ORIGINAL ARTICLE

Kiyonori Miura · Osamu Miyoshi · Kankatsu Yun
Johji Inazawa · Toshinobu Miyamoto · Hiroaki Hayashi
Hideaki Masuzaki · Shuichiro Yoshimura · Norio Niikawa
Yoshihiro Jinno · Tadayuki Ishimaru

Repeat-directed isolation of a novel gene preferentially expressed from the maternal allele in human placenta

Abstract Using a repetitive sequence of tandemly arrayed pentanucleotides in the human *H19* 3′-flanking region, we isolated a phage clone (λ PEN11) which localized to chromosome 11p15.5. The λ PEN11 phage encodes a 2.3-kb cDNA consisting of seven exons at least. The gene was mainly expressed in brain and pancreas (and less abundantly in testis), and demonstrated differential allele usage, with maternal expression being predominant in placenta, which indicates the gene is an atypical imprinted gene. While the pentamer repeat might contribute to this effect, it is also possible that the differential allele usage might reflect the local chromosomal structure known as the imprinting domain.

Key words Pentamer repeats · *H19* · *PEN11B* · Imprinting · Placenta

Introduction

While imprinting status is in general well conserved between two species, minor modifications or divergences are found in human and murine imprinted genes. The mouse *H19* gene is one of the best studied imprinted genes. *H19* imprinting is conserved in humans. The gene is exclusively expressed from the maternal allele and paternally methylated (Zhang and Tycko 1992; Moulton et al. 1994; Steenman et al. 1994). A unique sequence stretch of 8 or 9 base pairs (bp) repeated a number of times has been identified within the paternally methylated region (Tremlay et al. 1995). The human counterpart that seemed to meet prerequisites for the methylation imprint of *H19* had a 400-bp CpG-rich sequence reiterated several times (Jinno et al. 1996). While the apparent sequence divergence (no sequence homology and remarkable difference of repeated sequence sizes, 400 vs 9 bp) within the regions might explain the difference in *H19* expression patterns in human and mouse placenta (Jinno et al. 1995; Tremlay et al. 1995), it raises a question whether those repeats can be a target of methylation by themselves or instead whether they might need other specific sequences to be methylated.

Prompted by this question, we searched for any characteristic sequences in the human *H19* downstream region. In addition to two sequences homologous to the mouse endoderm-specific enhancers (Yoo-Warren et al. 1988), we found several kinds of repeated sequences including a CA dinucleotide repeat (unpublished data). Among them, we were especially interested in multiple tandem repeats of the pentanucleotide elements, CCC(A/T)G, located 0.7 kb downstream of the enhancer-like sequence, because it shared homology with the immunoglobulin switch regions. It has been shown that methylation and expression of the murine transgene *RSVIgmyc*, which is constructed from the rous sarcoma virus long terminal repeat, a truncated c-*myc* gene, and the α constant and switch regions of the immunoglobulin heavy chain locus, are determined in a parent-origin-dependent manner (Swain et al. 1987). A subsequent study identified the smallest region within the *RSVIgmyc* transgene that contained the repeated pentamers, C(C/T)CAG, as an essential *cis*-acting element for this parent-origin-dependent effect (Chaillet et al. 1995). Furthermore, it was recently demonstrated that the mouse 3′-flanking region including the endoderm-specific enhancers was necessary for *H19* transgene methylation imprinting (Elson and Bartolomei 1997).

K. Miura · O. Miyoshi · N. Niikawa · Y. Jinno (✉)
Department of Human Genetics, Nagasaki University School of
Medicine, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan
Tel. +81-95-849-7176; Fax +81-95-849-7121
e-mail: siriusyj@net.nagasaki-u.ac.jp

K. Miura · H. Masuzaki · S. Yoshimura · T. Ishimaru
Department of Obstetrics and Gynecology, Nagasaki University
School of Medicine, Nagasaki, Japan

K. Yun
Department of Pathology, Dunedin School of Medicine, University
of Otago, Dunedin, New Zealand

J. Inazawa
Department of Molecular Cytogenetics, Medical Research Institute,
Tokyo Medical and Dental University, Tokyo, Japan

T. Miyamoto · H. Hayashi
Department of Obstetrics and Gynecology, Asahikawa Medical
College, Asahikawa, Japan

To explain the discrepancy between imprinting conservation and sequence divergence in human and mouse, and to investigate whether the *H19* pentanucleotide repeat has a role for imprinting in humans, we isolated genomic clones with the homologous repeats. A novel gene was identified in one of them, which was localized to human chromosome 11p15.5 and showed maternal-predominant expression in placenta.

Materials and methods

Nucleic acid preparation

DNA and RNA were extracted as described previously (Jinno et al. 1994). Poly(A)$^+$ RNA was isolated using Oligotex-dT30 (Takara, Tokyo, Japan).

Genomic library screening

A human placental genomic library (EMBL3 SP6/T7) was obtained from Clontech (Palo Alto, CA, USA). A total of 6 × 10^5 phage clones were screened with a pentamer repeat probe. The final washing condition was 0.1 × standard sodium citrate (SSC)/0.1% sodium dodecylsulfate (SDS) at 55°C for 15 min. The probe was prepared by polymerase chain reaction (PCR) amplification of a plasmid DNA carrying the human *H19* enhancer region. The 100 μl reaction volume contained standard reaction buffer consisting of 1.5 mM MgCl$_2$/50 mM KCl/10 mM Tris–HCl (pH 8.3)/10% glycerol/200 mM each dNTP/0.5 mM each primer, PENA, and PENB/30 ng of the plasmid/3 U Taq polymerase. Fourteen cycles of PCR were performed at 95/57/72°C for 40/30/45 s (GeneAmp PCR System 9600; Perkin-Elmer, Norwalk, CT, USA). Primer sequences are as follows: PENA, 5′-CTGTGAGAGGAAGCCACTAAG; PENB, 5′-CTAAGGTCAAGCCTGTGCTC. The PCR product was electrophoresed on a 4% polyacrylamide gel and a 394-bp fragment was purified.

cDNA library screening and reverse transcription (RT)-PCRs for cDNA isolation

A cDNA library was constructed from a 14-week gestation fetal brain RNA using ZAP-cDNA Gigapack II Gold Cloning Kit (Stratagene, La Jolla, CA, USA). The entire original library (4 × 10^6 phages) was screened with a probe prepared from an exon-trapping clone which included exons 3–5. Twenty membranes (a total of 4 × 10^5 phages) were later re-utilized for a re-screening with the 3′ end-sequences of 3′ rapid amplification of cDNA ends (RACE) clones (3R3 and 3R4), which corresponded to a roughly 200-bp segment from nucleotide 2149 to the 3′ end and a 150-bp segment from nucleotides 1195 to 1345 of the cDNA, respectively.

