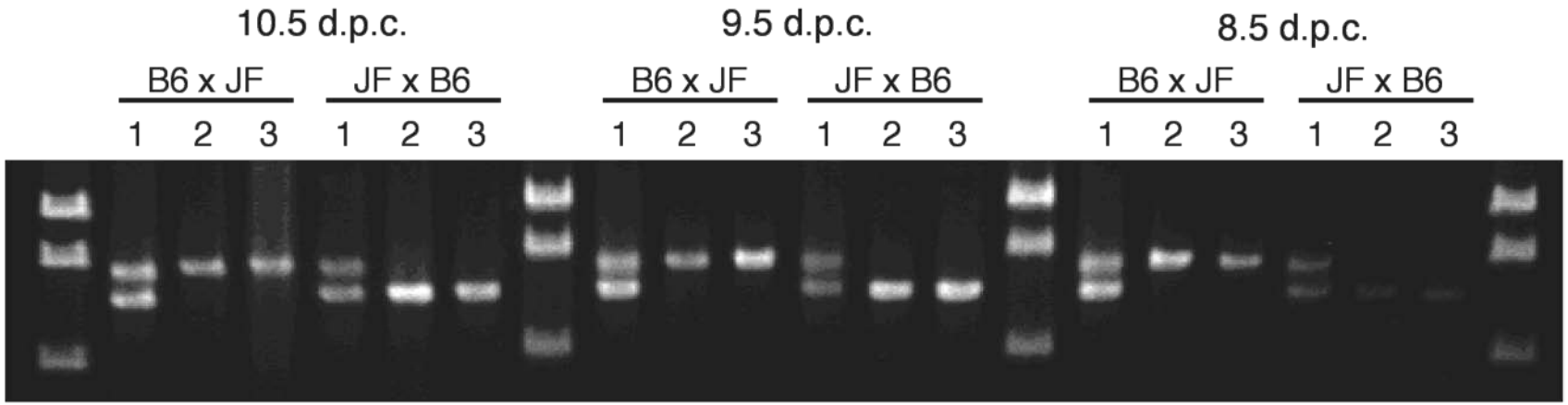


Figure 4.2 DNA methylation analysis of the *Impact* CpG island in embryonic stages. The methylation-specific PCR assay was performed to know when the DMR is established. Genomic DNAs were extracted from whole embryos of the reciprocal hybrid mice. PCR products from native genomic DNA and those digested with *Hha* I or *Hpa* II (lanes 1 to 3, respectively) were subjected to agarose gel electrophoresis. λ DNA-*Sty* I fragments were used as size standards. The result suggests that the DNA methylation pattern has already been established by 8.5 d.p.c.

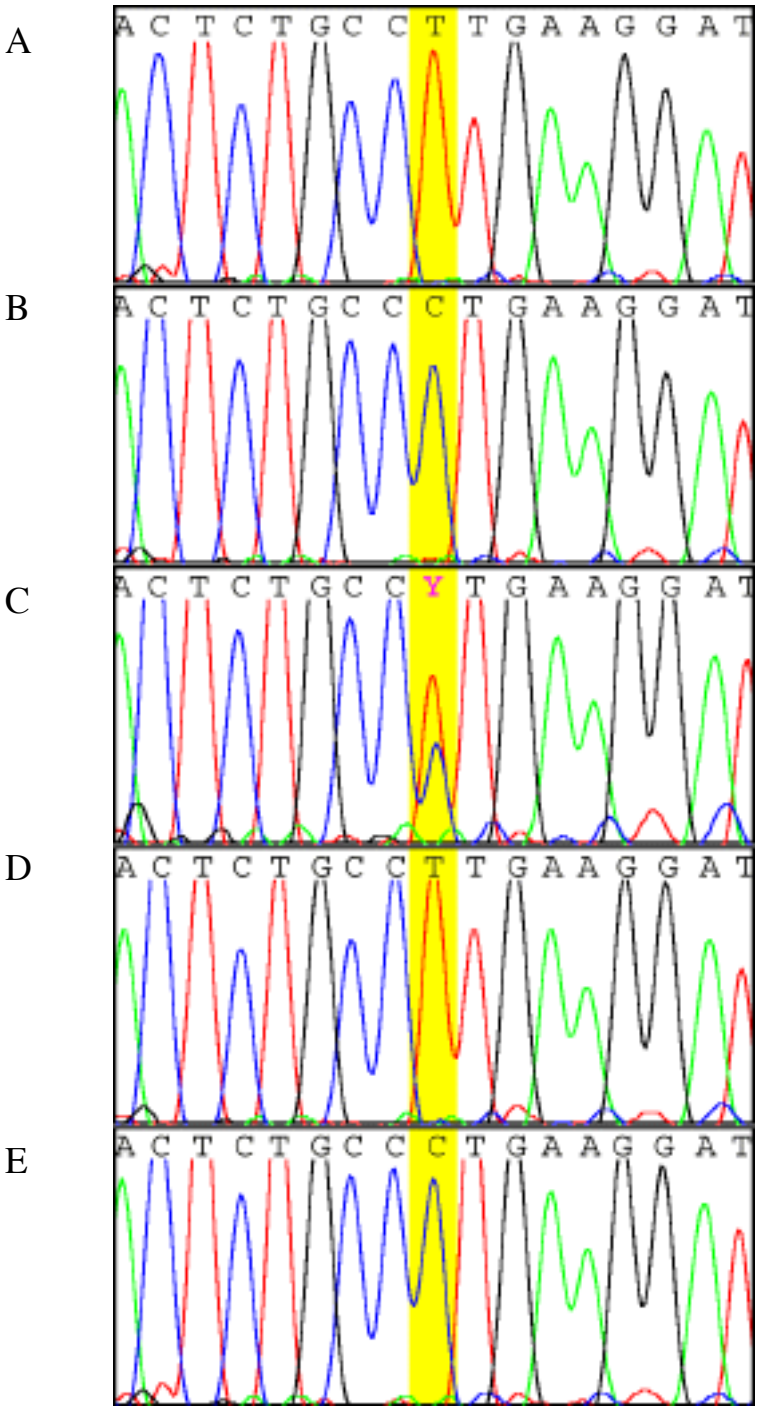


Figure 4.3 Methylation-specific PCR assays for the promoter region of *Impact*. PCR products were subjected to direct sequencing. Each panel shows the chromatogram for the sequence around the 350 bp-upstream region from the initiation codon. Undigested genomic DNAs from B6, JF, and (JF x B6)F_1 were used as the template in A to C, respectively. *Hpa* II-digested genomic DNAs from (B6 x JF) F_1 and (JF x B6) F_1 were used in D and E, respectively.

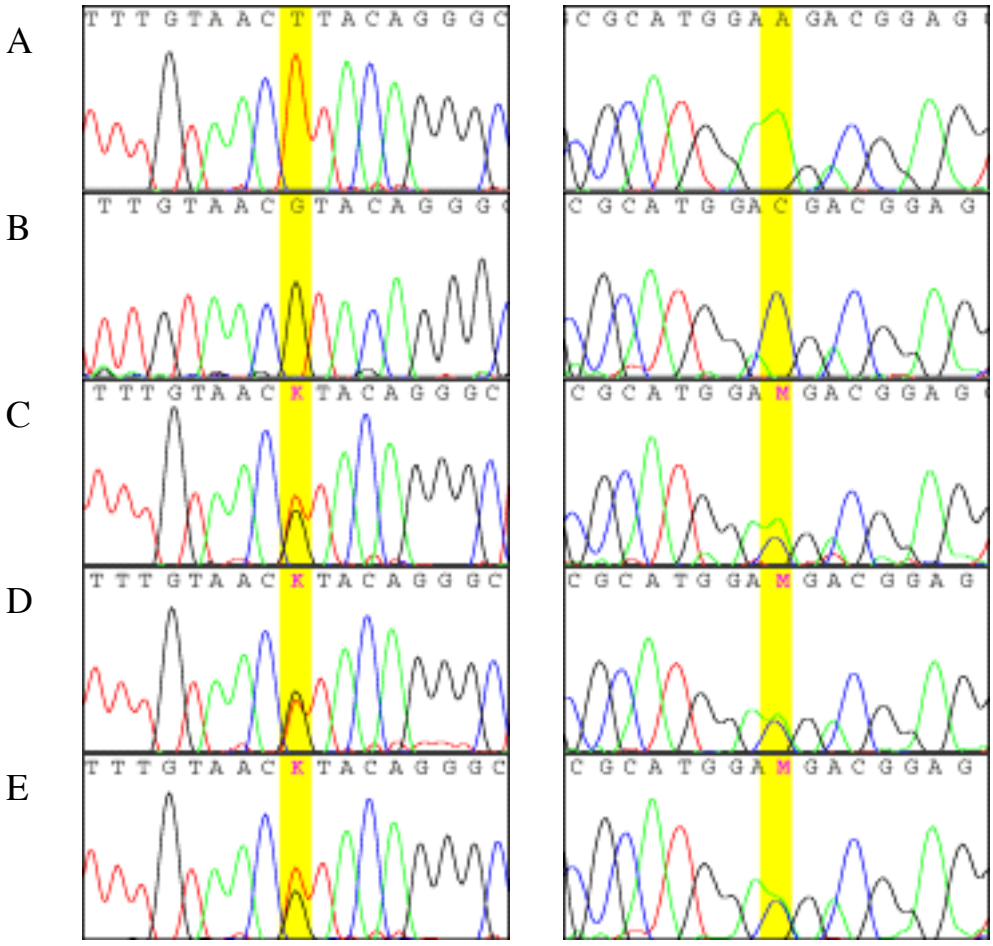
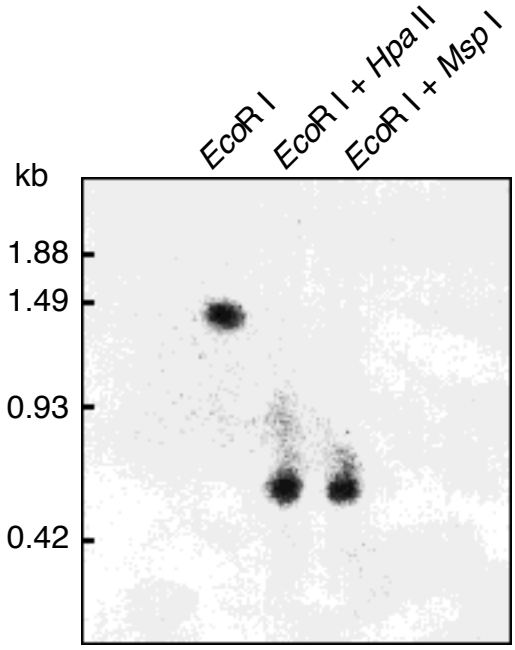


Figure 4.4 Methylation-specific PCR assays for upstream (left) and 3'-UTR (right) regions of mouse *Impact*. PCR products were subjected to direct sequencing. Each panel shows the chromatogram for the sequence around the 350-bp upstream region from the initiation codon. Undigested genomic DNAs from B6, JF, and (JF x B6) F1 were used as the template in A to C, respectively. *Hpa* II-digested genomic DNAs from (B6 x JF) F_1 and (JF x B6) F_1 were used in D and E, respectively. These regions are hypermethylated on both the maternal and paternal alleles.

A



B

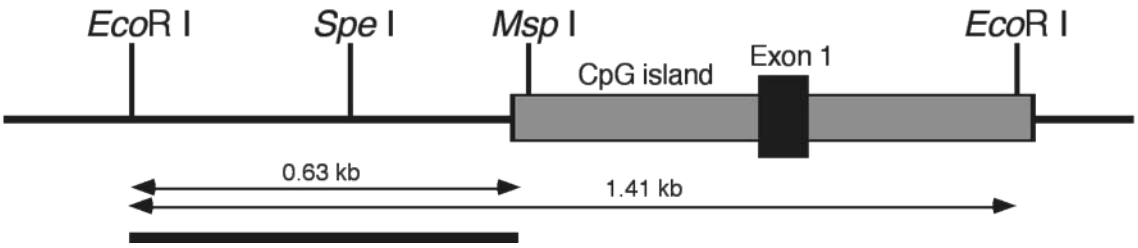


Figure 4.5 Southern blot hybridization analysis of the human *IMPACT* CpG island. Human genomic DNA was digested with *Eco*R I, *Eco*R I and *Hpa* II, or *Eco*R I and *Msp* I and electrophoresed in an agarose gel. The blot was probed with the ^{32}P-labeled 623-bp PCR product (A). A schematic representation of the genomic structure around the CpG island (B). The island has many *Msp* I sites, however, only the most 5' upstream *Msp* I site is shown. The sequence of the probe (solid bar) was checked by *Spe* I digestion, whose site is also indicated.

5. Identification of genes lying next to *IMPACT* and *Impact*

At the beginning of 2001, the draft sequences of the human genome were published in detail (Lanader *et al*. 2001; Venter *et al*. 2001). And a physical map of the mouse genome was recently constructed (Gregory *et al*. 2002). Utilizing this information, the NCBI Annotation Project has been developing model reference sequences for mRNAs *etc*. They are predicted by automated computational analysis using a gene prediction method: GenomeScan (Yeh *et al*. 2001), supported by EST evidence. I identified some genes lying next to human *IMPACT* and mouse *Impact* using the information generated by the project. Because almost every imprinted gene is embedded in a large imprinted region (Reik and Maher 1997; Vu and Hoffman 2000), it was quite interesting to elucidate genomic imprinting of genes around mouse *Impact*.

The first gene I found was the human *HRH4* (histamine H_4 receptor) gene, which resides only 7-kb downstream from 3' end of human *IMPACT*, and currently serves as the nearest neighbor of *IMPACT* (Fig. 5.1). It is transcribed in the same orientation as *IMPACT*. The *HRH4* gene encodes for an orphan G-protein-coupled receptor (GPCR) having the highest (37.4%) homology with H_3 receptor (Arrang *et al*. 1983) among known GPCRs. The *HRH4* was hunted making use of the similarity and it was already characterized (Nakamura *et al*. 2000; Oda *et al*. 2000).

A homology search using a protein-coding nucleotide sequence of human *HRH4* demonstrated that 365M4 harbored the mouse homologue in itself. Then I expanded the mouse contig of the subclones to clarify the genomic structure of the mouse *Hrh4* gene. Comparing with the human transcript, the genomic organization which has three exons spanning less than 20 kb was revealed, in which all the splice acceptor and donor sequences obey the Chambon rule (Breathnach and Chambon 1981) and all the junctions are conserved between the two species, although contiguous finished data has not been obtained yet. The gene locates about 14-kb downstream from 3' end of mouse *Impact*, and no other genes have been found in this region for the time being. It is also transcribed in the same orientation. Recently the mouse gene was identified together with rat and guinea pig ones and they were characterized comparatively (Liu *et al*. 2001).

Later the project mapped an intracellular lipid receptor in the upstream region of human *IMPACT*. It was the *OSBPL1* (oxysterol binding protein-like 1) gene. Human *OSBPL1* was cloned based on primary sequence similarity to the ligand-binding domain of *OSBP* (Levanon *et al*. 1990) and characterized in detail (Xu *et al*. 2001). This gene has a complicated means of expression, producing two quite different transcripts (Jaworski *et al*. 2001). *OSBPL1A* is made up of 15 exons. It encodes a protein of

437 amino acids and, like *OSBPL2*, contains only an OSBP domain. An alternative transcript, *OSBPL1B*, is produced from an upstream starting point. It includes 15 additional exons at the 5' end, skipping the initial exon (exon 16) of *OSBPL1A*, and gives rise to a 950-amino-acid peptide containing both the OSBP and PH domains. It adheres in a head-to-head orientation putting 29-kb intergenic region between the 5' ends of *OSBPL1B* and *IMPACT*. No other genes have been found in the region.

After the physical map of the mouse genome released, the mouse homologue *Osbpl1* gene was also annotated to the corresponding region in a head-to-head orientation to mouse *Impact*. But the model reference sequence is split into two transcripts. One is designated as *Osbpl1a* that corresponds to human *OSBPL1A*. And the other, whose tentative name is *LOC211564*, has sequence similarity to N-terminus of human *OSBPL1B*. These two transcripts are separated by 50-kb region in which a ribosomal protein gene is mapped. There have been no publications of the mouse transcripts yet, although some researchers submitted their ESTs or cDNA sequences. If the gene structures were conserved between mouse and human, mouse would also have the longer transcripts, *Osbpl1b*. *LOC211564* was predicted by computational methods supported by some EST evidence. One cDNA sequence which has a similarity to *LOC211564* has been submitted (GenBank acc. no. BC031735), but its ochre codon suggests that this gene encodes a protein of only 338 amino acids unlike human *OSBPL1B* (950 amino acids). The length of the cDNA is 1.2 kb and there is something like the poly(A) tail at the 3' terminus. It seemed that mouse didn't have the longer transcript that encodes both the PH and the OSBP domains. On the contrary Northern blot hybridization described in the next section revealed that the length of *LOC211564* is up to 4 kb, but not 1 or 2 kb. Subsequent RT-PCR analysis using the brain mRNAs identified the longer transcript, mouse *Osbpl1b*, whose hypothetical protein consists of 950 amino acids (Fig. 5.2). The deduced primary sequences were 98% identical and 92% similar to *OSBPL1B* and *Osbpl1b*. This gene consists of 29 exons and I designated all of them from exon 1 to exon 30 in proportion to human *OSBPL1B*. Both *Osbpl1b* and *OSBPL1B* do not have exon 16 which is the initial exon for *Osbpl1a* and *OSBPL1A*. Nucleotide sequences at all exon/intron junctions displayed similarity to the consensus boundary sequence with the GT-AG rule applied to the splicing point and are completely conserved in the two species. This identification clearly demonstrated that *LOC211564* is a part of *Osbpl1b* that is the mouse ortholog of human *OSBPL1B*. Also, I discovered five SNPs in the ORF, one of which alters an amino acid from asparagine to serine in JF, and a repetitive polymorphism just after the ochre codon.

