 Human Molecular Genetics, 2000, Vol. 9, No. 12 **1829–1841**

Sequence conservation and variability of imprinting in the Beckwith–Wiedemann syndrome gene cluster in human and mouse

Martina Paulsen[1,2], Osman El-Maarri[1,+], Sabine Engemann[1,+], Martin Strödicke[1,+], Olivia Franck[1], Karen Davies[3], Richard Reinhardt[1], Wolf Reik[3] and Jörn Walter[1,§]

[1]Max-Planck-Institut für Molekulare Genetik, Ihnestraße 73, D-14195 Berlin, Germany, [2]University of Cambridge, Department of Anatomy, Downing Street, Cambridge CB2 3DY, UK and [3]Laboratory of Developmental Genetics and Imprinting, The Babraham Institute, Cambridge CB2 4AT, UK

Received 27 March 2000; Revised and Accepted 15 May 2000 DDBJ/EMBL/GenBank accession nos AJ251835, AJ271092, AJ279791–AJ279797

In human and mouse most imprinted genes are arranged in chromosomal clusters. This linked organization suggests coordinated mechanisms controlling imprinted expression. We have sequenced 250 kb in the centre of the mouse imprinting cluster on distal chromosome 7 and compared it with the orthologous Beckwith–Wiedemann gene cluster on human chromosome 11p15.5. This first comparative imprinting cluster analysis revealed a high structural and functional conservation of the six orthologous genes identified. However, several striking differences were also discovered. First, compared with the mouse the human sequence is ~40% longer, mostly due to insertions of two large repetitive clusters. One of these clusters encompasses an additional gene coding for a homologue of the ribosomal protein L26. Second, pronounced blocks of unique direct repeats characteristic of imprinted genes were only found in the human sequence. Third, two of the orthologous gene pairs *Tssc4*/*TSSC4* and *Ltrpc5*/*LTRPC5* showed apparent differences in imprinting between human and mouse, whereas others like *Tssc6*/*TSSC6* were not imprinted in either organism. Together these results suggest a significant functional and structural variability in the centre of the imprinting cluster. Some genes escape imprinting in both organisms whereas others exhibit tissue- and species-specific imprinting. Hence the control of imprinting in the cluster appears to be a highly dynamic process under fast evolutionary adaptation. Intriguingly, whereas imprinted genes within the cluster contain CpG islands the non-imprinted *Ltrpc5* and *Tssc6*/*TSSC6* do not. This and additional comparisons with other imprinted and non-imprinted regions suggest that CpG islands are key features of imprinted domains.

INTRODUCTION

Imprinted genes are those that are predominantly expressed from only one of the parental alleles (1–6). The imprinting mechanism involves epigenetic modifications such as DNA methylation and altered chromatin structures (7). Imprinted genes play important roles in development, in particular in fetal growth, placental development, cell proliferation and behaviour after birth. Deregulation of imprinting is implicated in a number of human cancers and genetic diseases, such as the Prader–Willi/Angelman syndromes (8) and Beckwith–Wiedemann syndrome (BWS) (9), which is associated with fetal overgrowth and a predisposition to childhood tumours.

A significant number of imprinted genes are organized in clusters that are characterized by epigenetic features such as differences in allele-specific DNA replication, chromosome pairing of the two parental copies and sex-specific meiotic recombination frequencies (10–13). This suggests an epigenetic co-regulation of neighbouring imprinted genes. Some aspects of regulation of imprinting could affect the whole cluster, either by the chromatin structure itself or by *cis*-acting elements such as imprinting centres or *cis*-acting transcripts.

One prominent imprinting cluster in the human is the BWS region on chromosome 11p15.5. In the mouse the imprinting cluster on distal chromosome 7 is homologous to the human BWS region. We and others have identified a number of mouse genes in this region at the same relative positions as their human orthologues, indicating a syntenic organization in both organisms (14–19). In the mouse and in the human the imprinting clusters contain the paternally expressed *Igf2*/*IGF2* (20–22) and *Ins2*/*INS* genes (23), the maternally expressed *H19* (24,25), *Mash2*/*ASCL2* (*HASH2*) (26,27), *Cdkn1c*/*CDKN1C* (28–31), *Tssc3*/*TSSC3* (14), *Tssc5*/*TSSC5* (15,16) and *Kcnq1*/*KCNQ1* (17–19,32) genes and *Cd81*/*CD81*, whose maternal expression has so far only been shown in mouse (19,33,34). Antisense transcripts have been identified for