For isolation of the 3′ end portion of the cDNA, 3′RACE was carried out by RT followed by first round PCR, then nested PCR, and the products were cloned into a plasmid vector. Then 10 μg of fetal brain total RNA was reverse-transcribed in a 20 μl reaction volume with 1 μl of Moloney murine leukeria virus (M-MLV) reverse transcriptase (200 U/μl; Gibco/BRL, Gaithersbury, MD, USA) and 0.1 μM primer TRVP at 37°C for 1 h. The first PCR was carried out in a 50 μl reaction volume using standard buffer provided by the manufacturer with 2 μl RT product, and primers T3EB and 3RCA, for 26 cycles at 94/54/72°C for 35/35/60 s. The nested PCR consisted of 23 cycles at 94/55/72°C for 35/30/60 s using primers T3Eb and 3RCa, and 1/50 the volume of the 20-fold diluted first PCR product as a template. The nested PCR products were separated on a 4% polyacrylamide gel, and several major distinct bands were cut and eluted from the gel. Purified products were cloned into pUC19 and sequenced.

For 5′RACE, RT was performed on 4 μg of fetal brain poly(A)$^+$ RNA with primer 5RCP. cDNA was A-tailed with 50 U terminal transferase (Boehringer Mannheim, Indianapolis, IN, USA) (37°C, 30 min) after purification through a QIAquick Spin Column (Qiagen, Chatsworth, CA, USA), and followed by five cycles of PCR with primers TRVP and 5RCP at 94/34/60°C for 90/90/120 s (DNA Thermal Cycler PJ2000; Perkin Elmer). After the primers were removed, half the first PCR product was subjected to second and third PCRs driven with primer sets T3EB/5RCB and T3Eb/5RCb, respectively, according to the similar procedures as the 3′RACE. PCR conditions were as follows: 28 cycles at 94/55/72°C for 35/35/40 s (second PCR) and 22 cycles at 94/55/72°C for 30/30/30 s (third PCR). The 5′RACE products were also cloned and sequenced.

RT-PCR connecting the 3′ and 5′RACE exons was carried out in order to confirm the continuity of them. RT products from 0.1 μg fetal brain poly(A)$^+$ RNA were subjected to 28 cycles of PCR with primers RT5A and RT3B or RT4B at 94/55/72°C for 35/30/50 s, and subsequently 4 μl of the 10-fold diluted first PCR products was amplified in 100 μl reaction volume with primers RT5a and RT3b or RT4b by 28 cycles at 94/54/72°C for 35/30/50 s. Cloning and sequencing was done as above. Primers used are as follows: TRVP, 5′-CTTGGAGATCCGTCGCAACAT-TGAGATCCGACCTGTTTTTTTTTTTTTTTTTT; T3EB, 5′-CTTGGAGATCCGTCGCAACA; T3Eb, 5′-CAACA-TTGAGATCCGACCTG; 3RCA, 5′-TGTTCCAGAAG-CCGGTCAA; 3RCa, 5′-ACCTACACGGAGGGTGGG; 5RCP, 5′-CCGTTCTCCTTCTGC; 5RCB, 5′-TGATGAC-CACGAAGATCTGC; 5RCb, 5′-GCTCCTCCTTCTCC-AGGCT; RT5A, 5′-AGGAGATGTCCAACCTGACA; RT5a, 5′-GAAGAAGTCCTGGTTTGGGAA; RT3B, 5′-CTGCTTCTCCACAGAGAGAGA; RT3b, 5′-GAAG-CAGGAGCAAGGGACTA; RT4B, 5′-CTGTTTCCA-CCTCAGCCAGA; RT4b, 5′-GGAGGCCAAGACTA-CACGGA.

Primary structure determination and characterization of cDNA

Exon–intron borders were determined from sequence comparisons between genomic clones and cloned RT products

or partial cDNAs isolated from the fetal brain cDNA library. All exon portions and the partial cDNA were sequenced several times in both directions using the Auto Read and the Cycle Sequencing kits (Pharmacia, Piscataway, NJ, USA). A full-length cDNA sequence (GenBank No. AF020089) was obtained by combining these exon sequences and the partial cDNA sequence as the 3′ end. Protein sequence similarity was analyzed with the BLAST software (BLASTP and BLASTX).

Northern analysis

A Clontech Multiple Tissue Northern containing 2 μg mRNA per lane was hybridized to the same 3′ end-sequences probe as was used for re-screening of the cDNA library. Hybridization and washing were performed according to the manufacturer's recommendations. The membranes were re-utilized in hybridization with the 5′ end probe after stripping signals.

Allelic analysis of expression

Allele-specific expression of the gene was analyzed taking advantage of the *Bst*UI–*Hha*I restriction fragment length polymorphism (RFLP) in the last exon. Radioactivity of PCR products and restriction digests was measured on respective bands electrophoresed in polyacrylamide gels using an imaging analyzer BAS1000 (Fujix, Tokyo, Japan). Initially, the logarithmic phase of each sample was ascertained in a pilot experiment in which small aliquots were removed from a tube and electrophoresed to be quantified at four consecutive points of PCR in each three-cycle interval. Quantification was carried out on PCR products and their restriction digests at three consecutive points in a (late) logarithmic phase to minimize the effect of re-association of amplified products in real experiments. Reverse transcription was performed on 6–12 μg of fetal brain total RNA or 1.3–2.8 μg of fetal liver, kidney, and placental villi poly(A)$^+$ RNA in a 20 μl reaction volume with or without 200 U M-MLV RT (Gibco/BRL) at 37°C for 60 min. Before the reaction, RNA was digested with 10 U RNase-free DNase I (Boehringer Mannheim) at 37°C for 20 min. A 2 μl RT sample of the product was subjected to first PCR with primers BSTB and ASEA in a 50 μl standard buffer using 28 cycles at 94/52/72°C for 40/30/45 s. Nested PCR was carried out in a 50 μl reaction volume in the presence of 1.5 μl [α-^{32}P]dCTP (NEN, Boston, MA, USA) using 2 μl of the 10-fold diluted first-round PCR product and primers BSTb and ASEa for 11–17 cycles (brain samples) or 19–28 cycles (other tissue samples) at 94/54/72°C for 40/30/45 s. After small aliquots were removed for measurement of the amount of the amplified products, the remaining products were extracted with phenol–chloroform, precipitated with ethanol, and restricted with 18 U *Bst*UI at 60°C for 5 h. The restriction digests were separated in a 5% polyacrylamide gel. The radioactivity of each polymorphic *Bst*UI band was counted using an imaging analyzer BAS1000 (Fujix). As comparing controls, genomic DNA (200 ng) was amplified by 20–26 cycles of PCR with a different forward primer, BSTA, and the same 3′ primer, BSTb, as used in the nested RT-PCR in a 50 μl reaction volume at 94/59/72°C for 40/35/35 s. For mixing tests, cDNAs were prepared from homozygous fetal brain RNA by RT-PCR driven with primers BSTB and ASEA, cloned to pGEM-T vector (Promega, Madison, WI, USA), and sequenced to assure the direction and both end sequences of the two clones. Plasmid DNAs were diluted to 1 fg/μl and then they were mixed at various ratios. Mixed DNA (2 fg) was subjected to PCR and restriction digestion under the same conditions as in the quantification experiments of various tissue samples. Primer sequences are: ASEA, 5′-CAAGAGGGTGGTGGAGA-CCA; ASEa, 5′-AACTGTATGGAAATGATGACGG; BSTA, 5′-GTCACCTGACCCCTCAGCAA; BSTB, 5′-TCTTTAAAGTTTACAGTAGCTCG; BSTb; 5′-CACCAGCTCTGTCCTCAGAG.