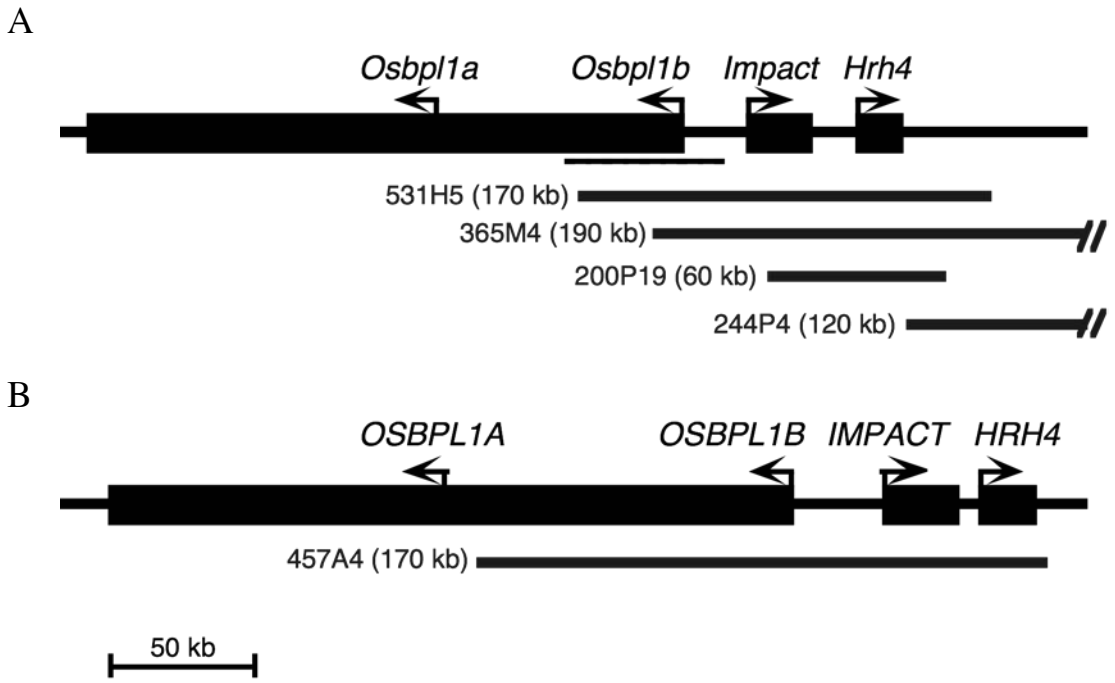


Figure 5.1 Overview of the mouse *Impact* (A) and human *IMPACT* (B) loci including their adjacent genes. Physical distance is indicated at the bottom of the diagrams. The units for each gene are shown as boxes, and the direction of transcription is denoted by arrows above the boxes. The region where *LOC211564* was mapped is indicated by a broken line. The locations of the screened BAC clones are depicted in blue. 244P4, 351H5, and 351H6 were obtained by the sequence of 365M4. But 351H6 turned out to be identical to 351H5 by their end sequences.

```
    GAGCGGCCCGCGTGGGTGACGGCCGCTCAGTGGGCGCGGCCGGGCCTGTGGTGGAGGCCG   60

    CCTCAGGAGCCGGGCCCGCGAGCACGCAGAAATGAACACAGAAGCAGAACAGCAGCTTCT  120
  1                                 M  N  T  E  A  E  Q  Q  L  L
    CCATCATGCCAGAAATGGCAATGCTGAAGAGGTGCGAAAGCTGCTGGCAGCCATGGCACG  180
 11  H  H  A  R  N  G  N  A  E  E  V  R  K  L  L  A  A  M  A  R
    GATGGAAGTGGTCGCTGACATCGACTGTAAAGGAAGAAGTAAGTCTAACTTGGGCTGGAC  240
 31  M  E  V  V  A  D  I  D  C  K  G  R  S  K  S  N  L  G  W  T
    ACCTCTTCACCTGGCATGCTATTTTGGACACAAACAAGTGGTCGAGGATCTGTTGAAGGC  300
 51  P  L  H  L  A  C  Y  F  G  H  K  Q  V  V  E  D  L  L  K  A
    TGGTGCAAAAGTGAACATGTTAAACGACATGGGGGACACGCCACTGCATCGAGCCGCCTT  360
 71  G  A  K  V  N  M  L  N  D  M  G  D  T  P  L  H  R  A  A  F
    CACAGGACGAAAGGAGCTGGTCTTGCTTCTTTTAGAATACGATGCTGACAGTACAGTTGT  420
 91  T  G  R  K  E  L  V  L  L  L  L  E  Y  D  A  D  S  T  V  V
    AAATGGAAGTGGACAGACAGCAAAAGAAGCCACCCATGACAAAGAAATCAGAAACATGCT  480
111  N  G  S  G  Q  T  A  K  E  A  T  H  D  K  E  I  R  N  M  L
    TGAAGCTGTGGAGAGGACTCAACAAAGAAAGCTTGAAGAATTGCTTTTAGGAGCAGCTAG  540
131  E  A  V  E  R  T  Q  Q  R  K  L  E  E  L  L  L  G  A  A  R
    AGAAGGAAGAACAGCCGAAGTATCAGCTCTGCTCAGCAGGCCCAATCCTCCTGATGTGAA  600
151  E  G  R  T  A  E  V  S  A  L  L  S  R  P  N  P  P  D  V  N
    CTGCTCAGATCAGCTAGGAAATACCCCCTTACACTGTGCCGCCTACCGGGCCCACAAGCA  660
171  C  S  D  Q  L  G  N  T  P  L  H  C  A  A  Y  R  A  H  K  Q
    GTGTGTCCTAAAGCTCCTGAGAAGCGGAGCTGACCCCAGCCTGAAGAACAAGAACGATCA  720
191  C  V  L  K  L  L  R  S  G  A  D  P  S  L  K  N  K  N  D  Q
    GAAACCTCTTGACCTTGCCCAGGGTGCTGAAATGAAGCATATTCTTGTTGGTAATAAGGT  780
211  K  P  L  D  L  A  Q  G  A  E  M  K  H  I  L  V  G  N  K  V
    TGTCCACAAAGCACTGAAACGCTACGAAGGCCCTCTCTGGAAGAGTTCAAGATTTTTTGG  840
231  V  H  K  A  L  K  R  Y  E  G  P  L  W  K  S  S  R  F  F  G
    CTGGAAGTTATTTTGGGTAGTGTTAGAACATGGAGTCCTTTCCTGGTATAGGAAGCAGCC  900
251  W  K  L  F  W  V  V  L  E  H  G  V  L  S  W  Y  R  K  Q  P
    TGATGCAGTCCATAATTCTTATCGCCAAGGATGCAAACACCTGACTCAAGCCGTGTGCAC  960
271  D  A  V  H  N  S  Y  R  Q  G  C  K  H  L  T  Q  A  V  C  T
    GGTGAAACCTACAGATAGCTGCCTCTTCTCCATTCGATGCTTTGATGACACTGTTCACTG 1020
291  V  K  P  T  D  S  C  L  F  S  I  R  C  F  D  D  T  V  H  C
    CTTCAGGGTTCCTAAGAACAGCGTACAGCAATCAAGAGAGAAGTGGCTGGAAGCAATTGA 1080
311  F  R  V  P  K  N  S  V  Q  Q  S  R  E  K  W  L  E  A  I  E
    AGAACACTCCGCATACAGCACTCACTACTGTTCCCAGGATCAGGTGACTGATGATGAAGA 1140
331  E  H  S  A  Y  S  T  H  Y  C  S  Q  D  Q  V  T  D  D  E  E
    AGAAGATGTAGTTTCGGCCATGGACTTGAAGGAGTCGTTAGCGAGAGCACAAACATGTCA 1200
351  E  D  V  V  S  A  M  D  L  K  E  S  L  A  R  A  Q  T  C  Q
    ACAGAGACTAGATAGGGAAATTTACAACTTTCTCAAAATGATTAAGGAGTGTGATGTGGC 1260
371  Q  R  L  D  R  E  I  Y  N  F  L  K  M  I  K  E  C  D  V  A
    CAAAGATATGCTTCCATCGTTTCTTCAGAAAGCTGATATTGTCTCCGAAGCTTCTAGAGA 1320
391  K  D  M  L  P  S  F  L  Q  K  A  D  I  V  S  E  A  S  R  E
    GACTTGTGTGGCGTTGAATGACTGTCTTAATCTCTTCACTAAACAGGAAGGGGTGAGGAA 1380
411  T  C  V  A  L  N  D  C  L  N  L  F  T  K  Q  E  G  V  R  N
    TTTTAAATTGGAACAGGAGCAAGAAAAAAACAAAATTTTGTCAGAAGCACTGGAGACTTT 1440
431  F  K  L  E  Q  E  Q  E  K  N  K  I  L  S  E  A  L  E  T  L
    AGCCACTGAGCATCATGAACTAGAACGATCCCTGGTTGAAGGATCCCCGCCTGTCAGCAT 1550
451  A  T  E  H  H  E  L  E  R  S  L  V  E  G  S  P  P  V  S  I
    CCTTAGTGAGGAGGAGTTCTATGATGCACTGTCAGGTTCCGAGTCCGAGGGGTCGCTGAC 1560
471  L  S  E  E  E  F  Y  D  A  L  S  G  S  E  S  E  G  S  L  T
    TTGCCTGGAAGCTGTGACAGCACACTCCTTTGAAGAGAACGAAGTGCCTGGGAGCAGCGG 1620
491  C  L  E  A  V  T  A  H  S  F  E  E  N  E  V  P  G  S  S  G
    AAAGCACAGAATGTCTGAAGGAAAAGACTGTGGTGGGGGAGATGCGCTCTCCAATGGCAT 1680
511  K  H  R  M  S  E  G  K  D  C  G  G  G  D  A  L  S  N  G  I
    CAAAAAGCACAGAACAAGCCTGCCATCACCTATGTTCTCCAGAAACGATTTCAGCATCTG 1740
531  K  K  H  R  T  S  L  P  S  P  M  F  S  R  N  D  F  S  I  W
    GAGCATCCTCAGAAAATGCATTGGGATGGAACTGTCCAAGATCACAATGCCAGTGATATT 1800
551  S  I  L  R  K  C  I  G  M  E  L  S  K  I  T  M  P  V  I  F
```

```
     TAATGAGCCTCTGAGCTTCCTGCAGCGGCTAACTGAATACATGGAGCACACGTACCTCAT 1860
571   N  E  P  L  S  F  L  Q  R  L  T  E  Y  M  E  H  T  Y  L  I
     CCACAAAGCCAGTTCACTTTCTGATCCTGTGGAAAGGATGCAGTGTGTGGCTGCATTTGC 1920
591   H  K  A  S  S  L  S  D  P  V  E  R  M  Q  C  V  A  A  F  A
     TGTGTCTGCTGTCGCCTCTCAGTGGGAGCGCACCGGAAAGCCCTTCAACCCGCTTCTGGG 1980
611   V  S  A  V  A  S  Q  W  E  R  T  G  K  P  F  N  P  L  L  G
     AGAGACTTATGAATTAGTTCGAGATGACCTTGGGTTTAGGCTCATCTCAGAACAGGTCAG 2040
631   E  T  Y  E  L  V  R  D  D  L  G  F  R  L  I  S  E  Q  V  S
     CCATCATCCCCCAATCAGTGCATTCCATGCAGAAGGGCTCAACAATGACTTCATCTTCCA 2100
651   H  H  P  P  I  S  A  F  H  A  E  G  L  N  N  D  F  I  F  H
     TGGCTCAATTTACCCCAAACTGAAGTTCTGGGGCAAGAGTGTAGAAGCAGAGCCTAAAGG 2160
671   G  S  I  Y  P  K  L  K  F  W  G  K  S  V  E  A  E  P  K  G
     AACCATCACCTTGGAGCTTTTAGACCACAACGAAGCATACACATGGACAAACCCCACCTG 2220
691   T  I  T  L  E  L  L  D  H  N  E  A  Y  T  W  T  N  P  T  C
     TTGTGTGCATAACATCATCGTGGGCAAGCTCTGGATTGAACAGTATGGCAACGTGGAAAT 2280
711   C  V  H  N  I  I  V  G  K  L  W  I  E  Q  Y  G  N  V  E  I
     CATAAACCACAAGACTGGGGACAAATGTGTGCTGAATTTTAAGCCATGTGGTCTTTTTGG 2340
731   I  N  H  K  T  G  D  K  C  V  L  N  F  K  P  C  G  L  F  G
     CAAGGAATTACACAAAGTTGAAGGCTACATACAAGATAAAAGCAAAAAGAAGCTCTGTGC 2400
751   K  E  L  H  K  V  E  G  Y  I  Q  D  K  S  K  K  K  L  C  A
     TCTCTATGGGAAGTGGACTGAGTGCTTGTACAGCGTAGACCCTGCCACTTTTGATGCTTA 2460
771   L  Y  G  K  W  T  E  C  L  Y  S  V  D  P  A  T  F  D  A  Y
     CAAAAAAAATGATAAGAAAAACACAGAAGAGAAGAAGAACAGCAAACAGACAAGCTCATC 2520
791   K  K  N  D  K  K  N  T  E  E  K  K  N  S  K  Q  T  S  S  S
     TGAGGAGTCTGATGAAATGCCAGTGCCAGATTCCGAGAGCGTATTCATTATCCCTGGAAG 2580
811   E  E  S  D  E  M  P  V  P  D  S  E  S  V  F  I  I  P  G  S
     TGTGCTCCTGTGGCGGATAGCCCCGCGGCCTCCCAACTCTGCTCAGATGTATAACTTTAC 2640
831   V  L  L  W  R  I  A  P  R  P  P  N  S  A  Q  M  Y  N  F  T
     CAGCTTTGCCATGGTTTTGAATGAAGTCGACAAGGAGATGGAGAGTGTGATTCCCAAGAC 2700
851   S  F  A  M  V  L  N  E  V  D  K  E  M  E  S  V  I  P  K  T
     TGACTGCAGGCTACGGCCTGACATCAGAGCCATGGAGAACGGAGAGATAGATCTAGCTAG 2760
871   D  C  R  L  R  P  D  I  R  A  M  E  N  G  E  I  D  L  A  S
     TGAAGAAAAGAAACGGCTTGAGGAAAAGCAAAGAGCAGCCCGCAAGAACAGGTCCAAGTC 2820
891   E  E  K  K  R  L  E  E  K  Q  R  A  A  R  K  N  R  S  K  S
     GGAAGAGGACTGGAAGACAAGGTGGTTCCATCAAGGTCCTAACCCCTACAGTGGAGCACA 2880
911   E  E  D  W  K  T  R  W  F  H  Q  G  P  N  P  Y  S  G  A  Q
     GGACTGGATTTATTCTGGCAGCTACTGGGACAGAAACTACTTCAATTTGCCTGATATTTA 2940
931   D  W  I  Y  S  G  S  Y  W  D  R  N  Y  F  N  L  P  D  I  Y
     TTAAAGTACAGAGAAGTCAAGGTGTTTGCTAATCTAAATAAGTCTTAAGCTTAAGATTTT 3000
951   *
     TAAATGTTTTTCCTTGGTTTCTACTCCTCTATAATTTGCGTTTCACCCAATAGGGCAAGG 3060

     CATCTGGTACAGTGGGGATGACCTGAAAGAGAAGAGGCAGTGCTGTGCCACAGCTGCCCC 3120

     C                                                             3121
```

Figure 5.2 The primary structure of the *Osbpl1b* cDNA. Nucleotide sequence and deduced amino acid sequence are shown. Numbers shown in the left and right indicate those for amino acid residues and nucleotides, respectively. Five SNPs are written in red. These C, A, C, G, and C (from 5' to 3') are replaced with the alternative pyrimidine or purine base, respectively, in JF. The repetitive polymorphism is indicated in blue. JF has two units of the 10-bp sequence.

6. Tissue distribution studies

To examine the tissue distribution of human *IMPACT* mRNA, I performed Northern blot hybridization using a probe derived from its ORF. The probe detected two messages. One is 4 kb long, showing good coincidence with the cDNA (accession. no. AB026264), and the other is 2 kb (Fig. 6.1). Both are detected in all tissues examined and display an identical tissue preference pattern. I found that a 3'-UTR probe detected only the longer RNA and thus assume that the shorter RNA is generated through differential polyadenylation or something else. Although a modest tissue preference was observed in the distribution of *IMPACT* mRNA, its expression is basically ubiquitous. This is in marked contrast with mouse *Impact* (Hagiwara *et al*. 1997), which is preferentially expressed in the adult brain. These results may raise the possibility that the human clone is derived from a paralog rather than the ortholog of *Impact*. However, the striking structural conservation observed not only in the ORF but also in the 3' UTR, the syntenic localization, and homological correspondence of location and transcriptional orientation of neighboring genes support the idea that *IMPACT* is the ortholog of mouse *Impact*. Also, every effort to find other homologues of *Impact* has so far been unsuccessful, except for a pseudogene found on human chromosomes 5 and 12 (Fig. 6.2). I thus assume that the *IMPACT* gene reported here is the human ortholog of mouse *Impact*.

Northern blot and PCR analyses of human *HRH4* were already performed (Oda *et al*. 2000). According to the report, the expression was detected only in peripheral blood leukocytes by Northern blot analysis, which forms a striking contrast to the tissue distribution of *IMPACT*. Human *IMPACT* mRNA was less expressed in leukocytes among the tissues examined. The regulatory mechanisms of gene expression seem to be quite different between the two genes. PCR analysis showed a little expression in thymus, small intestine, spleen, and colon (Oda *et al*. 2000), all of which are tissues related to the immune system. A similar result was presented in rodents (Liu *et al*. 2001). These facts are notable as contrasted with that of the *Hrh3* and *HRH3* genes, which were exclusively found in brains (Lovenberg *et al*. 1999).

Expressions of the human OSBP genes were surveyed extensively (Jaworski *et al*. 2001). Their RT-PCR analysis revealed that the *OSBPL1A* gene was expressed ubiquitously. In the article they reported that the gene has a complicated means of expression, producing two very different transcripts. They suggested the longer one included 15 additional exons at the 5' end, skipping the first exon of the alternative transcript, and presumptive promoters were upstream of each initial exon. The expression of the shorter valiant, in contrast to the longer product of the same gene, was

limited to brain and retina. I did not know whether these transcripts were expressed in peripheral blood leukocytes, so I performed Northern blot hybridization using a probe which can detect both *OSBPL1A* and *OSBPL1B* simultaneously (Fig. 6.3). The result showed that *OSBPL1A* (4 kb) is preferentially expressed in skeletal muscle, heart, and brain. On the other hand, *OSBPL1B* (5 kb) was barely detected on one membrane, but diffuse bands were observed in placenta, kidney, liver, and skeletal muscle on another filter (data not shown). This difference might be due to physiological condition when human RNAs were extracted. At any rate, little messages of these genes were detected in leukocytes, from which mRNA would be extracted for allelic expression analysis in this study. I also performed the same analysis for the mouse *Osbpl1a* and *Osbpl1b* (Fig. 6.4) and the probe detected both *Osbpl1a* (3 kb) and *Osbpl1b* (4 kb) in some tissues. *Osbpl1b* seemed to be much less expressed than *Osbpl1a* in the tissues where the shorter message was detected, *e.g.* heart and brain. Then I prepared an *Osbpl1b*-specific probe which didn't hybridize the shorter mRNA. The analysis using this ascertains that the mouse *Osbpl1* gene has an alternative longer transcript, *Osbpl1b*, which was first identified in this study.