[+]These authors contributed equally to this work

[§]To whom correspondence should be addressed. Tel: +49 30 84131274; Fax: +49 30 84131385; Email: walter@molgen.mpg.de

mouse *Igf2* (35), for human *TSSC5* (15,16,36) and in both organisms for *Kcnq1*/*KCNQ1* (37–39). The *Igf2* and *Kcnq1*/*KCNQ1* antisense transcripts *Kcnq1ot1*/*KCNQ1OT1* (*Lit1*/*LIT1*) are expressed from the paternal allele; imprinting of the *TSSC5* antisense transcript has not yet been investigated. The imprinted region is flanked by the non-imprinted genes *Nap1l4*/*NAP1L4* upstream of *Tssc3*/*TSSC3* (17,40) and *Rrpl23l*/*RPL23L* downstream of *H19* (41). In the mouse the *Nctc1* gene between *H19* and *Rrpl23l* has also been shown to be biallelically expressed (42).

Regionally coordinated control of imprinting has been shown for the *Ins2*, *Igf2* and *H19* genes in the mouse (43) and is likely in the human as well (44–46). However, the *Kcnq1* and *Cdkn1c* genes are apparently not affected by the *Igf2–H19* controls (19). On the other hand, *CDKN1C* mutations (29,47–49), loss of imprinting of *KCNQ1OT1* (37) and loss of imprinting of *IGF2* (46,50) are all associated with BWS, suggesting that there might be control mechanisms that involve the whole cluster. Indeed, translocations in BWS patients in the *KCNQ1* gene are associated with loss of imprinting of *IGF2* (39,51). Nevertheless, there seems to be a tripartite structure of the cluster as judged by the fact that the human *TSSC4* and *TSSC6* genes, which are located between *KCNQ1* and *ASCL2*, are not imprinted (52). Similarly, the mouse *Cd81* gene, which is located in the same region, only shows imprinting in very early development (19).

This region between the *Kcnq1* and *Mash2* genes therefore seems to be of particular interest in carrying out a systematic analysis of sequence features which may be involved in either the maintenance of or escape from imprinting in the cluster. This question lends itself to a comparative sequence analysis between mice and humans, which has been shown to be highly valuable for the detection of genes, organizational features and potential regulatory sequences (53,54).

Here we have carried out the first comparative sequence analysis in an imprinting cluster. Based on a previously generated BAC contig in the mouse (17) we sequenced a 250 kb region between *Kcnq1* and *Mash2*. We compared the gene organization between mouse and human and identified six orthologous genes in this region. Three of these were identified as novel genes in the mouse and were examined for their imprinting status. The results provide insights into the structural organization and the control mechanisms of imprinted expression in the BWS gene cluster. In particular, they highlight the importance of CpG islands for imprinted genes.

RESULTS

Sequencing of BAC 300P2 and identification of genes

In a previous work we generated a contig of BAC clones (17) covering 1 Mb of the imprinting cluster on distal chromosome 7 in the mouse between genes *Nap1l4* and *Mash2*. A 250 kb large BAC (300P2) covering a gene-rich region between exon 9 of *Kcnq1* and *Mash2* was chosen for sequencing. Analysis in the human suggested that some of the genes within this region, like *KCNQ1*, *LTRPC5* and *ASCL2*, are imprinted, whereas others, like *TSSC4* and *TSSC6*, are not (27,32,52,55). Our aim was to identify the mouse orthologues of these human genes, to analyse their imprinting status and to identify sequence features that characterize imprinted versus non-imprinted genes/regions. Therefore, the complete sequence of BAC 300P2 was determined by assembling 4205 sequence reads obtained from a 'shotgun' pUC18 library with an average insert size of 1.0–1.5 kb. The finished high quality sequence (GenBank accession no. AJ251835) consists of a single contig with a length of 249 487 bp starting 1887 bp 3′ of exon 9 of *Kcnq1* and ending 18 997 bp 3′ of the last exon of *Mash2* (Fig. 1). Genes in the BAC 300P2 sequence were identified either by homology searches against the partially annotated human genomic sequence (GenBank accession nos AC002536, AC003693 and AJ006345), transcribed sequences in the GenBank databases for non-redundant sequences and expressed sequence tags (ESTs) or by using a series of gene prediction programs (see Materials and Methods). The genomic locations and structures of five complete genes and ~90 kb of the 5′ part of the *Kcnq1* gene (including exon 9) were thus identified. *Cd81*, *Mash2* and *Kcnq1* have been described previously (17–19,26). *Ltrpc5*, *Tssc4* and *Tssc6* were newly identified in the mouse.