Results

Phage isolation, gene identification, and characterization

Using a 394-bp DNA encompassing the pentanucleotide repeat in the human *H19* enhancer region, 10 independent clones were isolated out of 5×10^5 plaques screened. They were further selected by two additional procedures: their chromosomal localizations were determined by fluorescence in-situ hybridization (FISH), and their gene-bearing potentials were examined by restriction with *Sac*II, *Eag*I, *Nae*I, and *Sma*I endonucleases, whose recognition sequences are often found within CpG islands. The clone with the highest number of those restriction sites was localized to chromosome 11p15.5 by FISH (data not shown). Later, the radiation hybrid mapping (Research Genetics, Huntsville, AL, USA) localized this clone 30.9 cR$_{10\,000}$ from the marker SHGC-10494 (D11S 4356). This clone (λPEN11) was chosen for further analysis.

A repetitive sequence with the pentanucleotide elements (GenBank No. AF020087) was found 3.2 kb upstream from the first exon of the gene (tentatively named as *PEN11B*) (Fig. 1). In the repetitious 375-bp stretch, a pentanucleotide element of CCCAG was most frequently reiterated (30 times) and the next abundant sequence was CTCAG repeated nine times. It seemed more degenerate when compared to the corresponding repeat in the *H19* 3′-flanking region, which mostly consisted of two major pentamer elements, CCCAG and CCCTG (reiterated 24 and 20 times, respectively, in the 280-bp sequence) (Fig. 2). The overall homology of the *PEN11B* and *H19* pentamer repeat region was 66%, which was unexpectedly low. Much higher local homology (75% over 162 bp and up to 90% over 41 bp) and/ or unusual properties of tandemly repeated sequences might have contributed to give a positive signal (Zakian 1989), although it was much less intense than the other clones.

BLAST searches identified two human, T18992 and W60403, and three mouse, W18584, W75642 and W41136,
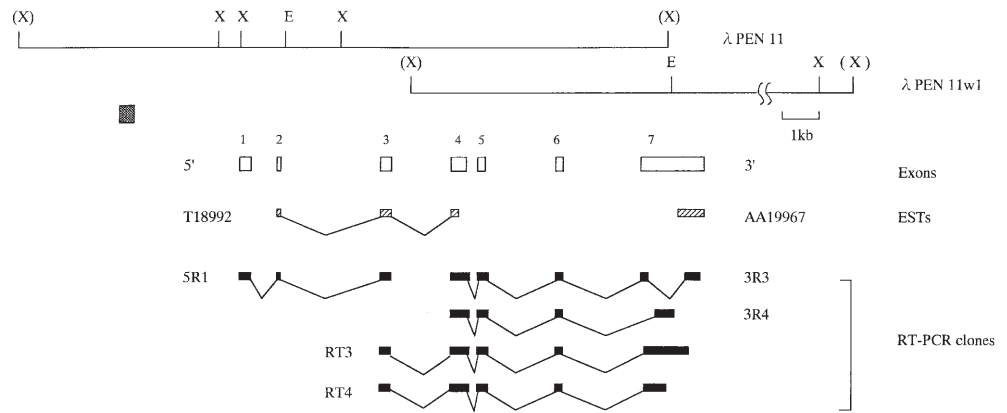
Fig. 1 Schematic representation of genomic organization of the *PEN11B* gene isolated with a pentamer repeat homologous to that in the *H19* 3′-flanking region. The phage clones are indicated with *Eco*RI (*E*) and *Xho*I (*X*) restriction sites at the top of the figure. The location of the pentamer repeat is shown by a *cross-hatched box*. The location and relative size of exons are indicated by *open boxes*. Two human expressed sequence tags corresponding to the gene are shown by *hatched boxes* below the exons. Reverse transcription-polymerase chain reaction clones are shown by *filled boxes*

PEN11B

CTCAGCCAAATCCAGTTTAGCCCAGCCCAGCCCTGCCCAGCTCAG
CCCAGCCCTGCCCAGCTCAGCCCAGTCAAATCCGTCTCAGCCCAT
TCTAGTGCAGCCTAGCCCAGCTCAGCCCAGCTCAGCCCAGCCCCG
CCCACCCTAATTCAACTCAACCCTGCTCAGCCGCGCAGTCCAGTA
CAGCCCAGCTCGGCTCAACCAAGTGCAGTCTAGCCCAGCCCAAAC
TAGCCCAGCCCAGCCCAGGCCAGCCCAGCTCAGCCTAGCCCAGTC
CAGTCCAGCACAGCCCAGCCCAGCTCAGCTCCCAGCCCAGCCCAG
CCCAGCTAAGCCCAGCCCAGCCCAGCCAAGCCCAGCCCAGCCCAG
CCCAGCTAAGCCCAG