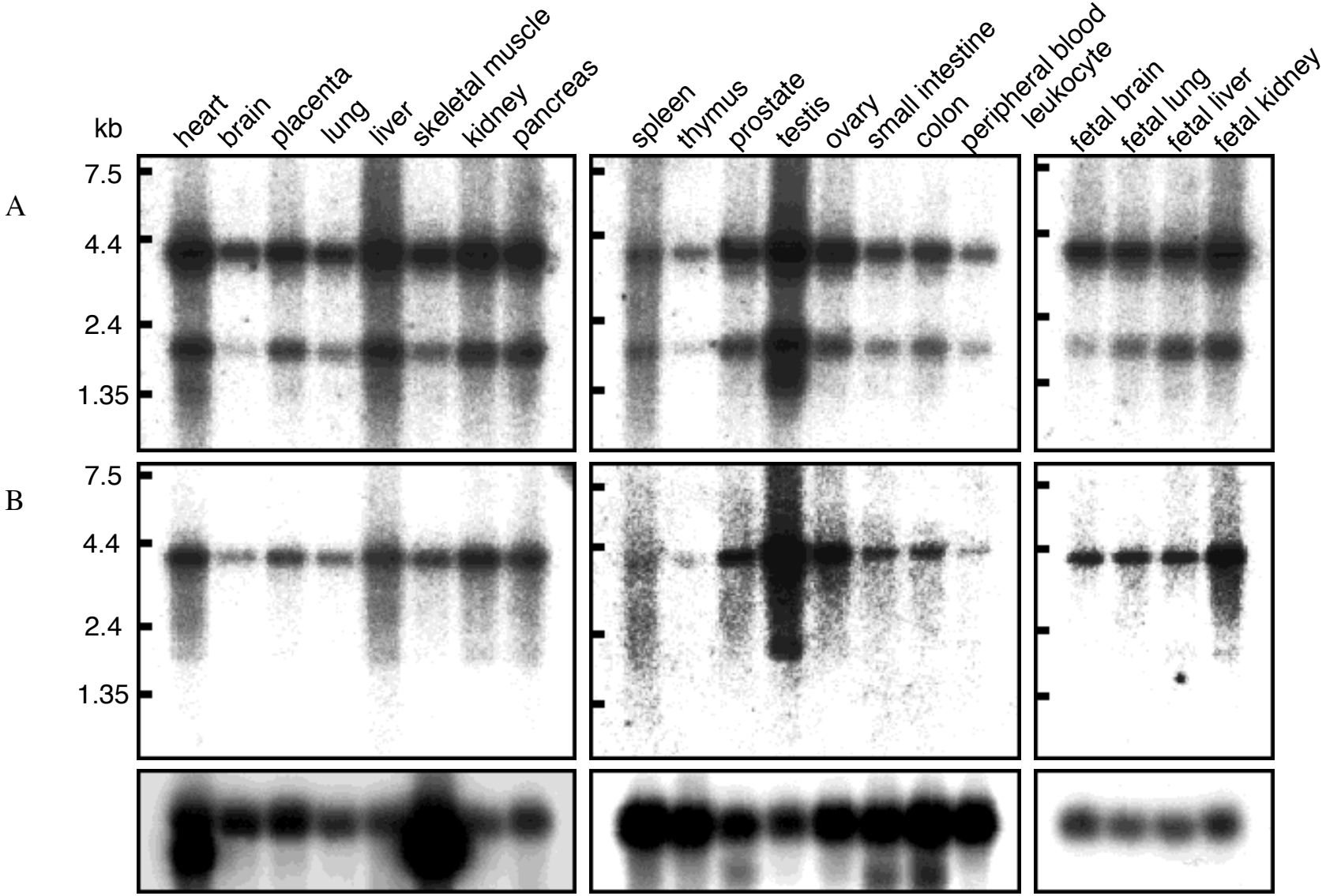

kb
heart
brain
placenta
lung
liver
skeletal muscle
kidney
pancreas
spleen
thymus
prostate
testis
ovary
small intestine
colon
peripheral blood leukocyte
fetal brain
fetal lung
fetal liver
fetal kidney
A
7.5
4.4
2.4
1.35
B
7.5
4.4
2.4
1.35

Figure 6.1 Tissue distribution of the human *IMPACT* mRNA. Northern blot hybridization using RNAs from the indicated tissues was performed with hybridization probe derived from the ORF (A) or 3' UTR (B) of *IMPACT*. The former probe detected two distinguishable mRNAs. Lower panels show the control hybridization with human *β-actin* probe.

```
          1            M  A  E  G  D  A  G  S  D  Q  R  Q  N  E  E  I
IMPACT cDNA TCCAGGGGCCACATGGCTGAGGGGGACGCAGGGAGCGACCAGAGGCAGAATGAGGAAATT 87
            |||| |||| |||||||||||||||||||||||||  |||| |||||| ||||||||||
pseudogene  TCCA-GGGCTACATGGCTGAGGGGGACGCAGGGAGTAACCAAAGGCAGGATGAGGAAATT

         17 E  A  M  A  A  I  Y  G  E  E  W  C  V  I  D  D  C  A  K  I
IMPACT cDNA GAAGCAATGGCAGCCATTTATGGCGAGGAGTGGTGTGTCATTGATGACTGTGCCAAAATA 147
            |||||||||| ||||||| |||||||||| ||||||||||||||||||||||||||||||
pseudogene  GAAGCAATGGTAGCCATTGATGGCGAGGAATGGTGTGTCATTGATGACTGTGCCAAAATA

         37 F  C  I  R  I  S  D  D  I  D  D  P  K  W  T  L  C  L  Q  V
IMPACT cDNA TTTTGTATTAGAATTAGCGACGATATAGATGACCCCAAATGGACACTTTGCTTGCAGGTG 207
            |||||||||||||||||| |||||||||||||||||||||||||||||||||||| |||
pseudogene  TTTTGTATTAGAATTAGCTACGATATAGATGACCCCAAATGGACACTTTGCTTGCAAGTG

         57 M  L  P  N  E  Y  P  G  T  A  P  P  I  Y  Q  L  N  A  P  W
IMPACT cDNA ATGCTGCCGAATGAATACCCAGGTACAGCTCCACCTATCTACCAGTTGAATGCTCCTTGG 267
            |||||||| |||||  |||||||||||||||||||| |||| |||||||||||||| ||||
pseudogene  ATGCTGCCAAATGAGGACCCAGGTACAGCTCCACCTGTCTATCAGTTGAATGCTCATTGG

         77 L  K  G  Q  E  R  A  D  L  S  N  S  L  E  E  I  Y  I
IMPACT cDNA CTTAAAGGGCAAGAACGTG-CGGATTTATCAAATAGCCTTGAGGAAATATATAT 320
            |||||||||||||||| || |  ||||||  |||||||| ||||||||||||||
pseudogene  CTTAAAGGGCAAGAACTTGTCC-ATTTATTGAATAGCCT-GAGGAAATATATAT
```

Figure 6.2 Alignment of the human *IMPACT* cDNA and its pseudogene found by genome walking. Because this pseudogene does not have a CpG island, it can be readily obtained by such a technique utilizing PCR. It covers the whole coding region, howerer, exons 1 to 4 are shown in this figure. Numbers written in the left and right indicate those for amino acid residues and nucleotides, respectively. There are many alterations of amino acids which are shown in red. Details are discussed later. This pseudogene is mapped to 5q13.3. And another one is discovered on 12p12.1.

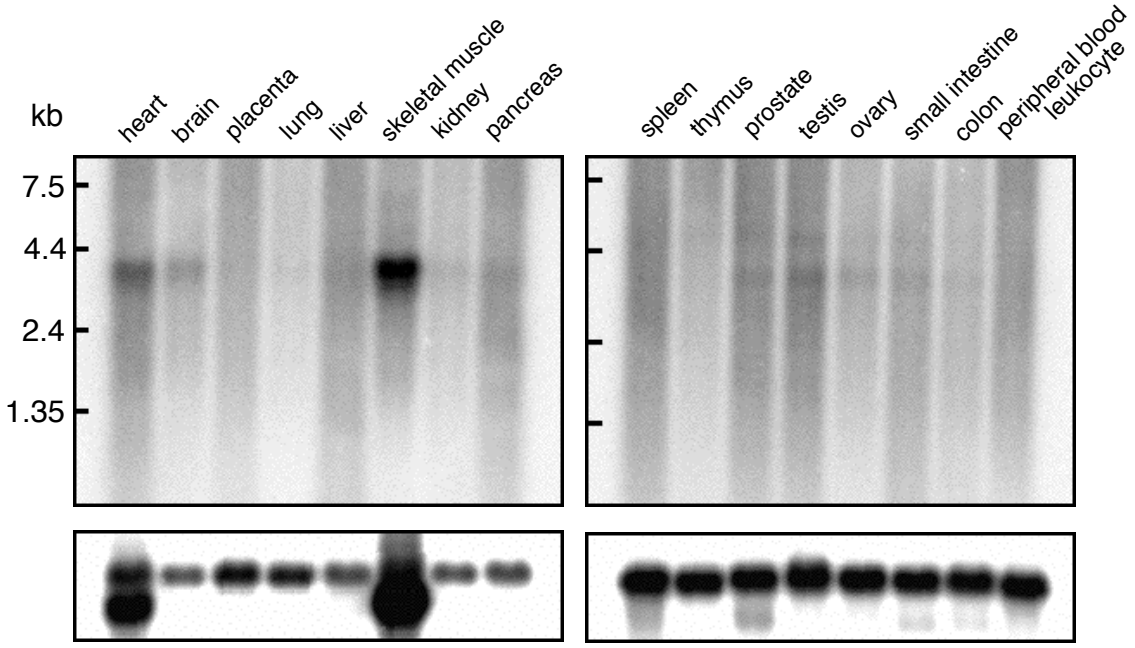


Figure 6.3 Tissue distribution of human *OSBPL1* mRNA by Northern blot hybridization. The probe used here is derived from a region that both the shorter and the longer transcripts have. The lower panels show the control hybridization with human *β-actin* probe.

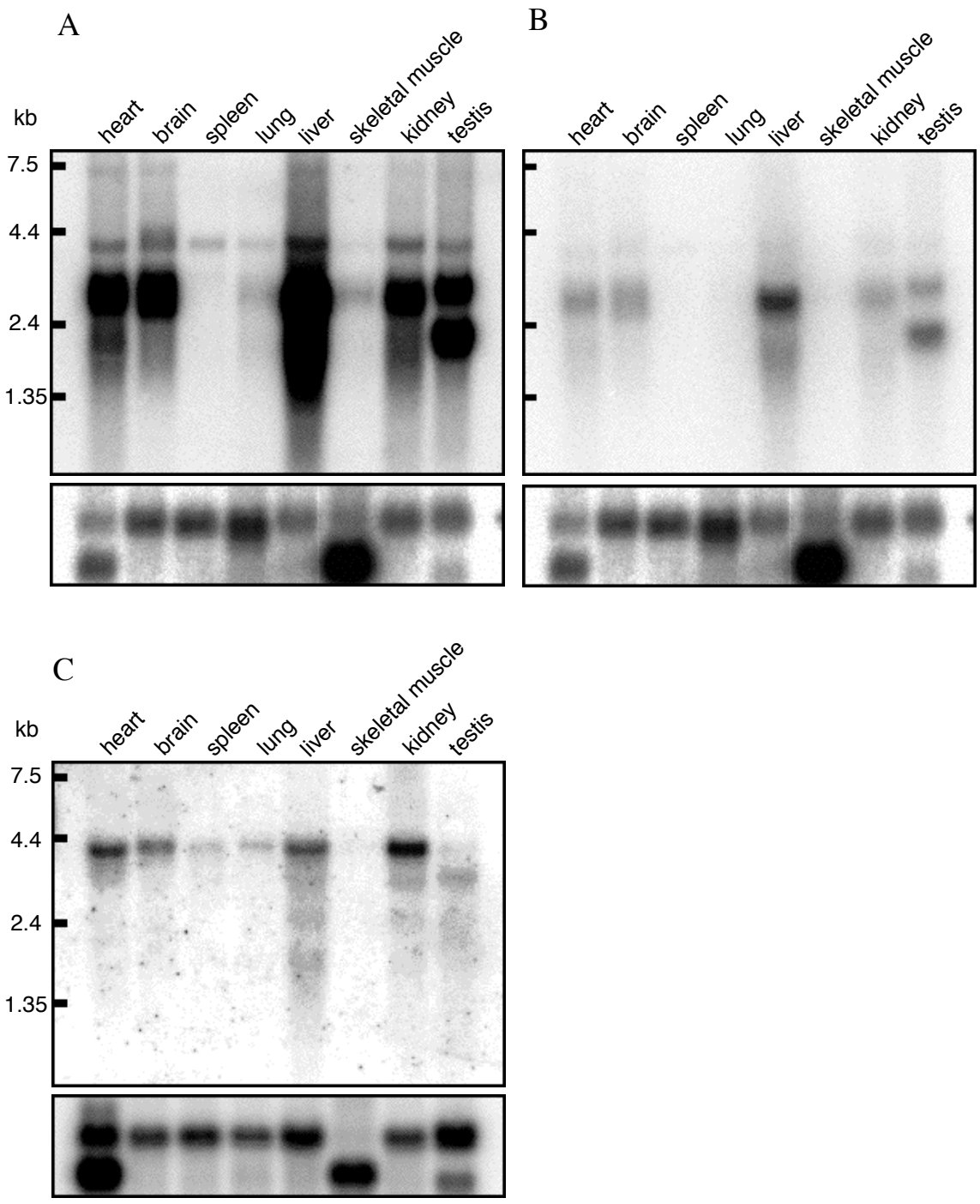


Figure 6.4 Tissue distribution of mouse *Osbpl1* mRNA by Northen blot hybridization. B is a shor-exposure result of A. The probe used in A and B is derived from a region that both *Osbpl1a* and *Osboplb* have. On the other hand, the probe used in C is derived from an *Osbpl1b*-specific region. Each lower panel shows the control hybridization with human *β-actin* probe.

7. Allelic expression and DNA methylation analyses of *Hrh4*

In order to examine genomic imprinting of human *HRH4*, I first had to find some polymorphisms in the gene. Fortunately, I could find a SNP in one Japanese family (family N). Since I failed to find this SNP in any databases, it seemed to be a novel one. Intriguingly, it was found in the ORF and alters an amino acid from a histidine to an arginine (Fig. 7.1). This gene encodes for a typical seven-transmembrane G-protein-coupled receptor. The region containing the SNP is remarkably conserved in rodents and humans (Liu *et al*. 2001), however, the histidine is not conserved for all these species. The arginine allele was inherited from the mother to her daughter. As this gene is preferentially expressed in leukocytes (Oda *et al*. 2000), I extracted total RNA from her blood.

According to the database, the human *HRH4* gene (accession no. AB044934) consists of only three exons. All the splice donors and acceptors are shown in Table 2. I designed PCR primers, one on the second exon and the other on the third exon in order to avoid amplification from genomic DNA. These exons are separated by a 7.8-kb intron. Then the RT-PCR product was directly sequenced (Fig. 7.2). As one can see in the figure, the result indicates that both alleles are transcribed, although her paternal allele seemed to be preferentially expressed. I also performed sequencing with a reverse primer, which supported the biallelic expression. I thus concluded human *HRH4* is expressed biallelically, although I cannot exclude the possibility of imprinted expression in other tissues and other individuals. This conclusion is conceivable taking account of the fact that neighboring *IMPACT* is expressed from the both alleles.

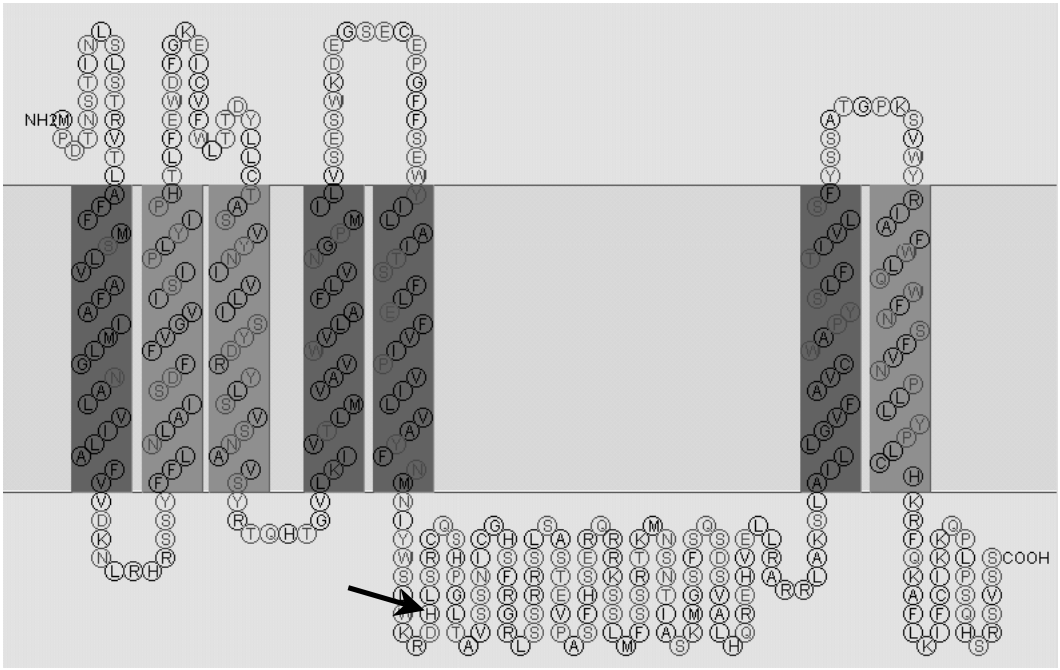
Human *HRH4* has no CpG island. Furthermore, its promoter region has few CpG dinucleotides. As for a 3-kb upstream region including the first exon, there is no *Hha* I and *Hpa* II sites and the ratio of observed versus expected CpG dinucleotides (Obs/Exp CpG) is 0.12 in contrast to 0.3 for the 30-kb *IMPACT* locus. It might be nonsense to discuss its DNA methylation status, however, I performed the methylation-specific PCR assay using McrBC (Sutherland *et al*. 1992). *Hpa* II-McrBC PCR assay was developed and two DMRs were successfully identified on the human chromosome 21 (Yamada *et al*. unpublished data). While the cleavage with *Hpa* II is blocked by methylation, McrBC cleaves only the methylated DNAs. Thus, the PCR from *Hap* II-digested and McrBC-digested DNAs amplifies methylated and unmethylated alleles, respectively. Practically, its application to the *Impact* CpG island clearly discriminates between the hypermethylated maternal allele and the undermethylated paternal allele (Fig. 7.3). As for the promoter region of *HRH4*, McrBC digestion diminished the subsequent PCR amplification (Fig. 7.4), which indicates that the upstream region of

human *HRH4* is hypermethylated.

Next I intended to know whether the mouse homolog is imprinted or not. I extracted spleen RNAs from four kinds of mice, JF, B6, and their reciprocal hybrid mice, JF x B6 and B6 x JF. I designed a forward PCR primer on the second exon and the other reverse one on the last exon. The RT-PCR product has a polymorphic *Hha* I site, so I performed a RFLP analysis that showed biallelic expression (Fig. 7.5). And the products were directly sequenced to confirm this result. The sequencing also supported that mouse *Hrh4* is expressed from both the maternal and paternal alleles. Hence, similar to its human homologue, mouse *Hrh4* is not imprinted.

Lastly I examined the DNA methylation status of mouse *Hrh4*. In contrast with *Impact*, this mouse gene also lacks any CpG islands. Because there are four *Hpa* II sites around the promoter region, I performed the methylation-specific PCR using spleen DNAs. After the treatment with *Hpa* II and McrBC, digested genomic DNAs were used as templates for the PCR assay. Each digestion didn't abolish the following amplification (Fig. 7.6). Unfortunately I couldn't find any polymorphisms enabling to perform RFLP analysis, so the PCR products were directly sequenced. The results indicated that not only *Hpa* II but also McrBC cleaved neither the maternal nor paternal alleles of the hybrid mice. It is difficult to tell the methylation status using McrBC digestion, so I adopted the consequence obtained by *Hpa* II digestion, *i.e*. both alleles are hypermethylated around the promoter region of mouse *Hrh4*.

A



B

```
                 ***::*:  *.**. :******** +**** **       **   **  *
Mouse            LEFLLPVISVAYFNVQIYWSLWKRRALSRCPSHAGF-STTSSSASGHLHR
Rat              LEFLLPVSLVVYFSVQIYWSLWKRGSLSRCPSHAGF-IATSSRGTGHSRR
Guinea Pig       LEFLIPILLVAYFSAHIYWSLWKREKLSRCLSHPVLPSDSSSSDHGHSCR
Human            LEFVIPVILVAYFNMNIYWSLWKRDHLSRCQSHPGL-TAVSSNICGHSFR
 (Family N)                                 R
```

Figure 7.1 Deduced amino acid sequences of the *HRH4* genes. The secondary structure of human *HRH4* (A) was predicted by SOSUI system (Mitaku and Hirokawa 1999). A SNP alters a histidine, which is indicated by an arrow, to an arginine. Multiple sequence alignment (B) was performed by CLUSTAL W (Thompson *et al*. 1994). Family N has an allele that encodes an arginine instead of a histidine in histamine H_4 receptor. Its position is indicated by a plus sign.