Sequence comparison between human and mouse

The first intriguing observation when comparing the mouse BAC 300P2 with the corresponding human (GenBank accession nos AC002536A, AC003693 and J006345) sequence is the almost completely conserved order of the genes located in the respective genomic regions (Fig. 1). The comparison revealed that not only the relative order of genes, *Kcnq1*/*KCNQ1*–*Ltrpc5*/*LTRPC5* (*MTR1*)–*Tssc4*/*TSSC4*–*Cd81*/*CD81* (*Tapa1*/*TAPA1*)–*Tssc6*/*TSSC6*–*Mash2*/*ASCL2* (*HASH2*), but also their relative transcriptional orientation and, for the majority of genes, the size and distribution of exons are highly conserved. This conservation is directly visible at the level of the genomic sequences, which can be aligned along the entire region with the exception of two major gaps (Fig. 2). These are caused by two remarkably large insertions of A/T-rich repeated sequences in the human chromosome (Figs 1 and 2), one in the 5′ part of the *KCNQ1* gene and the other between *CD81* and *TSSC6*. The latter includes an additional intronless gene coding for a homologue of the ribosomal protein L26 (Fig. 1). The absence of this gene in the mouse sequence and the fact that it is flanked by a number of retrotransposable elements suggests that it is a recently integrated copy of the already identified *RPL26* gene residing on chromosome 17p (GenBank accession no. AF083248) (56). Both repetitive insertions are rich in L1 elements and contribute to the overall higher proportion of repetitive elements (38.5% in the human sequence compared with only 23.5% in the mouse sequence). The two insertions are the main cause of the size difference of the homologous regions in human (344 kb) and mouse (249 kb).

Using prediction software (MarFinder) a few strong putative matrix attachment regions (MARs) (Fig. 1) were localized in both genomes. These MARs bracket the central part of the sequenced region including the *Kcnq1*/*KCNQ1* CpG island and the entire *Ltrpc5*/*LTRPC5*, *Tssc4*/*TSSC4* and *Cd81*/*CD81* genes.

Characteristic unique repetitive elements have previously been identified in the vicinity of some known imprinted genes and it has been argued that they might play a role in the epigenetic control of expression (57). When screening the entire region in human and mouse for such short unique tandemly arranged