H19

CCCAGTCCAGCCCAGCCCTGCCCTTCCCAGCCCTGCCCAGCCCAG
CTCATCCCTGCCCTACCCAGCCCAGCCCTGTCCTGCCCTGCCCAG
CCCAGCCCAGCCCAGCCCTGCCCTGCCCTGCCCTGCCCTTCCCAG
CCCTGCCCTTCCCAGCCCTGCCCAGCCCAGCTCATCCCTGCCCTA
CCCAGCTCAGCCCTGCCCTGCCCTGCCCTGCCCAGCCCTGCCCAG
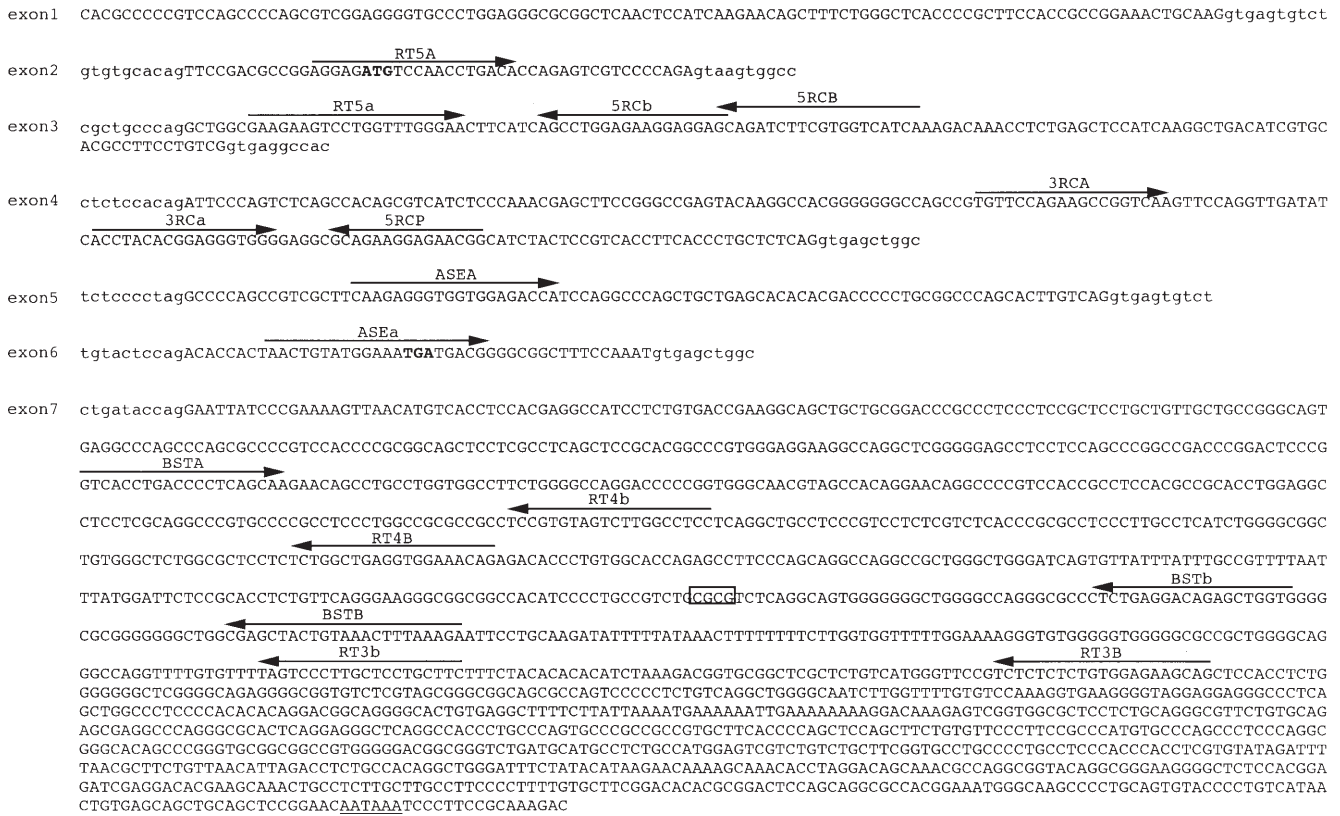CCCTGCCCTGCCCTGCCCAGCTCAGCCCTGCCCACCCCAGCCCAG
CCCAGCCCAG

Fig. 2 Nucleotide sequences of the pentamer repeats in the human *H19* and *PEN11B* genes. Most frequently repeated pentanucleotide elements are indicated with *shaded* or *open boxes*

homologous expressed sequence tags (ESTs). The mouse ESTs, W18584 and W75642, seemed to represent an identical gene. Whereas another mouse EST, W41136, revealed a high homology (97%) with the others at the 5′ portion, any homology was abruptly lost in the 3′ portion. The two human ESTs were less conserved with respect to each other (77%). Sequence comparison of the phage with these ESTs suggested that it encoded a gene from which T18992 derived. After the entire structure of the gene was revealed, BLAST identified another human EST, AA19967, which consisted of a part of the 3′ end sequence of the gene (Fig. 1). Thus, both human and mouse seem to have at least two homologous genes.