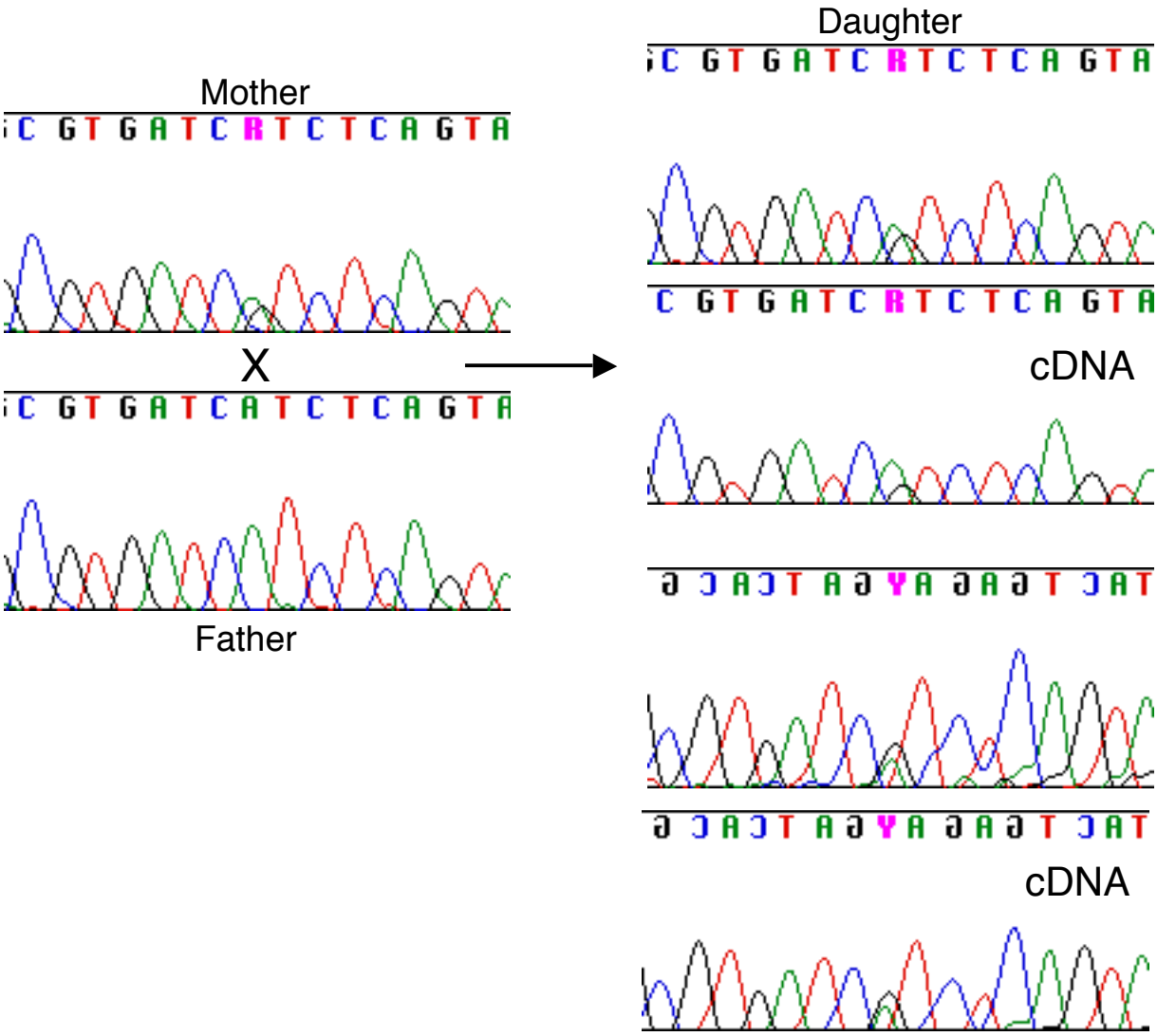


Figure 7.2 Allelic expression analysis of human *HRH4*. Father's, mother's, daughter's genomic DNA sequences, and daughter's cDNA sequences were obtained by direct sequencing of PCR products. Daughter's sequences were determined using both forward and reverse primers. The SNP can be seen in the center of each chromatogram.

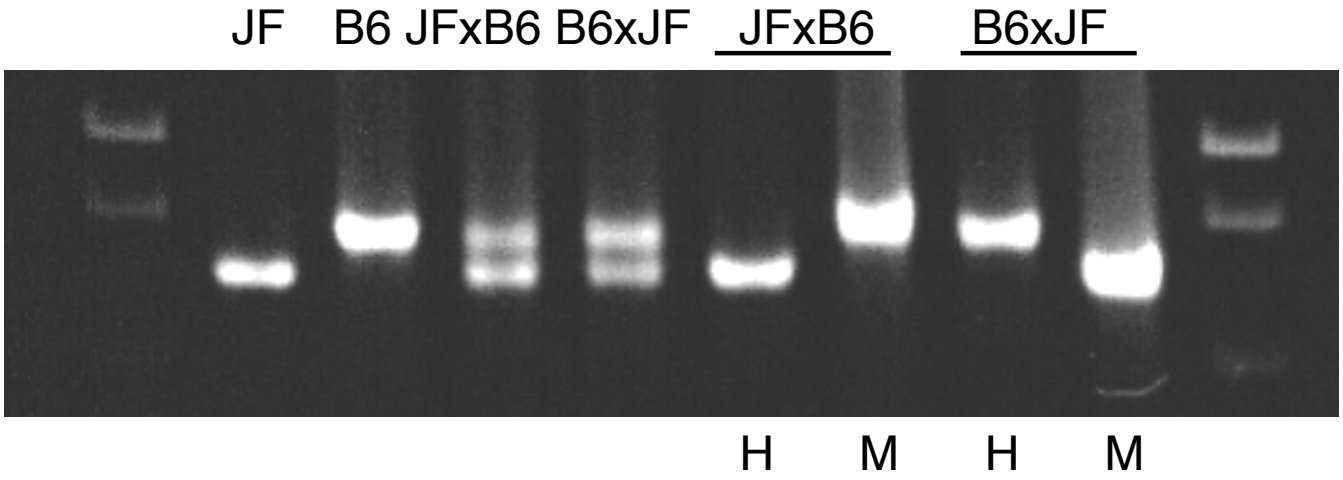


Figure 7.3 DNA Methylation analysis by *Hpa* II-McrBC PCR assay. Here I applied this assay to the *Impact* CpG island. The PCR from *Hap* II-digested and McrBC-digested DNAs amplifies methylated and unmethylated alleles, respectively. Capital letters beneath the picture, H and M, indicate PCR templates, *Hpa* II-digested and McrBC-digested genomic DNAs, respectively. λ DNA-*Sty* I digest was used as size standard.

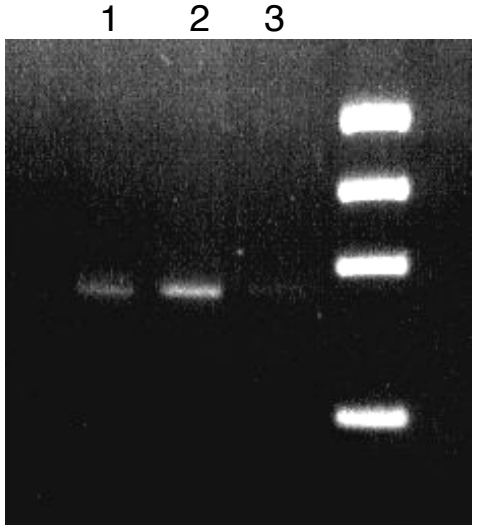


Figure 7.4 DNA methylation analysis of the promoter region of human *HRH4*. 807-bp PCR products amplified from non-digested, *Hpa* II-treated, and McrBC-treated human genomic DNAs were electrophoresed in lanes 1 to 3, respectively, on agarose gel. *Hae* III-digested ϕX174 DNA was used as size standards. This result indicates that the region is hypermethylated on the maternal and paternal alleles.

Table 2. Aligned Exon-Intron Organizations of Mouse *Hrh4* and Human *HRH4*

Gene	Exon	Size(bp)	5' Splice Donor	Intron	Size(bp)	3' Splice Acceptor
Hrh4	1	253	TTC CTC GTG G gtaagttatg...	1		...ccttgtacag GT TTG ATT TCC
HRH4	1	247	TTC TTT GTG G gtaagttata...	1	7871	...ccctgtgcag GT GTG ATC TCC
Hrh4	2	164	GTT TCA AAT GCT gtaagtcata...	2		...tttcttgtag GTG TCT TAT AGG
HRH4	2	164	GTC TCA AAT GCT gtaagtcgaa...	2	7802	...tttcttctag GTG TCT TAT AGA
Hrh4	3	1121				
HRH4	3	854				

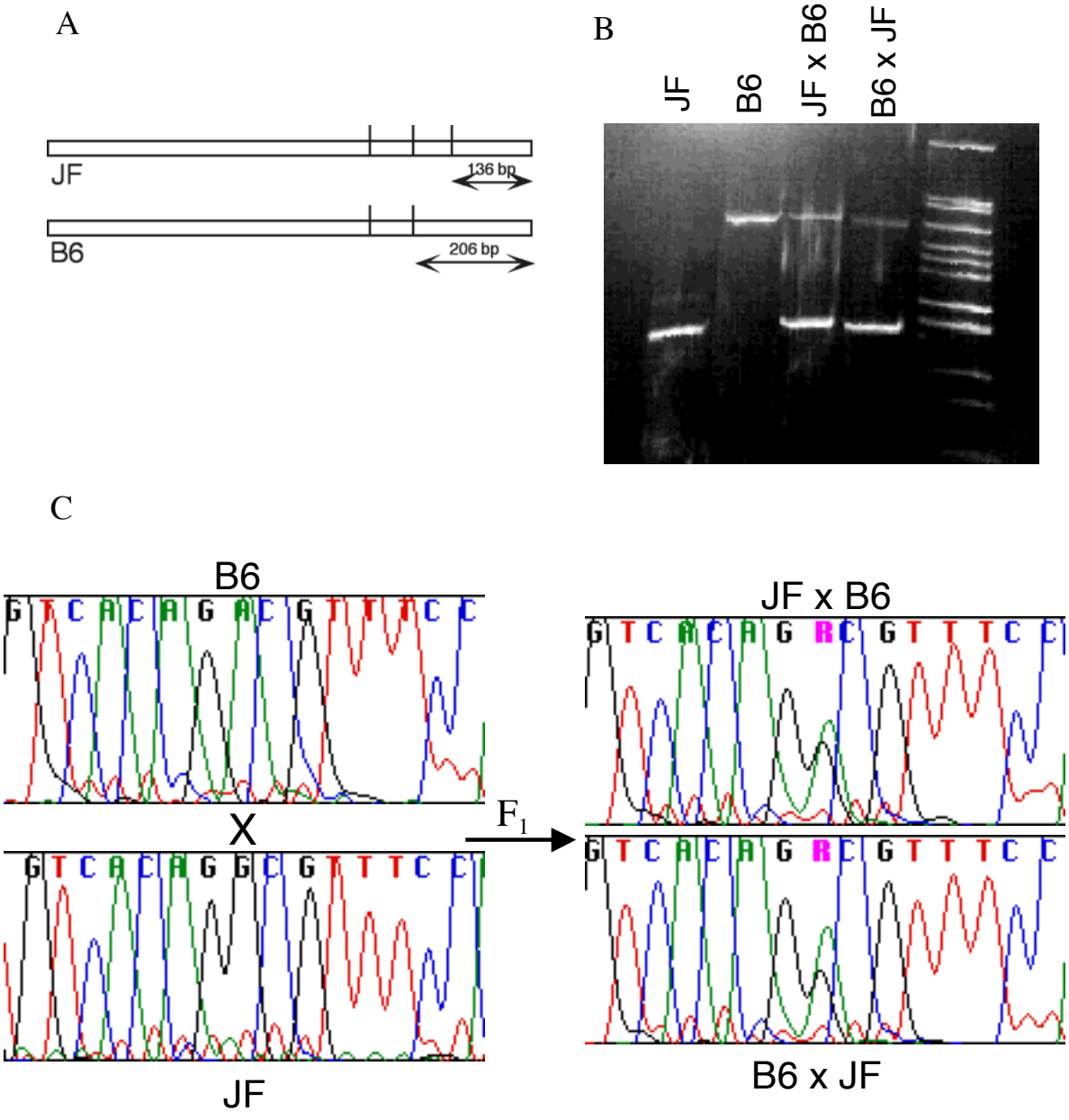


Figure 7.5 Allelic expression analyses of mouse *Hrh4*. *Hha* I sites in the 834-bp fragment amplified from the RT-PCR are delineated (A). *Hha* I treatment of B6 product gives a 206-bp fragment. On the other hand, JF gives a shorter one. They could be easily distinguished by electrophoresis (B). The RT-PCR products of JF, B6, JF x B6, and B6 x JF were directly sequenced (C) to confirm biallelic expression of this gene.

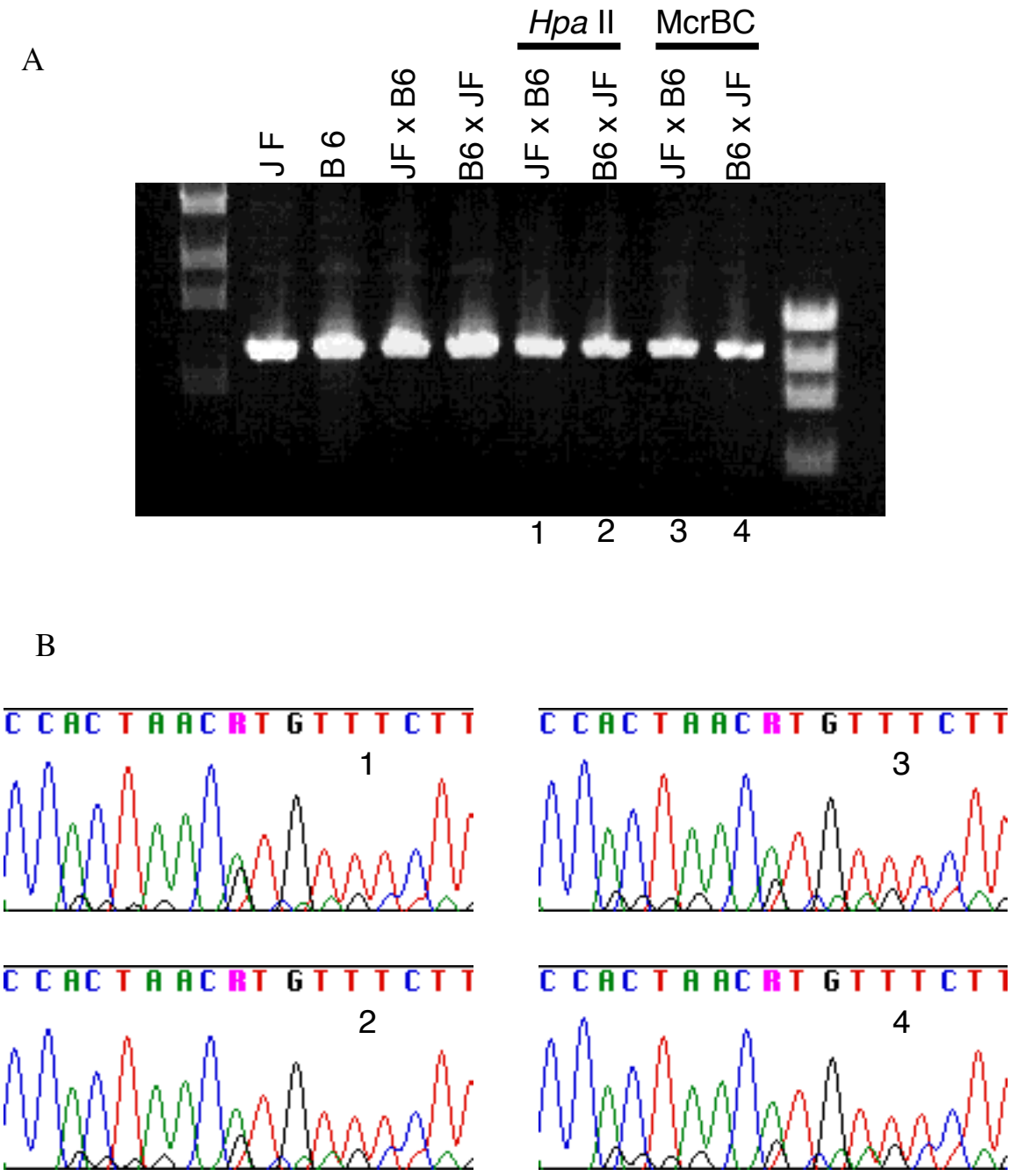


Figure 7.6 DNA methylation analysis of the promoter region of mouse *Hrh4*. Genomic PCR products are shown by electrophoresis (A). Right four are amplified from DNAs treated with specified endonucleases. *Sty* I-digested λ and *Hae* III-digested ϕX174 DNAs were used as size standards in the leftmost and rightmost lanes, respectively. The right four products, numbered 1 to 4, were directly sequenced to check parent-of-origin-specific methylation (B). An available SNP can be seen in the center of chromatograms.

8. Allelic expression and DNA methylation analyses of *Osbpl1*

To survey genomic imprinting of human *OSBPL1*, I tried to find polymorphisms in the transcribed region of this gene. I found an available SNP in two Japanese families (families N and S). They have the same polymorphism in the 3' UTR of human *OSBPL1*, which resides 803 bp downstream from the ochre termination codon. This region is shared between the shorter *OSBPL1A* and the longer *OSBPL1B* transcripts, so it is impossible to discriminate between these two, and expressional quantity in peripheral leukocytes is very low for both of them (Fig. 6.3). In any case, I extracted total RNAs from their blood and performed an allelic expression analysis. I designed PCR primers, one on exon 29 and the other on the last one, exon 30, to avoid amplification from genomic DNAs. Direct sequencing of the 975-bp RT-PCR products indicated biallelic expression of human *OSBPL1* (Fig. 9.1). Because both *OSBPL1A* and *OSBPL1B* have their own CpG island around each promoter region, their methylation status might give suggestion of their allelic expression.

The CpG island of human *OSBPL1A* spans the first exon (exon 16) and the first intron. Exon 16 is a part of 5' UTR of *OSBPL1A* and it doesn't serve as an exon of *OSBPL1B*. In other words, the CpG island resides in the 15th intron of *OSBPL1B*. Its length is 0.80 kb, %GC is 75, and Obs/Exp CpG is 0.85. It has a poly(A) stretch close by, but no apparent repeats. I performed DNA methylation analyses using *Hpa* II and McrBC. The region that would be amplified by the PCR has five *Hpa* II sites. The enzymatic digestion abolished the 621-bp amplification, which means that this region is undermethylated on both chromosomes (Fig. 8.2). Three nonspecific fragments were observed when non-treated genomic DNA was used as template. McrBC treatment also seemed to cut this region. This result may indicate that unmethylated region is not long, because the endonuclease can digest up to 2-kb region if both its ends have methylcytosines.

The CpG island associated with human *OSBPL1B* encompasses the first exon of this long transcript. Its length is 0.74 kb, %GC is 71, and Obs/Exp CpG is 0.95. The same DNA methylation assay was performed. The 444-bp PCR product covering the latter half of the island has five *Hpa* II sites. The digestion thoroughly abolished the PCR. Therefore the CpG island of *OSBPL1B* is also undermethylated on both chromosomes.

Then I examined allelic expression of the mouse transcripts using brain mRNA. I made PCR primers, one on exon 23 and the other on exon 29. This region is transcribed for both the short and long transcripts. However, the expressional quantity of *Osbpl1a* is twenty times as much as that of *Osbpl1b* in brain (Fig. 6.4). So the

contribution of *Osbpl1b* is negligible. The 793-bp RT-PCR products obtained with these primers contains a T/C SNP at position 2272 in exon 24 (Fig. 5.2). Direct sequencing of the SNP of the reciprocal hybrid mice revealed that mouse *Osbpl1a* was expressed biallelically (Fig. 8.3). I also designed *Osbpl1b*-specific primers, one on exon 3 and the other on exon 7, which gave a 326-bp RT-PCR product. This fragment has a T/C SNP at position 355 in exon 4, which serves as a polymorphic *Alu* I site, so I performed RFLP assay (Fig.8.4). It looked as though *Osbpl1b* was also expressed from both chromosomes, but the paternal allele was a little preferentially expressed. This inclination was clearly shown by direct sequencing of the RT-PCR products of not only a lot of JF x B6 and B6 x JF but also a JF x ICR and an ICR x JF hybrid mice (Fig. 8.5). Indeed mouse *Osbpl1b* was expressed biallelically, but the paternal allele was slightly more transcribed than the maternal one. A quantitative PCR analysis also demonstrated the allele-specific preference.