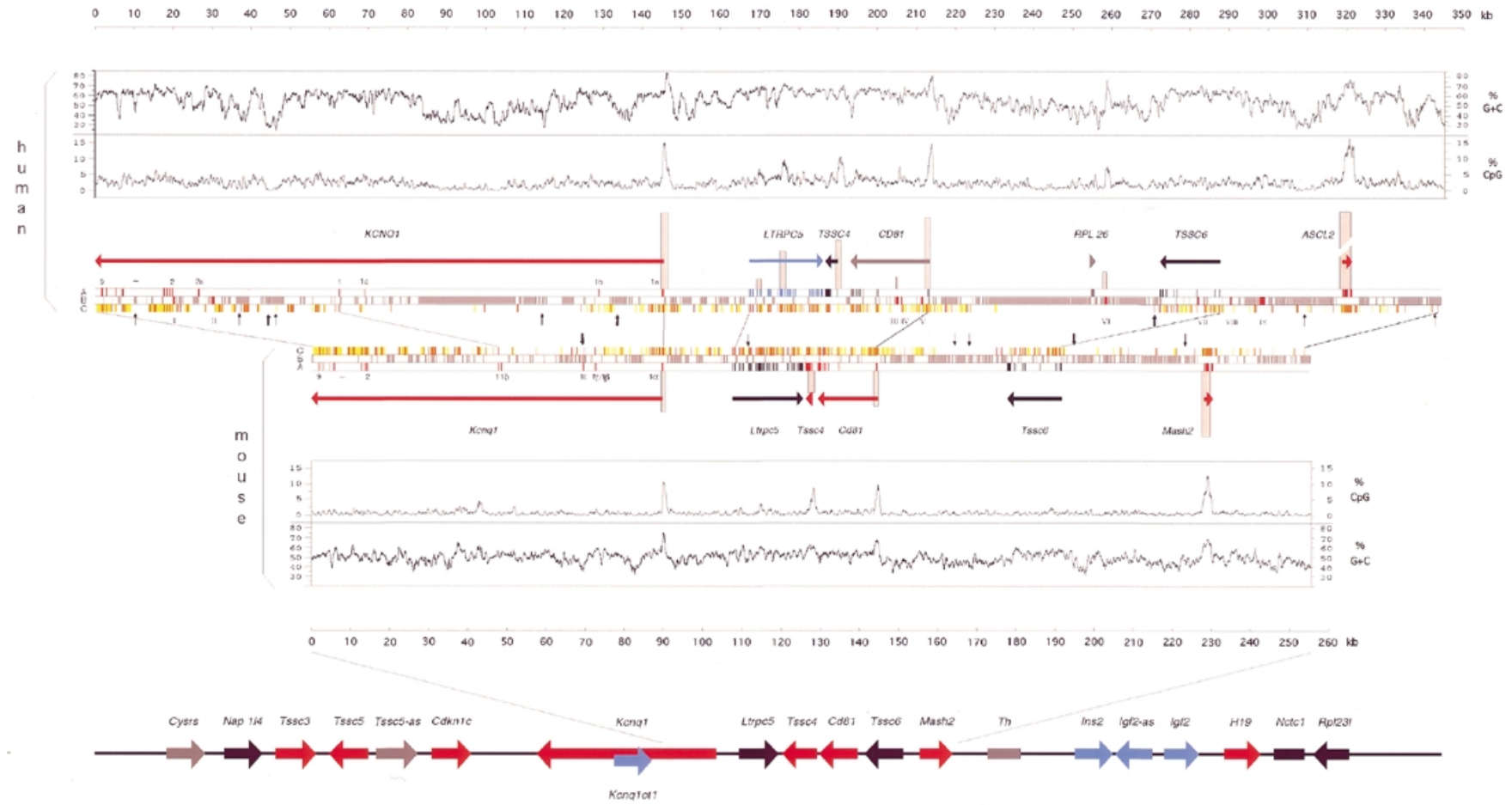

Figure 1. Localization of genes and genomic sequence features between *Kcnq1*/*KCNQ1* and *Mash2*/*ASCL2* on human chromosome 11p15.5 and distal mouse chromosome 7. An overview of the entire imprinting cluster on mouse distal chromosome 7 is shown below. Genes are marked by horizontal arrows indicating their relative transcriptional orientation. Red, blue and black colouring defines maternally, paternally or biallelically expressed genes, respectively, grey colouring those genes whose imprinting status has not been analysed. The upper part of the figure shows the detailed graphic map of the orthologous mouse and human sequences. The overall G+C and the CpG densities (window size 500 bp, values given as percentages) are presented as plots above and below the sequences. In the central part of the figure three graphic panels summarize sequence features of the respective human and mouse regions. Genes are marked by coloured arrows, CpG islands by grey columns with their height representing the relative CpGs island score and width the extension of the islands. (**A**) The intron–exon structure of each gene. For the human *ASCL2* gene, the intron–exon structure is shown as predicted by comparison with mouse and rat cDNA sequences. (**B**) The distribution of repetitive elements (grey); short unique sequences with a repetitive structure are marked as roman numbered red boxes (Table 1). (**C**) Sequence elements (>40 bp) with identities between 40 and 70% (yellow) and >70% (red). The localization of matrix attachments sites is given by thick (score/average strength >0.8) and thin (score/average strength <0.8) vertical arrows. Computer programs used for the various graphic representations are described in Materials and Methods.