Fetal brain cDNA library screening (a total of 4×10^6 plaques) only resulted in the isolation of the homologous gene corresponding to W60403 using a 5′ portion probe prepared from an exon-trapping clone (including exons 3–5). Then, we tried to isolate the cDNA by RT-PCR. 3′RACE eventually gave two kinds of cDNA clones. One (3R4) contained a sequence of the 3′ untranslated region (UTR) co-linearly corresponding to the genomic sequence in addition to other exon sequences, but poly(A) signals were not found within the clone. Another clone (3R3) contained a 3′ end sequence with a poly(A) signal followed by a poly(A) tract which was identified in an overlapping walking clone, λPEN11w1 (Fig. 1). The longest of the 5′RACE clones had one additional exon upstream to the T18992 exons (covering exons 2, 3, and a part of 4), and we tentatively consider it the first exon of the *PEN11B* gene. Continuity of exons obtained from 3′ and 5′ RACEs was confirmed by connecting RT-PCRs using primers designed from respective RACE clone sequences (Fig. 1) and by Northern blot analyses using two different probes (described below). Furthermore, we re-attempted to isolate the gene from a fetal cDNA library with the 3′ end probes. However, a full-length cDNA was not isolated. The longest clones all ended at the *Eco*RI site corresponding to that on the overlapping phage clone, λPEN11w1. This failure may be due to a secondary structure around the *Eco*RI site, which may be reflected by the fact that we experienced some difficulties in sequencing in the vicinity of the *Eco*RI site and that yields of RT-PCR were much poorer using primers downstream of the *Eco*RI site rather than 3′ primers. Thus, the λPEN11 and the overlapping λPEN11w1 phages seem to encode a gene consisting of six short exons and the long last exon, a total of 2.3 kb in length (Fig. 1). Sequences of them are shown in Fig. 3A along with boundary intron sequences (GenBank No. AF020089). The longest open reading frame codes for 149 amino acids. Neither characteristic motifs nor known homologous proteins could be found. Amino acids sequences deduced from the cDNA and the other three ESTs were well conserved with each other, particularly in the N-terminal portions. Among 73 amino acids from the N-termini, 67 were identical and six
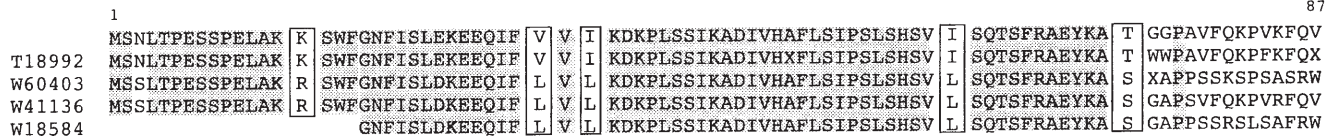
A.

```
exon1  CACGCCCCCGTCCAGCCCCAGCGTCGGAGGGGTGCCCTGGAGGGCGCGGCTCAACTCCATCAAGAACAGCTTTCTGGGCTCACCCCGCTTCCACCGCCGGAAACTGCAAGgtgagtgtct

                                            RT5A
exon2  gtgtgcacagTTCCGACGCCGGAGGAGATGTCCAACCTGACACCAGAGTCGTCCCCAGAgtaagtggcc

                              RT5a                      5RCb                      5RCB
exon3  cgctgcccagGCTGGCGAAGAAGTCCTGGTTTGGGAACTTCATCAGCCTGGAGAAGGAGGAGCAGATCTTCGTGGTCATCAAAGACAAACCTCTGAGCTCCATCAAGGCTGACATCGTGC
       ACGCCTTCCTGTCCGgtgaggccac

                                                                                              3RCA
exon4  ctctccacagATTCCCAGTCTCAGCCACAGCGTCATCTCCCAAACGAGCTTCCGGGCCGAGTACAAGGCCACGGGGGGGCCAGCCGTGTTCCAGAAGCCGGTCAAGTTCCAGGTTGATAT
          3RCa                    5RCP
       CACCTACACGGAGGGTGGGGAGGCGCAGAAGGAGAACGGCATCTACTCCGTCACCTTCACCCTGCTCTCAGgtgagctggc

                              ASEA
exon5  tctcccctagGCCCCAGCCGTCGCTTCAAGAGGGTGGTGGAGACCATCCAGGCCCAGCTGCTGAGCACACGACCCCCTGCGGCCCAGCACTTGTCAGgtgagtgtct

                  ASEa
exon6  tgtactccagACACCACTAACTGTATGGAAATGATGACGGGGCGGCTTTCCAAATgtgagctggc

exon7  ctgataccagGAATTATCCCGAAAAGTTAACATGTCACCTCCACGAGGCCATCCTCTGTGACCGAAGGCAGCTGCTGCGGACCCGCCCTCCCTCCGCTCCTGCTGTTGCTGCCGGGCAGT
       GAGGCCCAGCCCAGCGCCCCGTCCACCCCGCGGCAGCTCCTCGCCTCAGCTCCGCACGGCCCGTGGGAGGAAGGCCAGGCTCGGGGGAGCCTCCTCCAGCCCGGCCGACCCGGACTCCCG
           BSTA
       GTCACCTGACCCCTCAGCAAGAACAGCCTGCCTGGTGGCCTTCTGGGGCCAGGACCCCCGGTGGGCAACGTAGCCACAGGAACAGGCCCCGTCCACCGCCTCCACGCCGCACCTGGAGGC
                                                  RT4b
       CTCCTCGCAGGCCCGTGCCCCGCCTCCCTGGCCGCGCCGCCTCCGTGTAGTCTTGGCCTCCTCAGGCTGCCTCCCGTCCTCTCGTCTCACCCGCGCCTCCCTTGCCTCATCTGGGGCGGC
                         RT4B
       TGTGGGCTCTGGCGCTCCTCTCTGGCTGAGGTGGAAACAGAGACACCCTGTGGCACCAGAGCCTTCCCAGCAGGCCAGGCCGCTGGGCTGGGATCAGTGTTATTTATTTGCCGTTTTAAT
                                                                                                       BSTb
       TTATGGATTCTCCGCACCTCTGTTCAGGGAAGGGCGGCGGCCACATCCCCTGCCGTCTcGCGCTCTCAGGCAGTGGGGGGGCTGGGGCCAGGGCGCCCTCTGAGGACAGAGCTGGTGGGG
                         BSTB
       CGCGGGGGGGCTGGCGAGCTACTGTAAACTTTAAAGAATTCCTGCAAGATATTTTTATAAACTTTTTTTTCTTGGTGGTTTTTGGAAAAGGGTGTGGGGGTGGGGCGCCGCTGGGGCAG
                 RT3b                                                                              RT3B
       GGCCAGGTTTTGTGTTTTAGTCCCTTGCTCCTGCTTCTTTCTACACACACATCTAAAGACGGTGCGGCTCGCTCTGTCATGGGTTCCGTCTCTCTCTGTGGAGAAGCAGCTCCACCTCTG
       GGGGGGCTCGGGGCAGAGGGCGGTGTCTCGTAGCGGGCGGCAGCGCCAGTCCCCCTCTGTCAGGCTGGGGCAATCTTGGTTTTGTGTCCAAAGGTGAAAGGGTAGGAGGAGGGCCCTCA
       GCTGGCCCTCCCCACACACAGGACGGCAGGGGCACTGTGAGGCTTTTCTTATTAAAATGAAAAAATTGAAAAAAAAGGACAAAGAGTCGGTGGCGCTCCTCTGCAGGGCGTTCTGTGCAG
       GGGCACAGCCCGGGTGCGGCGGCCGTGGGGGACGGCGGGTCTGATGCATGCCTCTGCCCATGGAGTCGTCTGTCTGCTTCGGTGCCTGCCCCTGCCTCCCACCCACCTCGTGTATAGATTT
       TAACGCTTCTGTTAACATTAGACCTCTGCCACAGGCTGGGATTTCTATACATAAGAACAAAAGCAAACACCTAGGACAGCAAACGCCAGGCGGTACAGGCGGGAAGGGGCTCTCCACGGA
       GATCGAGGACACGAAGCAAACTGCCTCTTGCTTGCCTTCCCCTTTTGTGCTTCGGACACACGCGGACTCCAGCAGGCGCCACGGAAATGGGCAAGCCCCTGCAGTGTACCCCTGTCATAA
       CTGTGAGCAGCTGCAGCTCCGGAACAATAAATCCCTTCCGCAAAGAC
```

B.

```
               1                                                                                      87
               MSNLTPESSPELAK K SWFGNFISLEKEEQIF V V I KDKPLSSIKADIVHAFLSIPSLSHSV I SQTSFRAEYKA T GGPAVFQKPVKFQV
T18992         MSNLTPESSPELAK K SWFGNFISLEKEEQIF V V I KDKPLSSIKADIVHXFLSIPSLSHSV I SQTSFRAEYKA T WWPAVFQKPFKFQX
W60403         MSSLTPESSPELAK R SWFGNFISLDKEEQIF L V L KDKPLSSIKADIVHAFLSIPSLSHSV L SQTSFRAEYKA S XAPPSSKSPSASRW
W41136         MSSLTPESSPELAK R SWFGNFISLDKEEQIF L V L KDKPLSSIKADIVHAFLSIPSLSHSV L SQTSFRAEYKA S GAPSVFQKPVRFQV
W18584                       GNFISLDKEEQIF L V L KDKPLSSIKADIVHAFLSIPSLSHSV L SQTSFRAEYKA S GAPPSSRSLSAFRW
```

Fig. 3 Nucleotide and predicted amino acids sequences of the *PEN11B* gene. **A** Exon sequences are represented in *capital letters* with adjacent intron sequences (*lower case*). Initiation and termination codons of the longest open reading frame are shown in *bold*. Primers protein sequence is given in the one-letter code. Primers for Reverse transcription-polymerase chain reaction and detection of the *Bst*UI polymorphism are indicated by *arrows* above the sequence. The poly- morphic *Bst*UI site is *boxed*. The polyadenylation signal is *underlined*. **B** The predicted protein sequence in the N-terminal portion of the gene is compared with those deduced from human and mouse expressed sequence tags (ESTs) in the database. Identical sequences between them are *shaded* and highly conserved substitutions are *boxed*. T18992 and W60403 are human ESTs. W18584 and W41136 are mouse ESTs

were highly conserved substitutions between the four genes (Fig. 3B).