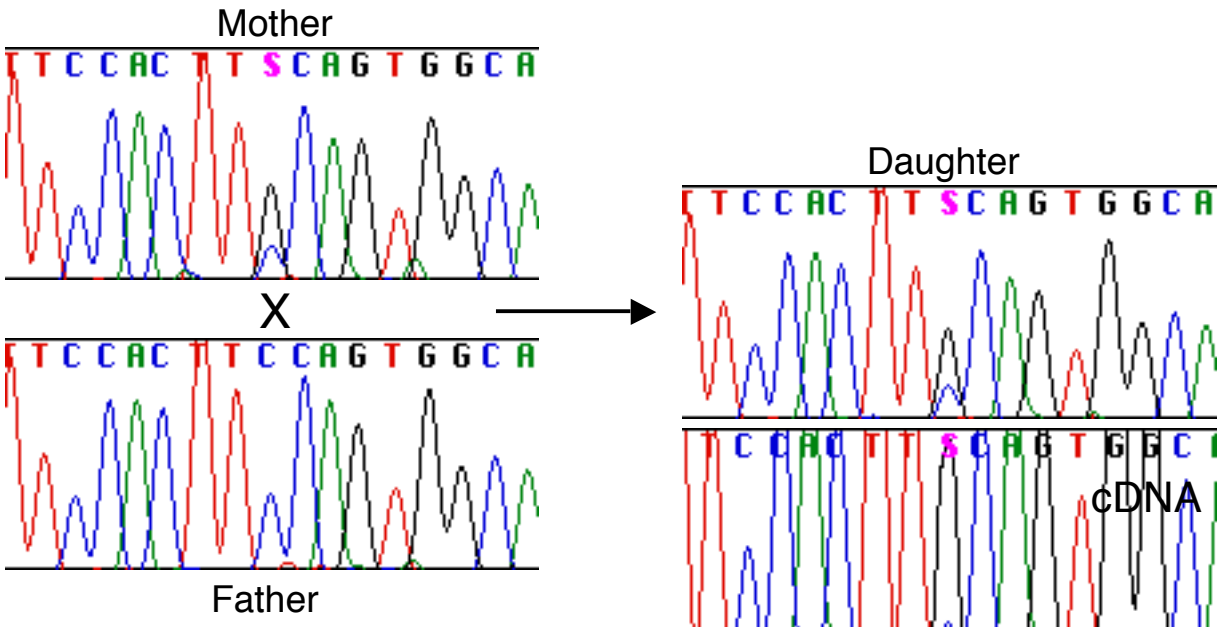
The CpG island of mouse *Osbpl1a* spans the promoter region and the first exon (exon 16). The island resides in the 15th intron of *Osbpl1b* like the human genome. Its length is 0.38 kb, %GC is 71, and Obs/Exp CpG is 0.80. The same DNA methylation assay was performed with a PCR product containing four *Hpa* II sites (Fig. 8.6). This 555-bp region covers the whole CpG island. *Hpa* II digestion diminished the amplification from genomic DNAs, which suggested both chromosomes were undermethylated in the CpG island.

The CpG island associated with mouse *Osbpl1b* lies encompassing the first exon. Its length is 0.56 kb, %GC is 66, and Obs/Exp CpG is 0.91. The methylation assay was done with a 300-bp PCR fragment which has three *Hpa* II sites and three *Hha* I sites (Fig. 8.7). *Hpa* II or *Hha* I treatment abolished the amplification from genomic DNAs. On the other hand, this locus survived McrBC digestion to serve as the templates for subsequent PCR. The PCR products amplified from McrBC-treated genomic DNAs were directly sequenced in order to check allele-specific methylation status. But the sequencing result showed McrBC cut neither the maternal nor the paternal alleles. A similar assay was performed with a B6-specific primer combination. The methylation status of this CpG island was also examined by sodium bisulfite treatment (Fig. 8.8). Bisulfite-treated B6 genomic DNA was used as the template and the PCR product was directly sequenced. All cytosines in CpG dinucleotides were detected as thymines like other cytosines. All these results suggested that both chromosomes escaped from methylation in the *Osbpl1b* CpG island. That is usual for a common CpG island, although this gene is preferentially expressed from the paternal allele.

These allelic expression and DNA methylation analyses are summarized

schematically (Fig. 8.9). Apart from the *Impact* DMR, no more characteristic tandemly repeated structures were found in not only CpG islands but all of the regions examined in this study.

A



B

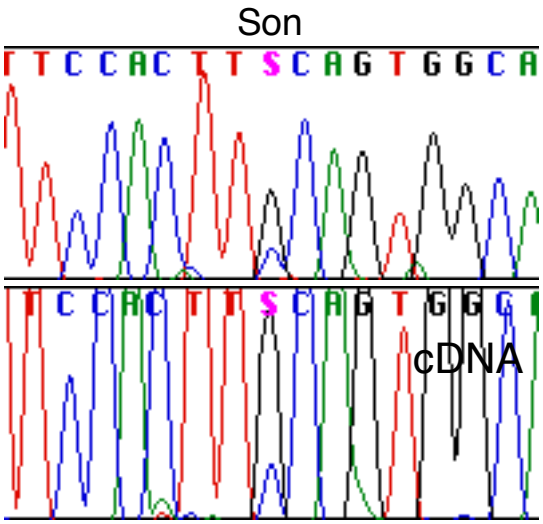


Figure 8.1 Allelic expression analysis of human *OSBPL1*. Direct sequencing results of genomic DNAs and children's cDNAs were depicted by chromatograms. Family N (A) and family S (B) are shown. The SNP can be seen in the center of each chromatogram.

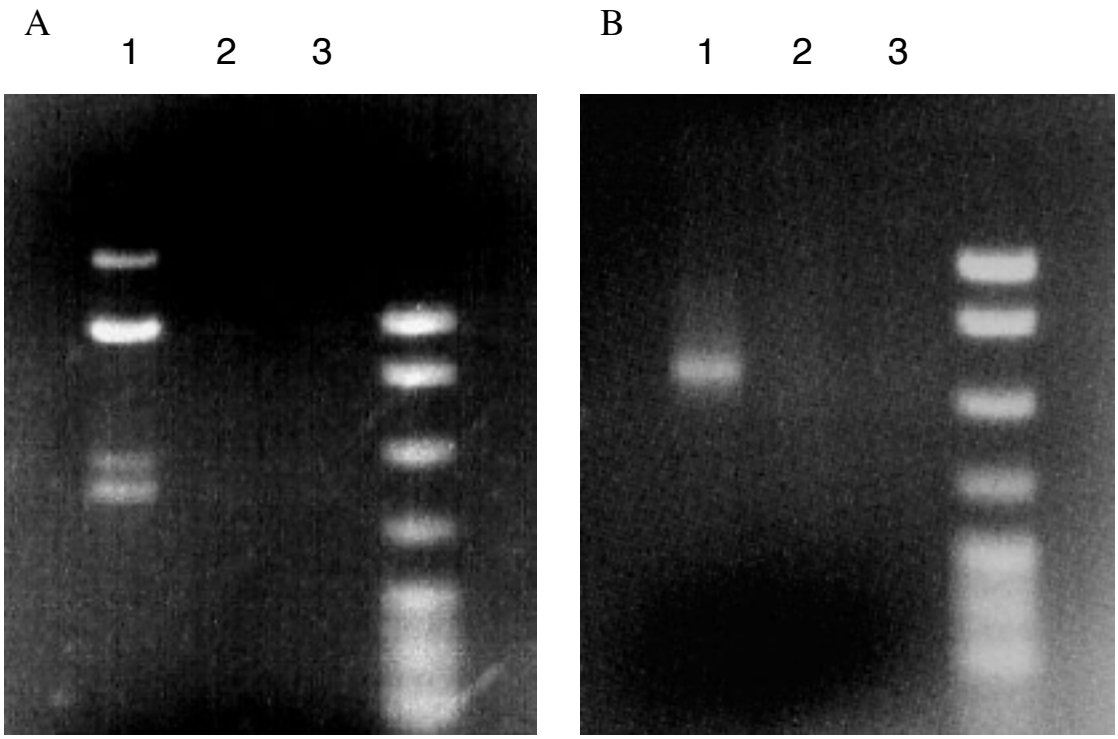


Figure 8.2 Methylation-specific PCR assay for the CpG islands of human *OSBPL1*. PCR products amplified from native genomic DNA and those digested with *Hpa* II or McrBC were subjected to agarose gel electrophoresis in lanes 1 to 3, respectively. pBR322 DNA-*Msp* I digest was used as a size standard in the rightmost lanes. No amplification from *Hpa* II-treated genomic DNAs implies undermethylation of both the *OSBPL1A* (A) and *OSBPL1B* (B) CpG islands.

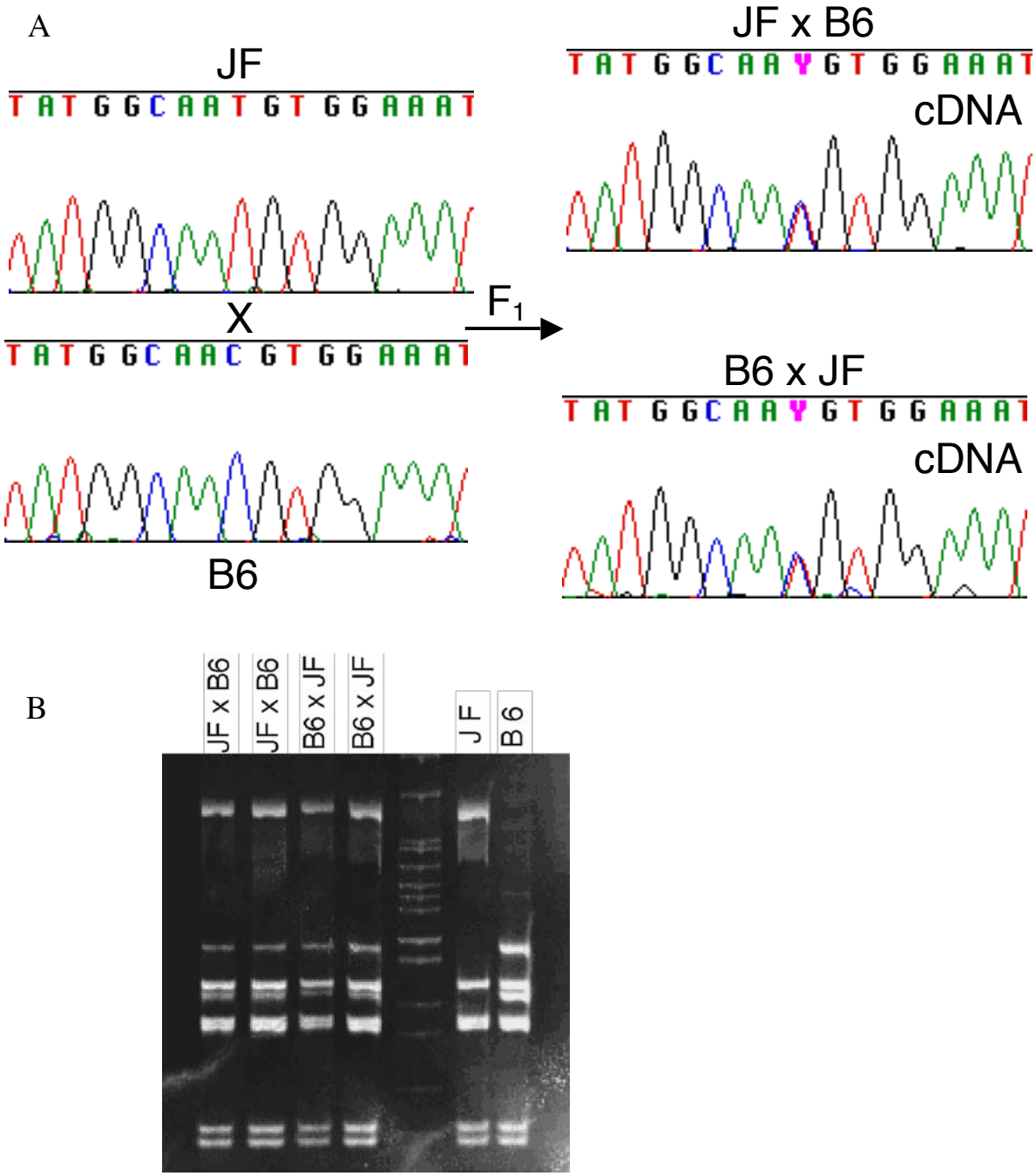
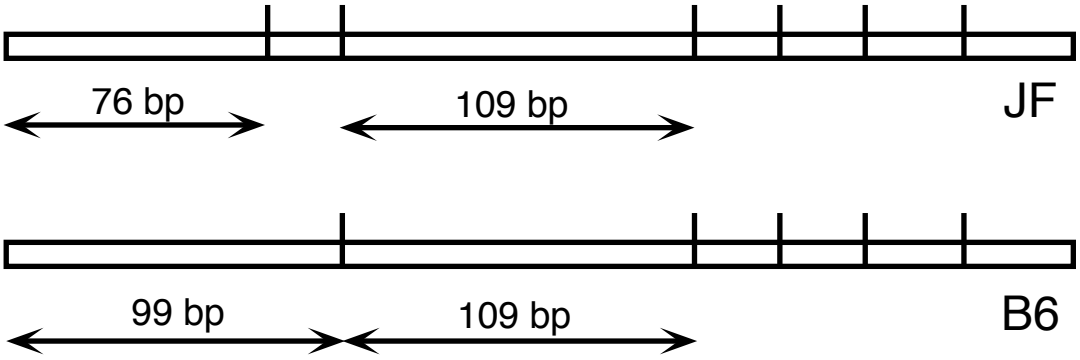


Figure 8.3 Allelic expression analysis of mouse *Osbpl1a*. The RT-PCR products were directly sequenced and investigated using the SNP that can be seen in the center of each chromatogram (A). RFLP analysis was also done (B). The PCR products amplified from B6 has an additional polymorphic *Alu* I site that divides the JF's longest 266-bp fragment into 144-bp and 122-bp fragments. pBR322 DNA-*Msp* I digest was used as a size standard. In reciprocal F_1 hybrid mice, JF x B6 and B6 x JF, both the maternal and paternal alleles were transcribed.

A



B

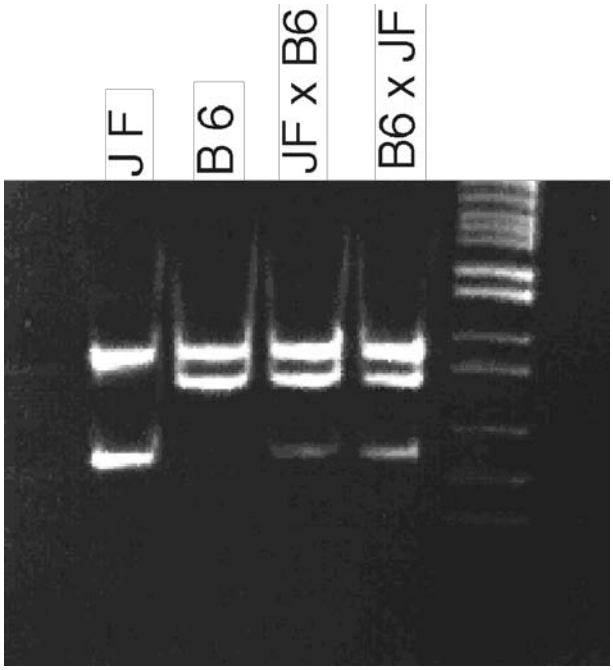


Figure 8.4 RFLP analysis of mouse *Osbpl1b*. *Alu* I sites in the 326-bp fragment amplified from the *Osbpl1b* mRNA are shown (A). JF allele harbors an additional *Alu* I site which gives a 76-bp fragment. PCR products from brain cDNAs of JF, B6, JF x B6, and B6 x JF (from left to right) were digested with *Alu* I and subjected to PAGE (B). pBR322-*Msp* I fragments were used as a size standard. Note that small fragments whose sizes were not indicated were not resolved on the gel used.

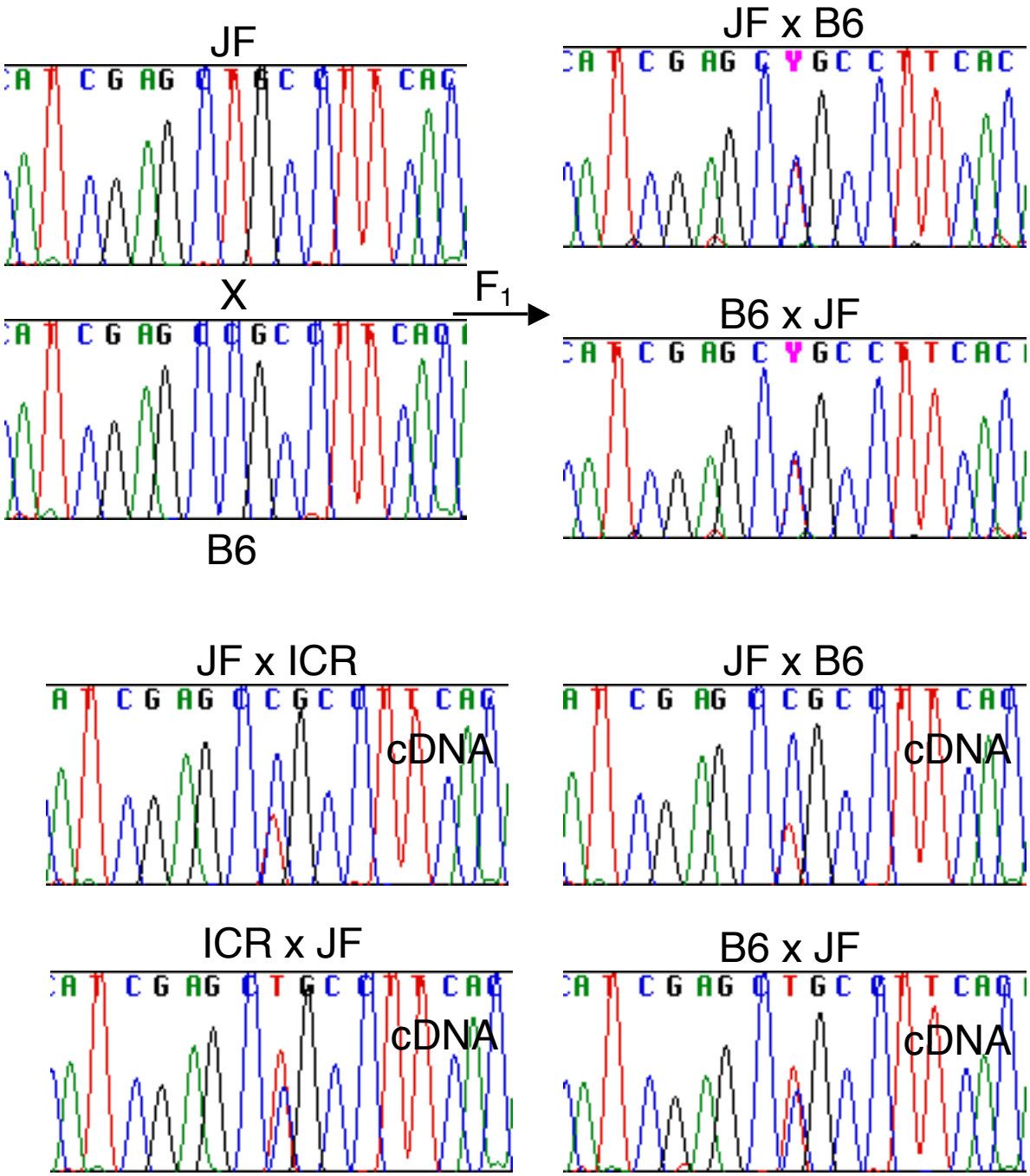


Figure 8.5 Allelic expression analysis of mouse *Osbpl1b*. The RT-PCR products were directly sequenced (lower four chromatograms) and investigated using the SNP that can be seen in the center of chromatograms. In reciprocal F_1 hybrid mice, the paternal alleles were preferentially expressed. Sequencing chromatograms of hybrid genomic DNAs were also shown on upper right for comparison.