Northern blots probed with the 3′ UTR DNA revealed 3.8-kb major and 4.2-kb minor positive bands mainly in the brain and pancreas, and less abundantly in the testis (Fig. 4). Transcripts of the gene were below the detectable level in the Northern hybridization in other tissues including the placenta. The 5′ end probe gave the same signal profiles in addition to a 2.8-kb intense signal in the brain which derived from the homologous gene (data not shown).

Allelic analysis of expression

A tentative goal in this study was to determine whether the gene was imprinted or not. We found a *Bst*UI–*Hha*I poly-

morphic site in the last exon for this purpose. Taking advantage of this polymorphism, allele usage in transcription of the gene was investigated by RT-PCR. While the same 3′ primer was used both in genomic and RT-PCRs, 5′ primers were designed from different portions for their PCRs. For RT-PCR, 5′ primers were prepared from exon 5 (first PCR) and exon 6 (nested PCR), which enables us to discriminate the amplification products of cDNA from those of residual contaminated genomic DNA (Fig. 3A). Thus, *Bst*UI digestions of PCR products yield 263-bp (a) and 208-bp (b) allelic bands common to both genomic and cDNA templates, and a 154-bp constant band, and 177- and 244-bp bands specific to genomic and cDNA PCRs, respectively (Fig. 5).

Transcripts were detected from both alleles, but close inspection of ethidium bromide-stained gels revealed a sig-

Adult tissues　　　　　　　　　　　　　　**Fetal tissues**

Fig. 4 *PEN11B* expression in adult and fetal tissues. The 3′ end-sequences of 3′ rapid amplification of cDNA ends clones were hybridized to Clontech Multiple Tissue Northern membranes containing 2 μg of mRNA per lane. *H*, heart; *B*, brain; *Pl*, placenta; *Lu*, lung; *Li*, liver; *M*, skeletal muscle; *K*, kidney; *Pa*, pancreas; *Pb*, peripheral blood lymphocytes; *C*, colon; *SI*, small intestine; *O*, ovary; *Te*, testis; *Pr*, prostate; *Th*, thymus; *S*, spleen
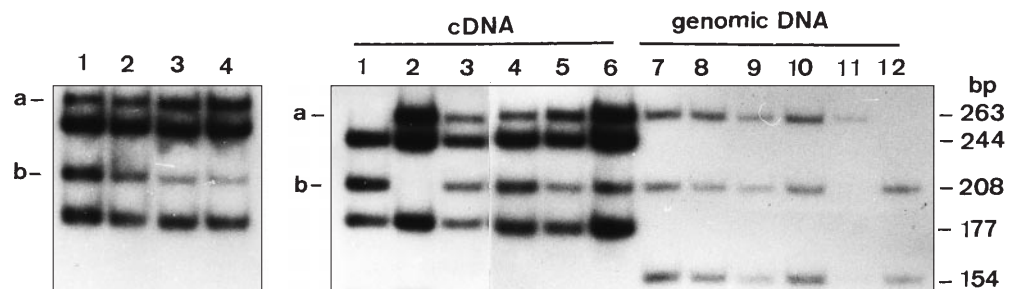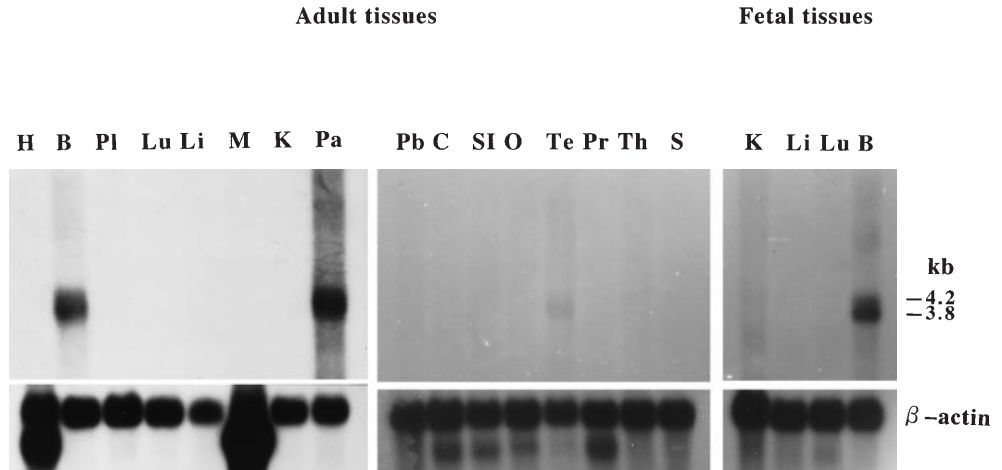
Fig. 5 Allelic analysis of *PEN11B* expression. (*Left*) A result of a mixing test is shown. Cloned cDNA without the polymorphic *Bst*UI site was mixed with that having the site at ratios of 0.5 (*lane 1*), 1 (*lane 2*), 2 (*lane 3*), and 3 (*lane 4*), amplified, and subjected to restriction with *Bst*UI. Allelic bands are indicated by *a* and *b* at the left. (*Right*) Two examples showing differential allele usage in placental villi with more intense allele b (*lanes 3* and *4*) and a (*lanes 5* and *6*), respectively, are presented along with their genomic polymerase chain reaction (PCR)/restriction profiles (*lanes 7–10*). *Lanes 1* and *2* and *lanes 11* and *12* are homozygous controls of reverse transcription (RT)-PCR and genomic PCR, respectively. Allelic bands are indicated by *a* and *b* at the left

nificant difference in intensity between allelic bands in placental specimens, whereas the genomic DNA gave a constant pattern with the uncleavable band (allele a) being somewhat more intense than the cleaved one (allele b). Then we compared transcription levels from each allele by quantifying radioactivity of allelic bands (a and b). As reassociation of amplified products in the plateau phase may make values incorrect, quantification was carried out as to PCR products at three consecutive points (each in at intervals of three cycles) in late logarithmic phase. Prior to restriction digestion, small aliquots of each PCR product were electrophoresed and counted in order to confirm their logarithmic amplification. Mixing tests using cloned cDNAs assured validity of this strategy, i.e. nested PCR correctly reflected the relative amount of starting materials without giving a bias (left panel in Fig. 5 and Table 1). While the ratio of allele a to b in genomic PCRs was 1.41 ± 0.04 ($n = 7$) (Table 2), that of RT-PCR in placental specimens varied from 0.71 to 5.17. When these values were corrected by corresponding genomic values, they ranged from 2.29 to 3.57 (in cases of a > b) and 0.50 to 0.61 (in cases of a < b). Thus, transcription levels showed approximately a 2-fold difference between each allele in human placenta (right panel in Fig. 5 and Table 3). Among seven cases examined, the parental origin of the more active allele could be ascer-

tained in five cases and it was all maternal. On the other hand, allelic difference was negligible, even if present, in other tissues, including brain, liver, and kidney (Table 3). These results were exactly reproducible in an additional experiment starting from a RT reaction. It is extremely unlikely that this preferential expression from the maternal allele in placenta resulted from maternal tissue contamination, because it was excluded by at least one of the polymorphic markers at multiple loci such as *WT1*, *TH*, *D11S29*, and *D16S423*, and because RT-PCR revealed that its expression was very low in decidua obtained from women who had an ectopic pregnancy. Therefore we conclude that the *PEN11B* gene is preferentially expressed from the maternal allele in human placenta and that it could be considered as an atypical imprinted gene.

Discussion

Stimulated by the discrepancy between conservation of imprinting status and remarkable divergence of the upstream sequences in the human and mouse *H19* genes, we searched *cis*-elements possibly involved in the human *H19* imprinting and found tandemly repeated pentamer sequences in the

Table 1 Results of mixing test in two separate experiments

Plasmid mix ratio (a/b)		Experiment 1			Experiment 2		
		19 cycles	22 cyles	25 cycles	19 cycles	22 cyles	25 cycles
0.5	product amount[a]	1	6.08	26.50	1	5.75	30.08
	a/b ratio[b]	0.94	0.82	0.86	0.83	0.71	0.82
	corrected ratio[c]	0.52	0.53	0.52	0.56	0.51	0.57
1	product amount	1	6.16	31.89	1	5.57	29.39
	a/b ratio	1.82	1.56	1.66	1.48	1.38	1.45
	corrected ratio	(1)	(1)	(1)	(1)	(1)	(1)
2	product amount	1	5.92	36.08	1	5.98	44.65
	a/b ratio	4.52	3.74	3.91	2.81	3.13	2.86
	corrected ratio	2.48	2.40	2.36	1.90	2.27	1.97
3	product amount	1	5.42	29.78	1	5.81	30.37
	a/b ratio	5.96	5.10	4.91	4.45	4.34	4.56
	corrected ratio	3.27	3.27	2.96	3.01	3.14	3.14

[a] Radioactivity of the target product at 19th cycle is arbitrarily represented as 1
[b] Ratio of radioactivity of allele a to that of allele b
[c] Actual a/b ratio was corrected by dividing it with that of the polymerase chain reaction of equally mixed DNA at each cycle

Table 2 Allelic quantification of genomic DNA

Case		Experiment			
		20 cycles	23 cycles	26 cycles	Average
1	product amount	1	6.17	31.55	
	a/b ratio	1.49	1.37	1.48	1.45
2	product amount	1	5.65	29.42	
	a/b ratio	1.38	1.40	1.48	1.42
3	product amount	1	6.58	30.68	
	a/b ratio	1.35	1.27	1.47	1.36
4	product amount	1	6.27	32.19	
	a/b ratio	1.46	1.48	1.44	1.46
5	product amount	1	5.75	36.28	
	a/b ratio	1.32	1.50	1.50	1.44
6	product amount	1	7.03	37.87	
	a/b ratio	1.27	1.47	1.38	1.37
7	product amount	1	6.33	46.97	
	a/b ratio	1.40	1.32	1.47	1.40

human *H19* 3′-flanking region. We were interested in investigating the role of the repetitive sequence for the human *H19* imprinting because of its unique structure and its similarity to the repetitious sequence essential for *RSVIgmyc* transgene imprinting. For this purpose, we isolated phage clones that contained sequences homologous to the repeat and identified a gene, *PEN11B*, in one of them. Whereas the *PEN11B* gene was transcribed from both maternal and paternal alleles in fetal tissues examined, preferential expression from the maternal allele was demonstrated in placenta, thus indicating an atypical imprinted gene. This imprinting effect appeared to be restricted to the placenta, which is reminiscent of the inactivation of murine X chromosome, i.e. paternal rather than random inactivation is kept only in extra-embryonic tissues (Migeon 1994). An intriguing feature of the placenta/extra-embryonic tissue is also observed in human *H19* imprinting (Jinno et al. 1995) and mouse *Ins* imprinting (Giddings et al. 1994). The placenta seems to have special status in imprinting mechanisms.

Incomplete imprinting might be explained by promoter-specific imprinting exemplified by *IGF2* imprinting (Vu and Hoffman 1994; Yun et al. 1998). The *PEN11B* gene has at least two forms of transcripts as shown in Northern blot analyses. While their precise structures remain to be determined, transcripts from one promoter would possibly mask allele usage of other transcripts from the other promoter. This possibility, however, seems less likely because the difference in the levels of the two transcripts surpasses the deviation of the allele-specific transcription. Alternatively, cell lineage-specific imprinting as in the case of the *Ube3a* gene (Albrecht et al. 1997) might have resulted in the apparent partial inactivation of the paternal allele. At present, this possibility cannot be examined in normal human specimens. Finally, the overall structure of the repeat itself might be the cause of the attenuated inactivation. The tandemly repeated pentanucleotide elements were more degenerate in the *PEN11B* gene than in *H19*.

We selected genomic clones by taking advantage of the mouse imprinting map (Cattanach and Beechey 1990) before identifying any genes in them. The selected clone is localized on chromosome 11p15.5, which forms one of the prominent human imprinting domains. Therefore it is also possible to speculate that the location of the gene was the decisive factor in its imprinting status, and that it was just a coincidence to have found the homologous pentanucleotide repeats both in *PEN11B* and *H19*. Many imprinted genes are indeed found in this region extending a few megabases along with biallelically expressed genes. It may be worth re-evaluating allele usage of those (non-imprinted) genes in detail, as in this study.

It may not be productive to use repetitive sequences to find novel imprinted genes, because the repeats identified so far in the vicinity of imprinted genes show no direct sequence homology to each other as suggested by Barlow et al. (Neumann et al. 1995; Smarzka et al. 1995). However, there are increasing examples of the association of repeated DNA sequences with gene regulation in many organisms (Henikoff and Matzke 1997). In addition, repetitive sequences are essential components of prominent structures of the genome, such as telomeres, centromeres, and the so-called heterochromatin, which are endowed with heritable