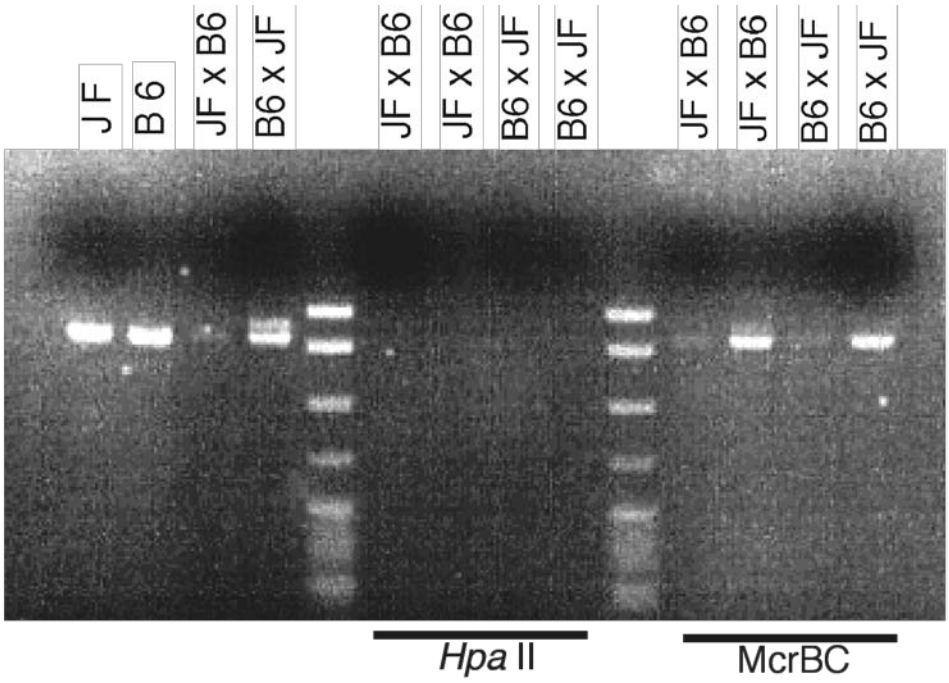
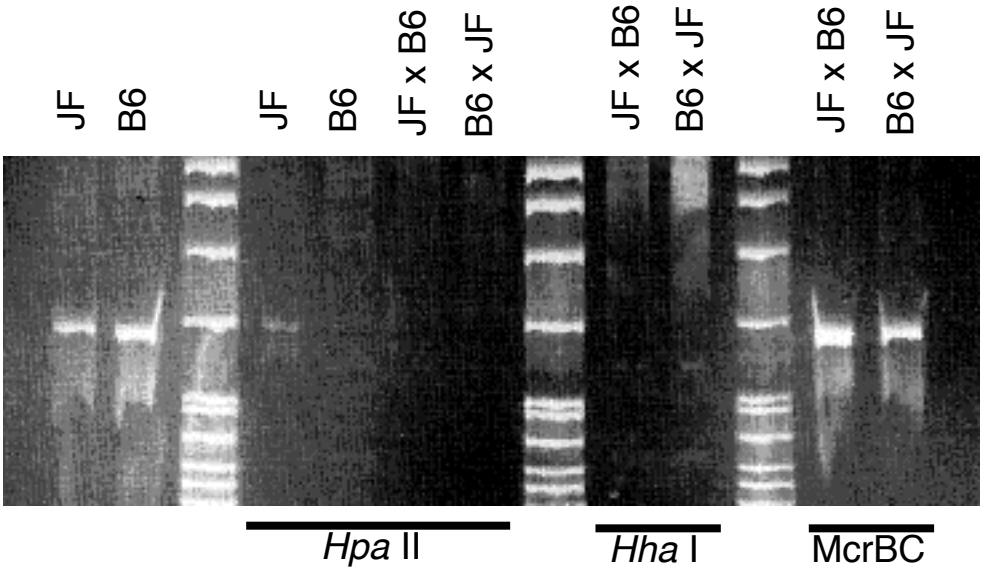


Figure 8.6 DNA methylation analysis of the mouse *Osbpl1a* CpG island. The DNA methylation status was examined by 555-bp PCR amplification of endonuclease-treated genomic DNAs. While McrBC did not cut this region, *Hpa* II digested and abolished the following amplification. pBR322-*Msp* I fragments were used as a size standard.

A



B

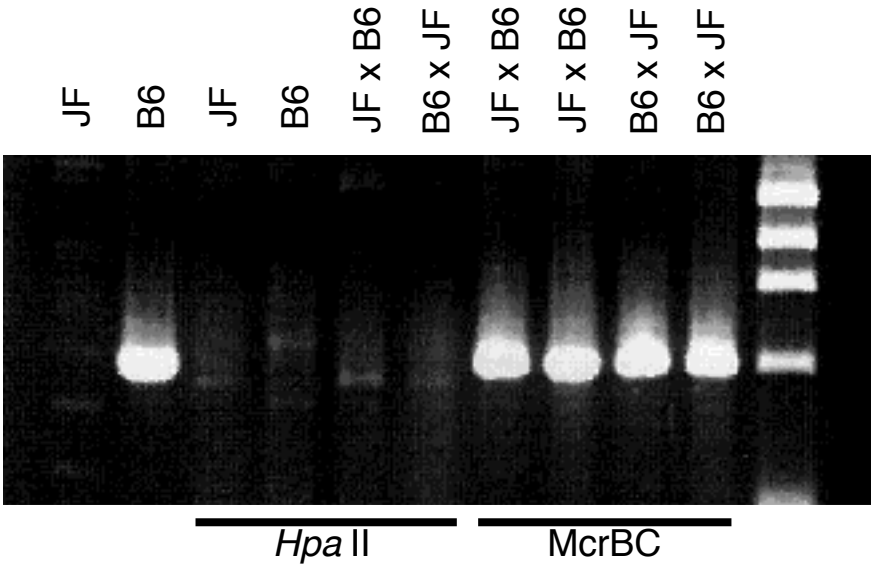


Figure 8.7 DNA methylation analysis of the mouse *Osbpl1b* CpG island. The DNA methylation status was examined by PCR amplification of endonuclease-treated genomic DNAs. While McrBC did not cut this region, both *Hpa* II and *Hha* I digested and abolished the following amplification (A). PCR was performed with B6-specific primer, too (B). These results suggested that the CpG island was undermethylated on both chromosomes.

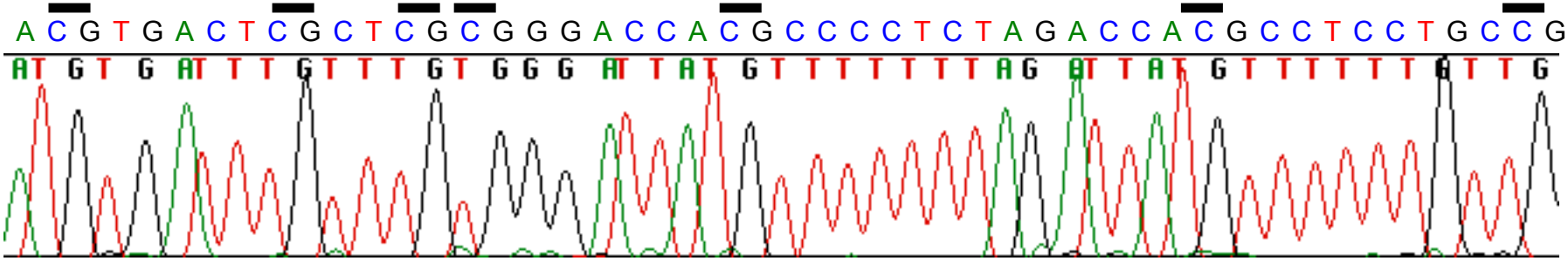


Figure 8.8 Bisulfite based cytosine methylation analysis of the mouse *Osbpl1b* CpG island. Since the paternal allele is preferentially expressed, the methylation status of the CpG island was examined by sodium bisulfite treatment. Bisulfite-treated B6 genomic DNA was used as the template and the PCR product was directly sequenced. The genomic DNA sequence is shown above the chromatogram, in which seven CpG dinucleotides are indicated by solid bars. All these cytosines were detected as thymines like other cytosines. This result indicates that the both maternal and paternal alleles of the *Osbpl1b* CpG island are not methylated in the same manner of usual CpG islands.

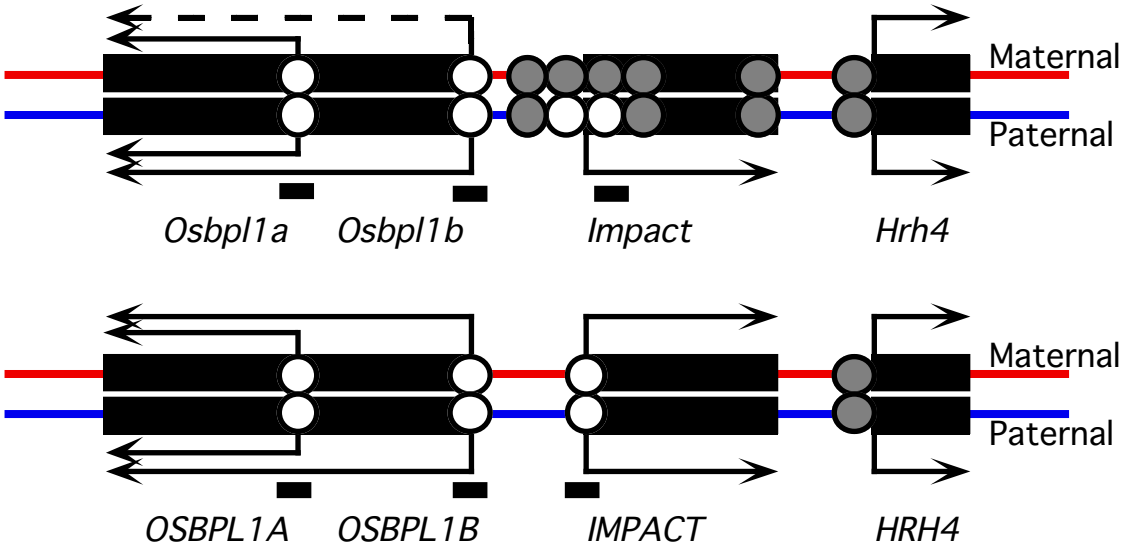


Figure 8.9 Summary of the allelic expression and DNA methylation analyses of all genes examined in this study. Mouse (upper) and human (lower) genomic regions spanning about 300-400 kb are schematized. Arrows above and below the boxes denote the directions of transcription for maternally and paternally expressions, respectively. Weak transcription is illustrated by a broken arrow. Black circles stand for hypermethylation and white circles stands for undermethylation. The locations of CpG islands are indicated by bold lines.

Discussion

A pseudogene and a cryptic exon found by genome walking

At the beginning of this study, I intended to obtain genomic clones using STSs which were developed from cDNA sequences. For the PCR screening of BAC libraries, an STS should be at least 150-bp long. However, lengths of mammalian exons are usually shorter than the minimum with the exception of the last exons. Since no information was available on the exon-intron organizations of the genes, I performed genome walking from the 5' ends of the cDNA clones. As for human *IMPACT*, I first identified a genomic region which had similarity to the *IMPACT* cDNA (Fig. 6.2). It, however, had many alterations of amino acid residues and seemed to have no intron, suggesting that it might be a pseudogene of *IMPACT*. I failed to walk from its first exon after all. The exon turned out to be a part of a CpG island whose %GC is nearly 70, after I determined the genomic DNA sequence. When I determined the sequence of the mouse *Impact* CpG island, I had some difficulty. In various genome projects, sequences of such regions tend to be underrepresented in early stages, because it is technically difficult to perform PCR on regions that have high G+C content. In such a case, plasmid DNAs are prepared and used as templates for the sequencing reaction. Later the pseudogene was mapped on human chromosome 5. It was inserted into an intron of the *DMGDH* gene. The pseudogene does not have any introns but an additional 96-bp fragment between exons 4 and 5. A sequence homologous to the inserted fragment was found in the 4th intron of both *Impact* and *IMPACT*. Intriguingly, in both mouse and human, the inserted sequences are flanked by AG and GT at their 5' and 3' ends, respectively, suggesting that they function as cryptic exons. However, the 96-bp sequence in *Impact* contains a conserved termination codon, and the one in human *IMPACT* is 95-bp long to result in a frame shift leading to protein truncation. The 4th introns are the longest ones in both mouse and human genes. The conserved sequences, which is unique lacking any other homologous regions in the genomes, may have some important roles. The shorter transcript detected by Northern blot hybridization (Fig. 6.1) could be related to an mRNA with the cryptic exon which might be subsequently generated through differential polyadenylation. The smaller protein coded by the mRNA would contain

most part of GI domain (Kubota *et al*. 2000), which could repress the function of IMPACT via a dominant negative effect. I did not examined the transcription in this study. However, analyses of the mRNA and the smaller protein should be required for the functional analysis of the Impact and IMPACT proteins.

The sequencing strategy

The shot-gun strategy has been most widely used for large-scale sequencing because its process in general is simple. However, it requires a high redundancy of sequencing, which often causes difficulty in assembling sequencing data. And the process using DNA sequencers had been time-consuming before capillary electrophoresis was applied to the machines. To reduce the process, I used a unique nested deletion strategy (Hattori *et al*. 1997) for the sequencing of the BAC clones. The BAC clones seem to have a large number of repetitive sequences, for example SINEs and LINEs. Similar sequences would be misassembled together in the former strategy. In the nested deletion method, however, relatively small subclones, *i.e*. up to 10 kb, are thoroughly sequenced so that the possibility of misassembling is reduced compared to the shot-gun strategy. I can offer confident sequence data by the nested deletion method. And sequenced subclones can be used for further analyses, such as for gene targeting.

Intron sizes and SINEs in an imprinted locus

Although some researchers suggested that imprinted genes have few or no introns (Hurst *et al*. 1996; Haig 1996), these cognate genes consist of at least 11 exons. Average size of these exons is about 100 bp except for the last ones. The last exons containing the ochre codons are more than 2-kb long, which are eccentric for others. In their study, they also insisted that the average intron size in imprinted genes is significantly smaller than in the control group. They listed a parameter which was given by dividing total intron size by total exon size, combining mouse and human genes. The values are 1.4 and 7.6 for imprinted genes and the control set respectively. For *Impact* and *IMPACT*, they are 5.0 and 6.3 respectively. It is true that the value is lower for imprinted mouse gene, however, I cannot conclude that the difference of intron sizes causes the different expression pattern. An *in silico* study revealed that SINEs were excluded from imprinted regions in the human genome (Greally 2002),

which may retain short introns. Indeed, much more SINEs were found in the nonimprinted human *IMPACT* locus than the imprinted *Impact* locus (Fig. 1.1). Nowadays more and more imprinted genes have been identifeid and genomic sequences are being available. It is high time that the hypothesis were reconsidered. The situation has been significantly changed since I started this study five years ago.

The genomic organization of the Impact family

Recent advance of genome projects has revealed that hypothetical proteins homologous to *Impact* throughout the reading frame are found in the genomes of the nematode *C. elegans*, the fission yeast *S. pombe*, *etc*. These exons are predicted by computer software, Genefinder (Favello *et al*. 1995). While experimental data were not available, their deduced amino acid sequences were aligned with *Impact* and *IMPACT* to compare the splice junctions. All the splice junctions are well conserved between mouse and human genes. But the sixth junctions are slightly different (Table 1). This may have something to do with one amino acid deletion in the exon 6 of mouse *Impact*. The Impact family is known to consist of three domains (Yamada *et al*. 1999). But the junctions are scattered relatively equally on the whole amino acid sequences regardless of the domains. Comparing mammals with the nematode, one junction is located in the same position, but others are not conserved. As compared to the fission yeast whose ORF is predicted to be split into three exons *in silico*, no conservation was observed. Because these results make me suspicious whether they are orthologous, empirical evidences are desired for functional analysis of the gene family with totally unknown function.

Homologous sequences found by comparative analysis

This study paid attention to differences, however, comparative sequence analysis also identifies homologous sequences of currently unknown function. Some of them are highly conserved like the intronic sequence mentioned at the beginning of this chapter. These show no homology to any repetitive sequences known in mammalian genome, therefore they should play an important role in cells. It is intriguing that the splice donors of exon 6 are conserved, where the junctions are shifted. I could also find small conserved regions upstream of human CpG island, whereas G+C content has been considerably changed between them. Annotation of these conserved sequences

provides the basis for future experimental analysis of these potential functional elements.

A LINE-1 lying around the promoter region of *Impact*

More than half of all LINE-1 are believed to be inserted into the genome before the mammalian radiation, based on their presence at orthologous sites in different mammalian genomes (Smit *et al*. 1995). However, I discovered that only mouse *Impact* has a LINE-1 fragment upstream of the first exon. It encodes a piece of reverse transcriptase but includes several termination codons. I cannot deny that the gene expression of *Impact* is affected by the parasitic sequence. Presence of a LINE-1 around the promoter region may cause the different expression manner between mouse and human. FISH analysis indicated that 365M4 seemed to be rich in repetitive sequences. But I have not found any LINE-1 fragment in the human *IMPACT* Locus. Because imprinted genes might be regulated by chromosomal domains, investigation between these parasitic elements and genomic imprinting could disclose its molecular mechanism.

CpG islands and their methylation status

Besides this LINE-1 fragment, there is another significant difference around the 5' region of *Impact* and *IMPACT*. As in most cases, human *IMPACT* has a CpG island located around 5' portion of the gene. This island encompasses exon 1 and extends into the 1st intron. On the other hand, mouse *Impact* has a CpG island within the 1st intron. A CpG island is a genomic region with a high G+C content and a high ratio of observed versus expected CpG dinucleotides (Gardiner-Garden *et al*. 1987). While the human *IMPACT* CpG island meets the criteria, the corresponding region of mouse does not (Fig. 1.1). The %GC of the mouse region is 43 and the Obs/Exp CpG is 0.35. Although 0.35 is higher than average (0.25 for the 37,954-bp region sequenced in this study), it may not work as a CpG island any longer. Even though introns of nonimprinted human *IMPACT* are longer than those of mouse *Impact* on the whole, the 1st intron of *IMPACT* is shorter than that of mouse gene. And the mouse and human introns do not share any homology at all. These facts tell that the mouse CpG island was generated or the human sequence was deleted after the mammalian radiation. This unique *Impact* CpG island has a highly repeated structure, which is one of the curious

characteristics of imprinted genes (Neumann *et al*. 1995). Repetitive sequences in genomic DNA have been suggested to induce gene silencing by pairing or heterochromatin formation. This study detected another reiterated structure in the human genome. The 4th intron of *IMPACT* has such a 4-kb region whose %GC is 42. The sequence consists of a number of *Alu* and some of which are inserted reversely. An *Alu* element is usually much longer those that often found around imprinted genes (Neumann *et al*. 1995) including mouse *Impact* (Fig. 3.1). Hence, the reiterate structure of the 4-kb region is different from the tandem repeats that are considered to be relative to genomic imprinting. Any other repeat-rich regions were not found in the mouse and human genomic sequences surveyed in this study.