Table 3 Allelic quantification of reverse transcribed cDNA

Case		Experiment 1				Experiment 2			
		1	2	3	Average	1	2	3	Average
1 (villi)	maternal genotype a/a								
	product amount[a]	1	5.31	32.60		1	6.65	21.4	
	a/b ratio[b]	4.05	4.39	5.45	4.63	4.90	5.44	(9.37)	5.17[d]
	corrected ratio[c]				3.19				3.57
2 (villi)	maternal genotype a/b								
	product amount	1	5.03	37.71		1	5.56	23.18	
	a/b ratio	0.67	0.70	0.75	0.71	0.73	0.66	0.77	0.72
	corrected ratio				0.50				0.51
3	maternal genotype b/b								
(i) villi	product amount	1	6.58	30.68		1	6.60	36.9	
	a/b ratio	0.92	0.77	0.80	0.83	0.71	0.86	0.71	0.76
	corrected ratio				0.61				0.56
(ii) brain	product amount	1	6.59	39.86					
	a/b ratio	1.42	1.60	1.68	1.57				
	corrected ratio				1.15				
(iii) liver	product amount	1	5.33	27.5		1	6.65	32.79	
	a/b ratio	1.39	1.50	1.49	1.46	1.48	1.26	1.41	1.38
	corrected ratio				1.07				1.01
4	maternal genotype a/b								
(i) villi	product amount	1	6.43	46.46					
	a/b ratio	3.34	3.25	3.55	3.38				
	corrected ratio				2.32				
(ii) brain	product amount VC	1	5.75	33.5					
	a/b ratio	1.49	1.64	1.54	1.56				
	corrected ratio				1.07				
(iii) liver	product amount	1	4.42	37.48					
	a/b ratio	1.67	1.98	1.85	1.83				
	corrected ratio				1.25				
(iv) kidney	product amount	1	6.08	30.06					
	a/b ratio	1.95	2.03	2.04	2.01				
	corrected ratio				1.38				
5 (villi)	maternal genotype a/a								
	product amount	1	6.41	27.05					
	a/b ratio	3.89	3.47	3.67	3.68				
	corrected ratio				2.56				
6 (villi)	maternal genotype a/a								
	product amount	1	6.02	24.32					
	a/b ratio	3.30	3.10	3.02	3.14				
	corrected ratio				2.29				
7 (villi)	maternal genotype b/b								
	product amount	1	6.42	27.99					
	a/b ratio	0.74	0.78	0.78	0.77				
	corrected ratio				0.55				

[a] Radioactivity of the target product at the first point is arbitrarily represented as 1 PCR cycles ranged from 11 to 17 (brain) or 19 to 28 (other tissue)

[b] Ratio of radioactivity of allele a to that of allele b

[c] Actual a/b ratio was corrected by dividing it with that of the polymerase chain reaction of the corresponding DNA

[d] Average of the values at the first and second points

epigenetic modifications and affect the activity of adjacent genes. Therefore, repetitive sequences could continue to be a target in searching *cis*-elements involved in the imprinting mechanism.

Note added in proof
During submission of this manuscript, three ESTs, two human (A1014254 and T23947) and one rat (A1071294), and a human partial cDNA (AJ006701) were further identified. From comparison with the cDNA, the *PEN11B* gene may have more exons further upstream.

References

Albrecht U, Sutcliffe JS, Cattanach BM, Beechey CV, Armstrong D, Eichele G, Beaudet AL (1997) Imprinted expression of the murine Angelman syndrome gene, *Ube3a*, in hippocampal and Purkinje neurons. Nature Genet 17: 75–78

Cattanach BM, Beechey CV (1990) Autosomal and X-chromosome imprinting. Development Suppl: 63–72

Chaillet JR, Bader DS, Leder P (1995) Regulation of genomic imprinting by gametic and embryonic processes. Genes Dev 9: 1177–1187

Elson DA, Bartolomei MS (1997) A 5′ differentially methylated sequence and the 3′-flanking region are necessary for *H19* transgene imprinting. Mol Cell Biol 17: 309–317

Giddings SJ, King CD, Harman KW, Flood JF, Carnaghi LR (1994) Allele specific inactivation of insulin 1 and 2, in the mouse yolk sac, indicates imprinting. Nature Genet 6: 310–313

Henikoff S, Matzke MA (1997) Exploring and explaining epigenetic effects. Trends Genet 13: 293–295

Jinno Y, Yun K, Nishiwaki K, Kubota T, Ogawa O, Reeve AE, Niikawa N (1994) Mosaic and polymorphic imprinting of the WT1 gene in humans. Nature Genet 6: 305–309

Jinno Y, Ikeda Y, Yun K, Maw M, Masuzaki H, Fukuda H, Inuzuka K, Fujishita A, Ohtani Y, Okimoto T, Ishimaru T, Niikawa N (1995) Establishment of functional imprinting of the *H19* gene in human developing placenta. Nature Genet 10: 318–324

Jinno Y, Sengoku K, Nakao M, Tamate K, Miyamoto T, Matsuzaka T, Sutcliffe JS, Anan T, Takuma N, Nishiwaki K, Ikeda Y, Ishimaru T, Ishikawa M, Niikawa N (1996) Mouse/human sequence divergence in a region with a paternal-specific methylation imprint at the human *H19* locus. Hum Mol Genet 5: 1155–1161

Migeon BR (1994) X-chromosome inactivation: Molecular mechanisms and genetic consequences. Trends Genet 10: 230–235

Moulton T, Crenshaw T, Hao Y, Moosikasuwan J, Lin N, Dembitzer F, Hensle T, Weiss L, McMorrow L, Loew T, Kraus W, Gerald W, Tycko B (1994) Epigenetic lesions at the *H19* locus in Wilms' tumour patients. Nature Genet 7: 440–447

Neumann B, Kubicka P, Barlow DP (1995) Characteristics of imprinted genes. Nature Genet 9: 12–13

Steenman MJC, Rainier S, Dobry CJ, Grundy P, Horon IL, Feinberg AP (1994) Loss of imprinting of *IGF*2 is linked to reduced expression and abnormal methylation of *H19* in Wilms' tumour. Nature Genet 7: 433–439

Smarzka OW, Fae I, Stoger R, Kurzbauer R, Fischer GF, Henn T, Weith A, Barlow DP (1995) Conservation of a maternal-specific methylation signal at the human *IGF2R* locus. Hum Mol Genet 4: 1945–1952

Swain JL, Stewart TA, Leder P (1987) Parental legacy determines methylation and expression of an autosomal transgene: A molecular mechanism for parental imprinting. Cell 50: 719–727

Tremlay KD, Saam JR, Ingram RS, Tilghman SM, Bartolomei MS (1995) A paternal-specific methylation imprint marks the alleles of the mouse *H19* gene. Nature Genet 9: 407–413

Vu TH, Hoffman AR (1994) Promoter-specific imprinting of the human insulin-like growth factor-II gene. Nature 371: 714–717

Yoo-Warren H, Pachnis V, Ingram RS, Tilghman SM (1998) Two regulatory domains flank the mouse *H19* gene. Mol Cell Biol 8: 4707–4715

Yun K, Jinno Y, Sohda T, Niikawa N, Ikeda T (1998) Promoter-specific insulin-like growth factor 2 gene imprinting in human fetal liver and hepatoblastoma. J Pathol 183: in press

Zakian VA (1989) Structure and function of telomeres. Annu Rev Genet 23: 579–604

Zhang Y, Tycko B (1992) Monoallelic expression of the human *H19* gene. Nature Genet 1: 40–44