In addition to the six CpG islands that I examined DNA methylation status, there is one more small CpG island in the mouse genome surveyed in this study. A ribosomal protein gene, which has a 0.23-kb CpG island, is mapped in the 11th intron of *Osbpl1b*. No such a gene is found in the human locus. Because I cannot exclude the possibility of misassebling, I have not analyzed it yet. But this difference might cause the biased expression. The mouse genomic sequence is being elucidated day by day (Mural *et al*. 2002), and the results allow us to identify novel CpG islands. CpG islands still intrigue a number of researchers (Takai and Jones 2002), especially if they are differentially methylated (Strichman-Almashanu *et al*. 2002). For allelic DNA methylation analysis, I used McrBC to support the result obtained by methylation sensitive restriction enzymes. McrBC digests DNA containing methylcytosine on one or both strands (5'...$R^mC(N_{40\text{-}2000})R^mC$...3'). It worked well at the *Impact* CpG island (Fig. 7.3), but it did not for the assay in some cases (Fig. 7.6). I need an additional experiment, *e.g*. bisulfite based analysis, to evaluate their methylation status, but they seemed to be usual CpG islands. As polymorphisms are often found in repeated structured regions, I found many polymorphisms in the *Impact* CpG island between B6 and JF. The length polymorphism (181 bp) easily led me to the clear demonstration of the differential methylation (Fig. 4.1). This finding introduces a very convenient technique for researchers. Some mechanisms have been proposed for gene silencing by methylation, but in this case allele-specific differential methylation was found in an intron. The mechanisms do not seem to be so simple. But I can safely say that this region plays an important role in the imprinting of *Impact*. A model has been proposed to explain how the allele-specific methylation patterns are established at

imprinted loci (Constancia *et al*. 1998). According to the model, allele-specific methylation patterns are the result of the interaction of opposing methylation and demethylation signals. Tandem repeats may act as methylation centers, attracting or inducing the spread of methylation over a stretch of DNA. The extent of methylation may be limited by counteracting demethylation signals induced by CpG-rich environment. In the case of *Impact*, for instance, a male germ-line-specific *trans*-acting factor could demonstrate the parent-of-origin-specific methylation status. It is likely that such a factor interfere with the methylation signal in spermatogenesis and the methylation status is maintained throughout the subsequent fertilization and development. To verify the model and this hypothesis, I need to get much more data about regional and temporal methylation status. This methylation-specific PCR assay will help understand more about the imprinting of *Impact*. This can be applied to various examinations, for example changes of methylation levels during development (Fig. 4.2). It has been shown that DNA methylation status changes during aging (Issa *et al*. 1994) and the change can induce tumorigenesis (Jones and Laird 1999). Such islands may provide comprehensive access to these regions. Since some genes exhibit stage-specific imprinting (Latham 1999), genomic imprinting may play some important roles in aging and subsequent carcinogenesis.

I showed only the *Impact* CpG island contains characteristic tandem repeats and serves as a differentially methylated region. Intriguingly, this intronic island is missing from the nonimprinted human *IMPACT*. These result suggest that the intronic DMR plays a crucial role in the imprinting of *Impact*.

Trans-acting factors involved with genomic imprinting

Until now a number of studies on *cis*-acting centers have been done and successfully identified many elements responsible for genomic imprinting. For example, in case of the *H19* gene, a transgenic study was performed to search the minimal sequence required for the imprinting (Cranston *et al*. 2001). Although *trans*-acting factors encoded on different chromosomes are presumed to be equally important, very few factors have been discovered. CTCF, which was originally found as a DNA-binding protein to a boundary element at the chicken globin locus, is one of a few candidates of *trans*-acting factors (Bell *et al*. 1999). It was shown to regulate enhancer access to the *H19-Igf2* imprinted genes (Bell *et al*. 2000) and recently to be

implicated in X-inactivation choice (Chao *et al*. 2002). Another factor, BOLIS, which is a male germ-line-specific insulator protein, was also discovered (Loukinov *et al*. 2002). It is more and more important to identify such factors and their binding sites. In the case of the *U2af1-rs1* gene, the multicopy transgene (8.3-kb genomic fragment that contained the whole coding of *U2af1-rs1* and a repeat-rich CpG island) seemed to titrate the *trans*-acting factor which was involved in genomic imprinting, and as a result the machinery was disturbed so that imprinting of endogenous *Usaf1-rs1* was affected (Hatada *et al*. 1997). Hence, a transgenic study was performed for mouse *Impact* to define the region where a putative *trans*-acting factor binds (data not shown). If the factor interacted with the *Impact* CpG island, additional multicopy islands would absorb it and endogenous ones would be freed from the factor. As the result of this titration, the allele-specific expression of *Impact* would be expected to be disturbed. However, no aberration was observed in this study. While a large number of researchers insist upon the importance of such CpG islands, this region has been demonstrated to be dispensable by a gene targeting technique as for *U2af1-rs1* (Sunahara *et al*. 2000). This is consistent with this transgenic study, although the mechanisms between *U2af1-rs1* and *Impact* might be different.

The histamine receptor H_4 genes

Histamine is an important physiological amine that works as a chemical messenger to exert numerous functions in central and peripheral tissues. These effects are mediated through three pharmacologically distinct subtypes of receptors, *i.e*. the H_1, H_2, and H_3 receptors, which are all members of the GPCR family (Hill *et al*. 1997). With hundreds of members populating the family and many more awaiting discovery in the human genome, they are of interest to the pharmaceutical industry because of the opportunities they afford for yielding potentially lucrative drug targets. The *HRH4* gene was one of them and mapped to just downstream of *IMPACT*. Unexpectedly, I found a SNP, which was a prerequisite for this study, in the ORF which alters an amino acid. While the amino acid residue is not evolutionarily conserved among mammals, it may intrigue some investigators. I successfully identified its mouse homologue and knew that both the mouse and human genes were expressed biallelically. They have neither CpG islands nor regions whose %GC is high. According to the tissue distribution analyses, the regulation of *Impact* and *Hrh4* transcription seems to be

independent each other. It is not so much a boundary of an imprinted region as that *Impact* is an isolated imprinted gene.

The oxysterol binding protein like 1 genes

Oxysterols are oxygenated derivatives of cholesterol. Their physiological functions are regulation of cholesterol synthesis (Kandutsch *et al*. 1978), modulation of vesicular movement (Fang *et al*. 1996), induction of differentiation (Hanley *et al*. 2000), and involvement with cell cycle regulation and apoptosis (Schroepfer 2000). Presumably, these functions are mediated by proteins that bind the lipid. Oxysterol binding proteins are known to be part of multi-gene families in eukaryotes including yeasts, nematodes, fruit fries, and mammals. As the human *OSBP* genes show many examples of alternative splicing (Jaworski *et al*. 2001), the *Osbpl1* genes that lie just upstream of *Impact* use alternative promoters. One gene product is twice as large and features N-terminal PH domain. Similar diversity in the structures of OSBPs is found in the genomes of many eukaryotes, which have forms with and without PH domains. Intriguingly, the two initiation sites make the two different forms for this locus. This was first reported in human genes (Jaworski *et al*. 2001), and I showed the mechanism was conserved in mice.

It is of particular interest that the *Osbpl1* genes and the imprinted *Impact* gene reside in head-to-head orientation. *Osbpl1a* is expressed equally from the maternal and paternal alleles, however, *Osbpl1b* is preferentially expressed from the paternal one. That three transcripts have their own CpG islands suggests the transcription of each gene is regulated independently to one another. But Northern blot analysis showed similar tissue distribution for *Osbpl1a* and *Osbpl1b* except for that in spleen. The transcription might be coordinated. The *Osbpl1a* promoter region is far from *Impact*, but that of *Osbpl1b* is located much closer to the *Impact* one. It is conceivable that mouse *Impact* is an isolated imprinted gene and its allele-specific difference of chromatin structure has an influence on expression of a nearby gene, *Osbpl1b*. It is likely that the allele-specific expression of *Impact* involves *Osbpl1b* in its skewed transcription.

Some theories, *e.g*. the genetic conflict theory (Constancia *et al*. 2002), have been presented for biological meanings of genomic imprinting. Now that many imprinted genes which have various functions have been identified, it is becoming difficult to get

the manifold functions into one theory. It seems that several adjacent genes get involved with parent-of-origin-specific expression manner regardless of the physiological significance, because such genes usually flank each other on a chromosomal region. In any case, the allelic expression manner is different between the two transcripts of *Osbpl1*. The structural difference is whether the protein has a PH domain or not. It is a kind of functional domain containing about 120 amino acids, which exists on many proteins that are involved in cellular signal transduction. And the OSBP PH domains have been shown to bind to the Golgi apparatus in yeast and mammalian cells. In spleen, the *Osbpl1b* mRNA that encodes PH domain is exclusively expressed. If there were a meaning of the biased expression of the gene, the functional difference would take on importance for the study of genomic imprinting.

Species-specific imprinting of *Impact*

I have difficulty in enumerating the actual number of imprinted genes because of the ambiguous definition of a gene. According to a catalogue of imprinted genes (Morison *et al*. 2001), 41 genes were reported to be preferentially or exclusively expressed from one of the two parental alleles in mouse and human. For most of them, parent-of-origin-specific expression is evolutionarily conserved between the two organisms. However, some human homologues, *e.g*. those of imprinted mouse *U2af1-rs1* (Pearsall *et al*. 1996), *Igf2r* (Kalscheuer *et al*. 1993), and *Tssc4* (Paulsen *et al*. 2000), have been documented as nonimprinted genes. The imprinting of human *IMPACT* is also leaky, polymorphic, or lost, therefore mouse gene exhibits species-specific imprinting (Okamura *et al*. 2002). As I mention in the next section, the human *U2AF1-RS1* is thought to be the homolog of nonimprinted *U2af1-rs2* on the X chromosome rather than *U2af1-rs1* (Nabetani *et al*. 1997). As for *IGF2R*, a differentially methylated region that is not accompanied by allele-specific transcription was reported (Riesewijk *et al*. 1996). On the other hand, no parent-of-origin effect has been observed for *IMPACT* yet, hence mouse *Impact* appears to be an exceptional case. To understand the mechanisms underlying genomic imprinting, several research groups have focused on the identification and characterization of genomic regions that are conserved between mouse and human (Vu and Hoffman 2000). This comparative study, however, sheds light on the mechanisms by the difference between the two species. The imprinted *Impact* and the nonimprinted *IMPACT* genes will serve as a

nice model to solve the mystery.

A Unique type of genomic imprinting

Apart from the biased expression of *Osbpl1b*, mouse *Impact* appears to be a solitary imprinted gene, whereas clustering into chromosomal domains is one of the characteristic features of imprinted genes. Only *U2af1-rs1* (Nabetani *et al*. 1997) and *NNAT* (Evans *et al*. 2001) have been recognized as such genes so far. The mouse *U2af1-rs1* gene is located within an intron of the *Murr1* gene that is transcribed biallelically. The isolated imprinted gene is thought to originate in a retrotransposition of *U2af1-rs2* on another chromosome into the intron of *Murr1* due to the following reasons. First, *U2af1-rs1* is intron less, which is one of the characteristic features of imprinted genes, while *U2af1-rs2* has introns. Second, the human *U2AF1-RS1* does not sit within the human *MURR1* locus on human chromosome 2, which has conserved synteny to the mouse *Murr1* locus. And *U2AF1-RS1* is not so much the human homologue of *U2af1-rs1* as that of *U2af1-rs2*, judging from sequence similarity. The transposition might have occurred after mice and humans diverged. This view could be concordant with the absence of imprinting of *U2AF1-RS1* (Pearsall *et al*. 1996). Third, because it has been shown that some transgenes exhibit parent-of-origin-specific expression and methylation (Reik *et al*. 1987), the neomorphic *U2af1-rs1* gene can be thought to bahave like an imprinted transgene. The human *NNAT* gene that is transcribed specifically from the paternal allele lies within the singular intron of *BLCAP* which is not imprinted (Evans *et al*. 2001). In contrast to *U2af1-rs1*, the genomic structure and imprint status seem to be conserved between mouse and human, and both mouse *Nnat* and human *NNAT* have two introns. Although *BLCAP* homologues are present in species such as *D. rerio* and *D. melanogaster*, *NNAT* homologues are not found outside the mammalian lineage. These indicate that the gene structure could originate through a retrotransposition event early in mammalian evolution. Surprisingly, tandem repeat sequences have not been detected within this domain.

I have demonstrated here that mouse *Impact* is also a solitary imprinted gene as above. However, there are two remarkable differences in their genomic structures. One is the numbers of introns, and the other is whether one lies within an intron of another gene. The *Impact* gene has ten introns unlike typical imprinted genes. And the sizes (Table 1) are not smaller than those of genes whose origins were thought as

transposition. In contrast to *U2af1-rs1* and *NNAT*, *Impact* does not reside in introns of other genes but has nonimprinted genes close on the chromosomal domain. It seems far from a retrotranspositional gene, while such an insertion is thought as a common mechanism for the formation of imprinted genes (Nabetani *et al*. 1997). These findings indicate the mouse *Impact* has a unique mechanism for genomic imprinting compared to not only the two retrotranspositional genes but also most imprinted genes clustered. Fortunately, the human orthologue *IMPACT* does not exhibit allele-specific expression. Hence, differences restricted in this locus, *e.g*. the intronic CpG island, would represent the modification necessary for *Impact* to be imprinted. The neomorphic island itself could be derived from a transposition after the mammalian radiation.

This study presented the first analysis of this unique type of genomic imprinting elucidating the genomic structures comparing to its nonimprinted human homologue. The comparative analysis gave an intronic sequence as the *cis*-acting region responsible for the imprinting, however, there might be other regions playing roles in it. I could also show the *Impact* gene is suited for investigating the control elements required for localized regulation of genomic imprinting.

References

Alt, F.W., Yancopoulos, G.D., Blackwell, T.K., Wood, C., Thomas, E., Boss, M., Coffman, R., Rosenberg, N., Tonegawa, S., and Baltimore, D. 1984. Ordered rearrangement of immunoglobulin heavy chain variable region segments. *EMBO J*. 3(6):1209-19. [PMID:6086308]

Arrang, J.M., Garbarg, M., and Schwartz, J.C. 1983. Auto-inhibition of brain histamine release mediated by a novel class (H_3) of histamine receptor. *Nature* 302(5911):832-7. [PMID:6188956]

Barlow, D.P., Stoger, R., Herrmann, B.G., Saito, K., and Schweifer, N. 1991. The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the *Tme* locus. *Nature* 349(6304):84-7. [PMID:1845916]

Barlow, D.P. 1997. Competition - a common motif for the imprinting mechanism? *EMBO J*. 16(23):6899-905. [PMID:9384569]

Baroux, C., Spillane, C., and Grossniklaus, U. 2002. Genomic imprinting during seed development. *Adv. Genet*. 46:165-214. [PMID:11931224]

Bell, A.C., and Felsenfeld, G. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* 405(6785):482-5. [PMID:10839546]

Bell, A.C., West, A.G., and Felsenfeld, G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98(3):387-96. [PMID:10458613]

Boumil, R.M., and Lee, J.T. 2001. Forty years of decoding the silence in X-chromosome inactivation. *Hum. Mol. Genet*. 10(20):2225-32. [PMID:11673405]

Breathnach, R., and Chambon, P. 1981. Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem*. 50:349-83. [PMID:6791577]

Chao, W., Huynh, K.D., Spencer, R.J., Davidow, L.S., and Lee, J.T. 2002. CTCF, a candidate *trans*-acting factor for X-inactivation choice. *Science* 295(5553):345-7. [PMID:11743158]

Chess, A., Simon, I., Cedar, H., and Axel, R. 1994. Allelic inactivation regulates olfactory receptor gene expression. *Cell* 78(5):823-34. [PMID:8087849]

Clark, S.J., Harrison, J., Paul, C.L., and Frommer, M. 1994. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res*. 22(15):2990-7. [PMID:8065911]

Constancia, M., Pickard, B., Kelsey, G., and Reik, W. 1998. Imprinting mechanisms. *Genome Res*. 8(9):881-900. [PMID:9750189]

Constancia, M., Hemberger, M., Hughes, J., Dean, W., Ferguson-Smith, A., Fundele, R., Stewart, F., Kelsey, G., Fowden, A., Sibley, C., and Reik, W. 2002. Placental-specific IGF-II is a major modulator of placental and fetal growth. *Nature* 2002 417(6892):945-8. [PMID:12087403]

Cranston, M.J., Spinka, T.L., Elson, D.A., and Bartolomei, M.S. 2001. Elucidation of the minimal sequence required to imprint *H19* transgenes. *Genomics* 73(1):98-107. [PMID:11352570]

Crouse, H.V., 1960. The controlling elements in sex chromosome behavior in Sciara. *Genetics* 45:1425-43. [No PMID]

Deal, C.L. 1995. Parental genomic imprinting. *Curr. Opin. Pediatr.* 7(4):445-58. [PMID:7581651]

Evans, H.K., Wylie, A.A., Murphy, S.K., and Jirtle, R.L. 2001. The *neuronatin* gene resides in a "micro-imprinted" domain on human chromosome 20q11.2. *Genomics* 77(1-2):99-104. [PMID:11543638]

Fang, M., Kearns, B.G., Gedvilaite, A., Kagiwada, S., Kearns, M., Fung, M.K., and Bankaitis, V.A. 1996. Kes1p shares homology with human oxysterol binding protein and participates in a novel regulatory pathway for yeast Golgi-derived transport vesicle biogenesis. *EMBO J*. 15(23):6447-59. [PMID:8978672]

Favello, A., Hillier, L., and Wilson, R.K. 1995. Genomic DNA sequencing methods. *Methods Cell Biol*. 48:551-69. [PMID:8531742]

Feil, R., and Khosla, S. 1999. Genomic imprinting in mammals: an interplay between chromatin and DNA methylation? *Trends Genet*. 15(11):431-5. [PMID:10529801]

Gardiner-Garden, M., and Frommer, M. 1987. CpG islands in vertebrate genomes. *J. Mol. Biol*. 196(2):261-82. [PMID:3656447]

Greally, J.M. 2002. Short interspersed transposable elements (SINEs) are excluded from imprinted regions in the human genome. *Proc. Natl. Acad. Sci. USA* 99(1):327-32. [PMID:11756672]

Gregory, S.G., *et al*. 2002. A physical map of the mouse genome. *Nature* 418(6899):743-50. [PMID:12181558]

Hagiwara, Y., Hirai, M., Nishiyama, K., Kanazawa, I., Ueda, T., Sakaki, Y., and Ito, T. 1997. Screening for imprinted genes by allelic message display: identification of a paternally expressed

gene *Impact* on mouse chromosome 18. *Proc. Natl. Acad. Sci. USA* 94(17):9249-54. [PMID:9256468]

Haig, D. 1996. Do imprinted genes have few and small introns? *Bioessays* 18(5):351-3. [PMID:8639158]

Hanley, K., Ng, D.C., He, S.S., Lau, P., Min, K., Elias, P.M., Bikle, D.D., Mangelsdorf, D.J., Williams, M.L., and Feingold, K.R. 2000. Oxysterols induce differentiation in human keratinocytes and increase Ap-1-dependent involucrin transcription. *J. Invest. Dermatol.* 114(3):545-53. [PMID:10692116]

Harr, R., Hagblom, P., and Gustafsson, P. 1982. Two-dimensional graphic analysis of DNA sequence homologies. *Nucleic Acids Res.* 10(1):365-74. [PMID:7063405]

Hatada, I., Nabetani, A., Arai, Y., Ohishi, S., Suzuki, M., Miyabara, S., Nishimune, Y., and Mukai, T. 1997. Aberrant methylation of an imprinted gene *U2af1-rs1*(*SP2*) caused by its own transgene. *J. Biol. Chem.* 272(14):9120-2. [PMID:9083040]

Hattori, M., Tsukahara, F., Furuhata, Y., Tanahashi, H., Hirose, M., Saito, M., Tsukuni, S., and Sakaki, Y. 1997. A novel method for making nested deletions and its application for sequencing of a 300 kb region of human APP locus. *Nucleic Acids Res.* 25(9):1802-8. [PMID:9108164]

Hill, S.J., Ganellin, C.R., Timmerman, H., Schwartz, J.C., Shankley, N.P., Young, J.M., Schunack, W., Levi, R., and Haas, H.L. 1997. International Union of Pharmacology. XIII. Classification of histamine receptors. *Pharmacol. Rev.* 1997 Sep;49(3):253-78. [PMID:9311023]

Hollander, G.A., Zuklys, S., Morel, C., Mizoguchi, E., Mobisson, K., Simpson, S., Terhorst, C., Wishart, W., Golan, D.E., Bhan, A.K., and Burakoff, S.J. 1998. Monoallelic expression of the *interleukin-2* locus. *Science* 279(5359):2118-21. [PMID:9516115]

Hurst, L.D., McVean, G., and Moore, T. 1996. Imprinted genes have few and small introns. *Nat. Genet.* 12(3):234-7. [PMID:8589711]

Inoue, K., Kohda, T., Lee, J., Ogonuki, N., Mochida, K., Noguchi, Y., and Tanemura, K., Kaneko-Ishino, T., Ishino, F., and Ogura, A. 2002. Faithful expression of imprinted genes in cloned mice. *Science* 295(5553):297. [PMID:11786635]

Issa, J.P., Ottaviano, Y.L., Celano, P., Hamilton, S.R., Davidson, N.E., and Baylin, S.B. 1994. Methylation of the oestrogen receptor CpG island links ageing and neoplasia in human colon. *Nat. Genet.* 7(4):536-40. [PMID:7951326]

Jaworski, C.J., Moreira, E., Li, A., Lee, R., and Rodriguez, I.R. 2001. A family of 12 human genes containing oxysterol-binding domains. *Genomics* 78(3):185-96. [PMID:11735225]

Jones, P.A., and Laird, P.W. 1999. Cancer epigenetics comes of age. *Nat Genet*. 21(2):163-7. [PMID:9988266]

Kalscheuer, V.M., Mariman, E.C., Schepens, M.T., Rehder, H., and Ropers, H.H. 1993. The insulin-like growth factor type-2 receptor gene is imprinted in the mouse but not in humans. *Nat. Genet*. 5(1):74-8. [PMID:8220428]

Kandutsch, A.A., Chen, H.W., and Heiniger, H.J. 1978. Biological activity of some oxygenated sterols. *Science* 201(4355):498-501. [PMID:663671]

Kim, U.J., Birren, B.W., Slepak T., Mancino, V., Boysen, C., Kang, H.L., Simon, M.I., and Shizuya H. 1996. Construction and characterization of a human bacterial artificial chromosome library. *Genomics* 34(2):213-8. [PMID:8661051]

Kosaki, K., Suzuki, T., Kosaki, R., Yoshihashi, H., Itoh, M., Goto, Y., and Matsuo, N. 2001. Human homolog of the mouse imprinted gene *Impact* resides at the pericentric region of chromosome 18 within the critical region for bipolar affective disorder. *Mol. Psychiatry* 6(1):87-91. [PMID:11244491]

Kubota, H., Sakaki, Y., and Ito T. 2000. GI domain-mediated association of the eukaryotic initiation factor 2α kinase GCN2 with its activator GCN1 is required for general amino acid control in budding yeast. *J. Biol. Chem*. 275(27):20243-6. [PMID:10801780]

Lander, E.S., Rogers, J., Waterston, R.H., Hawkins, T., Gibbs, R.A., Sakaki, Y., *et al*. 2001. Initial sequencing and analysis of the human genome. *Nature* 409(6822):860-921. [PMID:11237011]

Latham, K.E. 1999. Epigenetic modification and imprinting of the mammalian genome during development. *Curr. Top. Dev. Biol*. 43:1-49. [PMID:9891882]

Levanon, D., Hsieh, C.L., Francke, U., Dawson, P.A., Ridgway, N.D., Brown, M.S., and Goldstein, J.L. 1990. cDNA cloning of human oxysterol-binding protein and localization of the gene to human chromosome 11 and mouse chromosome 19. *Genomics* 7(1):65-74. [PMID:1970801]

Liu, C., Wilson, S.J., Kuei, C., and Lovenberg, T.W. 2001. Comparison of human, mouse, rat, and guinea pig histamine H_4 receptors reveals substantial pharmacological species variation. *J. Pharmacol. Exp. Ther.* 299(1):121-30. [PMID:11561071]

Lloyd, V. 2000. Parental imprinting in *Drosophila*. *Genetica* 109(1-2):35-44. [PMID:11293793]

Loukinov, D.I., Pugacheva, E., Vatolin, S., Pack, S.D., Moon, H., Chernukhin, I., Mannan, P., Larsson, E., Kanduri, C., Vostrov, A.A., Cui, H., Niemitz, E.L., Rasko, J.E., Docquier, F.M., Kistler, M., Breen, J.J., Zhuang, Z., Quitschke, W.W., Renkawitz, R., Klenova, E.M., Feinberg, A.P., Ohlsson, R., Morse, H.C., and Lobanenkov, V.V. 2002. BORIS, a novel male germ-line-specific protein associated with epigenetic reprogramming events, shares the same 11-zinc-finger domain with CTCF, the insulator protein involved in reading imprinting marks in the soma. *Proc. Natl. Acad. Sci. USA* 99(10):6806-11. [PMID:12011441]

Lovenberg, T.W., Roland, B.L., Wilson, S.J., Jiang, X., Pyati, J., Huvar, A., Jackson, M.R., and Erlander, M.G. 1999. Cloning and functional expression of the human histamine H_3 receptor. *Mol. Pharmacol.* 55(6):1101-7. [PMID:10347254]

Lyon, M.F. 1961. Gene action in the X-chromosome of the mouse (*Mus musculus L.*). *Nature* 190:372-3. [No PMID]

Matsuoka, S., Thompson, J.S., Edwards, M.C., Bartletta, J.M., Grundy, P., Kalikin, L.M., Harper, J.W., Elledge, S.J., and Feinberg, A.P. 1996. Imprinting of the gene encoding a human cyclin-dependent kinase inhibitor, $p57^{KIP2}$, on chromosome 11p15. *Proc. Natl. Acad. Sci. USA* 93(7):3026-30. [PMID:8610162]

McMahon, F.J., Hopkins, P.J., Xu, J., McInnis, M.G., Shaw, S., Cardon, L., Simpson, S.G., MacKinnon, D.F., Stine, O.C., Sherrington, R., Meyers, D.A., and DePaulo, J.R. 1997. Linkage of bipolar affective disorder to chromosome 18 markers in a new pedigree series. *Am. J. Hum. Genet.* 61(6):1397-404. [PMID:9399888]

McMahon, F.J., Stine, O.C., Meyers, D.A., Simpson, S.G., and DePaulo, J.R. 1995. Patterns of maternal transmission in bipolar affective disorder. *Am. J. Hum. Genet.* 56(6):1277-86. [PMID:7762550]

Mitaku, S., and Hirokawa, T. 1999. Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length. *Protein Eng.* 12(11):953-7. [PMID:10585500]

Moore, T. and Haig, D. 1991. Genomic imprinting in mammalian development: a parental tug-of-war. *Trends Genet.* 7(2):45-9. [PMID:2035190]

Morison, I.M., Paton, C.J., and Cleverley, S.D. 2001. The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res.* 29(1):275-6. [PMID:11125110]

Mural, R.J., Adams, M.D., Myers, E.W., Venter, J.C., Pfannkoch, C.M., Barnstead, M., Stephenson,

L.D. *et al*. 2002. A comparison of whole-genome shotgun-derived mouse chromosome 16 and the human genome. *Science* 296(5573):1661-71. [PMID:12040188]

Nabetani, A., Hatada, I., Morisaki, H., Oshimura, M., and Mukai, T. 1997. Mouse *U2af1-rs1* is a neomorphic imprinted gene. *Mol. Cell. Biol.* 17(2):789-98. [PMID:9001233]

Nakamura, T., Itadani, H., Hidaka, Y., Ohta, M., and Tanaka, K. 2000. Molecular cloning and characterization of a new human histamine receptor, HH4R. *Biochem. Biophys. Res. Commun.* 279(2):615-20. [PMID:11118334]

Neumann, B., Kubicka, P., and Barlow D.P. 1995. Characteristics of imprinted genes. *Nat. Genet.* 9(1):12-3. [PMID:7704015]

Oakey, R.J., Matteson, P.G., Litwin, S., Tilghman, S.M., and Nussbaum, R.L. 1995. Nondisjunction rates and abnormal embryonic development in a mouse cross between heterozygotes carrying a (7, 18) Robertsonian translocation chromosome. *Genetics* 141(2):667-74. [PMID:8647401]

O'Brien, S.J., Womack, J.E., Lyons, L.A., Moore, K.J., Jenkins, N.A., and Copeland, N.G. 1993. Anchored reference loci for comparative genome mapping in mammals. *Nat. Genet.* 3(2):103-12. [PMID:8499943]

Oda, T., Morikawa, N., Saito, Y., Masuho, Y., and Matsumoto, S. 2000. Molecular cloning and characterization of a novel type of histamine receptor preferentially expressed in leukocytes. *J. Biol. Chem.* 275(47):36781-6. [PMID:10973974]

Okamura, K., Hagiwara-Takeuchi, Y., Li, T., Vu, T.H., Hirai, M., Hattori, M., Sakaki, Y., Hoffman, A.R., and Ito, T. 2000. Comparative genome analysis of the mouse imprinted gene *Impact* and its nonimprinted human homolog *IMPACT*: toward the structural basis for species-specific imprinting. *Genome Res.* 10(12):1878-89. [PMID:11116084]

Paulsen, M., El-Maarri, O., Engemann, S., Strodicke, M., Franck, O., Davies, K., Reinhardt, R., Reik, W., and Walter, J. 2000. Sequence conservation and variability of imprinting in the Beckwith-Wiedemann syndrome gene cluster in human and mouse. *Hum. Mol. Genet.* 9(12):1829-41. [PMID:10915772]

Pearsall, R.S., Shibata, H., Brozowska, A., Yoshino, K., Okuda, K., deJong, P.J., Plass, C., Chapman, V.M., Hayashizaki, Y., Held, W.A. 1996. Absence of imprinting in U2AFBPL, a human homologue of the imprinted mouse gene U2afbp-rs. *Biochem. Biophys. Res. Commun.* 222(1):171-7. [PMID:8630064]

Reik, W., Collick, A., Norris, M.L., Barton, S.C., and Surani, M.A. 1987. Genomic imprinting

determines methylation of parental alleles in transgenic mice. *Nature* 328(6127):248-51. [PMID:3600805]

Reik, W., Maher, E.R. 1997. Imprinting in clusters: lessons from Beckwith-Wiedemann syndrome. *Trends Genet*. 13(8):330-4. [PMID:9260520]

Rideout, W.M., Eggan, K., and Jaenisch, R. Nuclear cloning and epigenetic reprogramming of the genome. *Science* 293(5532):1093-8. [PMID:11498580]

Riesewijk, A.M., Schepens, M.T., Welch, T.R., van den Berg-Loonen, E.M., Mariman, E.M., Ropers, H.H., and Kalscheuer, V.M. 1996. Maternal-specific methylation of the human *IGF2R* gene is not accompanied by allele-specific transcription. *Genomics* 31(2):158-66. [PMID:8824797]

Schroepfer, G.J., Jr. 2000. Oxysterols: modulators of cholesterol metabolism and other processes. *Physiol. Rev.* 80(1):361-554. [PMID:10617772]

Shizuya, H., Birren, B., Kim, U.J., Mancino, V., Slepak, T., Tachiiri, Y., and Simon, M. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Proc. Natl. Acad. Sci. USA* 89(18):8794-7. [PMID:1528894]

Siebert, P.D., Chenchik, A., Kellogg, D.E., Lukyanov, K.A., and Lukyanov, S.A. 1995. An improved PCR method for walking in uncloned genomic DNA. *Nucleic Acids Res*. 23(6):1087-8. [PMID:7731798]

Smit, A.F., Toth, G., Riggs, A.D., and Jurka, J. 1995. Ancestral, mammalian-wide subfamilies of LINE-1 repetitive sequences. *J. Mol. Biol*. 246(3):401-17. [PMID:7877164]

Solter, D. 1988. Differential imprinting and expression of maternal and paternal genomes. *Annu. Rev. Genet*. 22:127-46. [PMID:3071246]

Strichman-Almashanu, L.Z., Lee, R.S., Onyango, P.O., Perlman, E., Flam, F., Frieman, M.B., and Feinberg, A.P. 2002. A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome Res*. 12(4):543-54. [PMID:11932239]

Sunahara, S., Nakamura, K., Nakao, K., Gondo, Y., Nagata, Y., and Katsuki, M. 2000. The oocyte-specific methylated region of the *U2afbp-rs/U2af1-rs1* gene is dispensable for its imprinted methylation. *Biochem. Biophys. Res. Commun*. 268(2):590-5. [PMID:10679248]

Sutherland, E., Coe, L., and Raleigh, E.A. 1992. McrBC: a multisubunit GTP-dependent restriction endonuclease. *J. Mol. Biol*. 225(2):327-48. [PMID:1317461]

Takai, D., and Jones, P.A. 2002. Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. USA* 99(6):3740-5. [PMID:11891299]

Thompson, J.D., Higgins, D.G., and Gibson, T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22(22):4673-80. [PMID:7984417]

Tilghman, S.M. 1999. The sins of the fathers and mothers: genomic imprinting in mammalian development. *Cell* 96(2):185-93. [PMID:9988214]

Venter, J.C., *et al.* 2001. The sequence of the human genome. *Science* 291(5507):1304-51. [PMID:11181995]

Vu, T.H., and Hoffman, A.R. 2000. Comparative genomics sheds light on mechanisms of genomic imprinting. *Genome Res.* 10(11):1660-3. [PMID:11076850]

Wakayama, T., Perry, A.C., Zuccotti, M., Johnson, K.R., and Yanagimachi, R. 1998. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* 394(6691):369-74. [PMID:9690471]

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. 1997. Viable offspring derived from fetal and adult mammalian cells. *Nature* 385(6619):810-3. [PMID:9039911]

Xu, Y., Liu, Y., Ridgway, N.D., and McMaster, C.R. 2001. Novel members of the human oxysterol-binding protein family bind phospholipids and regulate vesicle transport. *J. Biol. Chem.* 276(21):18407-14. [PMID:11279184]

Yamada, Y., Hagiwara, Y., Shiokawa, K., Sakaki, Y., and Ito, T. 1999. Spatiotemporal, allelic, and enforced expression of *Ximpact*, the *Xenopus* homolog of mouse imprinted gene *impact*. *Biochem. Biophys. Res. Commun.* 256(1):162-9. [PMID:10066441]

Yanagawa, Y., Kobayashi, T., Ohnishi, M., Kobayashi, T., Tamura, S., Tsuzuki, T., Sanbo, M., Yagi, T., Tashiro, F., and Miyazaki, J. 1999. Enrichment and efficient screening of ES cells containing a targeted mutation: the use of *DT-A* gene with the polyadenylation signal as a negative selection maker. *Transgenic Res.* 8(3):215-21. [PMID:10478491]

Yeh, R.F., Lim, L.P., and Burge, C.B. 2001. Computational inference of homologous gene structures in the human genome. *Genome Res.* 11(5):803-16. [PMID:11337476]

Acknowledgments

I'm grateful to all the people who made this work possible. Foremost, I'd like to express my sincere gratitude to my supervisor, Prof. Yoshiyuki Sakaki. When I first visited the laboratory five years ago, he welcomed me with hospitality and set aside my poor knowledge of life science. Since then he's supported all of my work. Also, I would like to thank Prof. Takashi Ito (Kanazawa Univ.) whom I am grateful for giving me a very interesting theme and appropriate suggestions. He was instrumental especially in my learning how my own research appealed to others effectively. I am grateful that they gave me opportunities to give presentations here and there. And I appreciate Dr. Yuriko Hagiwara who hunted *Impact* by AMD. She kindly taught me the ABC's of molecular biology.

I'd like to thank Dr. Hajime Tei, who has sometimes been easy to talk with, for helpful suggestions. And I thank these members of the laboratory, Drs. Tadayuki Takeda, Yuepeng Wang, Rika Nakamura, Shinji Sato, Dong-Kug Choi, Ayako Yamamoto, Yutaka Takaoka, Akiko Hida, Rika Numano, Kazuyuki Mizushima, Yoichi Yamada, Nobuya Koike, Hideki Noguchi, Shuji Shigenobu, Messrs. Hack-Young Maeng, Chihiro Uematsu, Kunihiko Takamatsu, Kohei Maekawa, Sven Dillenburger, Atsuki Shinozaki, Soshi Kawaguchi, Tomomi Aida, Misses. Shihoko Kojima, Yoko Sato, Satomi Shiozuka, Tomoyo Shirakawa, and others.

I should like to thank Prof. Masahira Hattori (Kitasato Univ.) for cooperation. I am also grateful to Mses. Mai Kondo, Kayo Kobayashi, Rie Arai, Minami Horishima, Maho Tanaka, Yuka Matsumura, Emi Isozaki, Yuko Odama, Noriko Yamamoto, Hiroko Morita, and Mr. Tomoyuki Aizu (Genomic Sciences Center, RIKEN) for help in construction of the nested deletion libraries, and to Mses. Yasue Yamashita, Saori Tsuto, Fumiwo Ejima, Kyoko Fukada, Hiroko Seki, Kaoru Yakushiji, Messrs. Rintaro Fukawa, and Kenshiro Oshima (Genomic Sciences Center, RIKEN) for help in DNA sequencing. And to Drs. Atsushi Toyoda, Hong-Seog Park, and Todd Taylor (Genomic Sciences Center, RIKEN) for kind suggestions.

I would like to thank Prof. Momoki Hirai for mapping the mouse and human BAC clones by FISH.

I'm very much obliged to Prof. Masahide Asano and Dr. Chie Nakajima (Kanazawa

Univ.) for teaching me how to treat mice in the laboratory environment. I thank Genetic Stock Research Center, National Institute of Genetics for giving me the JF mice. My heart goes out to all the cute mice who participated in this study.

A special thank you to Dr. Kazuhisa Ota (Kanazawa Univ.) for teaching me how to treat budding yeasts and so on.

I thank Prof. Nobuaki Yoshida and Dr. Hirotake Ichise for suggestion for the construction of a gene targeting vector. I also thank Dr. Takeshi Yagi (Osaka Univ.) for a generous gift of pMC1DT-ApA containing the *DT-A* gene with the polyadenylation signal. I am obliged to Ms. Yuki Kataoka who helped me culture mammalian cells.

I have been supported by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists for three years. I would like to acknowledge the society and its fellowship system.

I wish to thank the members of my thesis qualifying committee for all their guidance and recommendations in this research.

I'd like to Mr. Gary Miller (from Britain), whom I haven't met face-to-face yet, for the English corrections.

Thank you to Misses Ritsuko Ozawa, Tomoko Chiba, Miyuki Shimada, and Aya Nakayama at all events.

I'd like to say thank you very much to all the people who gave me their own DNA or RNA. They were my friends, my friends' families, and even members of my own family. I'd better apologize for all the pain caused when blood was drawn. Honestly, I'm not related to Dracula. Some presented me cells from their oral cavities or a navel string. Most of them were not of use for imprinting examination, however, I learned a little more about the brotherhood and sisterhood of humanity.

Last but not least, I would especially like to thank my better half, Michiko, for supporting me throughout this entire process with her extensive knowledge of life science and for giving me a daughter, Kozue, who has provided me with vitality in my daily life.