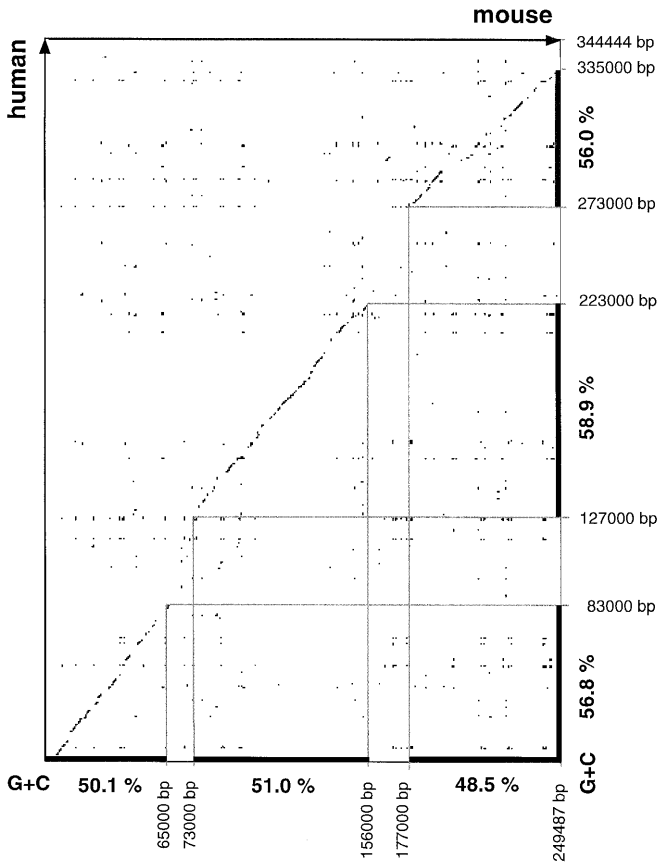

Figure 2. Dot plot comparison (PipMaker, http://globin.cse.psu.edu/cgi-bin/pipmaker) of the human and mouse genomic sequences shown in Figure 1. A colinear alignment is observed for three large regions (thick bars) between human and mouse. The G+C content is denoted for both organisms.

sequence elements (motif size >7, element size >200), nine repeated motifs could be identified in the human but no such elements were detected in the mouse (Table 1 and Fig. 1). The repeated motifs in the human extend to a total length of up to 1.3 kb, with the longest repeat unit being 56 bp. Except for one of the motifs (VI) no homologies to other genomic regions were detected in the GenBank database. The absence of such unique tandem repeats in the mouse sequence and the finding that these are not necessarily linked to imprinted genes in the human do not support the hypothesis that such repeats are defining hallmarks of imprinted genes (57).

Comparison of the gene structures in human and mouse

Cd81, Mash2 and Kcnq1. The precise placements of exons of the *Cd81* (34) and *Mash2* genes (58) and the 5′ part (up to exon 9) of *Kcnq1* (17) are shown in Figure 1. The exact coordinates of the exon–intron structures are deposited with the BAC 300P2 sequence (GenBank accession no. AJ251835). A comparison with the human sequence revealed a strong conservation of gene structures in mouse and human. In the human, however, the *ASCL2* gene has only been partially identified and its 5′-end was missing. A comparison of human sequences with mouse and rat cDNAs and genomic sequences (data not shown) suggests the location of the putative 5′ exon of the *ASCL2* gene (see Fig. 1 and legend).

In contrast to the majority of genes in this region, whose relative exon–intron structure is quite conserved between human and mouse, the 5′-ends of the *Kcnq1*/*KCNQ1* genes show remarkable differences between the two species (Fig. 1). The second intron in the human is ~40 kb larger than the corresponding intron in the mouse. This increase mainly results from the large insertion of repetitive sequences (Fig. 1) between exons 1b and 1c in the human. Exons homologous to the mouse exons 1β–1ε are not present in the human sequence (17). Conversely, homologues of the human exons 1b, 1c and 2a (32) were not found in the mouse. The differences in the 5′ portions of the mouse and human *Kcnq1*/*KCNQ1* genes may contribute to differences in expression and function in both organisms (17,32). In contrast to the differences in the 5′ region of *Kcnq1*, the sequences 3′ of exon 1 are highly homologous between mouse and human.

Tssc4. Tssc4 is located 1.4 kb 3′ of the last exon of *Cd81* and transcribed in the same orientation; the distance to the 3′-end of the oppositely orientated *Ltrpc5* is 0.7 kb. The *Tssc4* cDNA sequence of 1394 bp (GenBank accession no. AJ279796) was assembled from 73 ESTs matching the BAC 300P2 sequence.

Table 1. List of the short unique repetitive DNA sequence elements I–IX found in the human genomic sequence between *KCNQ1* and *ASCL2* (Fig. 1)

	Repeat motif	Motif length (bp)	Repeat length (bp)
I	CACCAGGAACNCCAGCTTGGGCCAGAGGGCGTCCCACGCCAGGAACCCNAGCNTGGGTCAGAGGGGTCCCA	72	216
II	CCCCCAGGATGGACGTCA	18	508
III	ACAGACGACGGGCA	14	233
IV	ACCCCACAAGCCGTCCC	17	270
V	GTCTCCCTCAGCCCCCACCCCCAGGTCCACA	31	313
VI	ACGGGGTCTCGGCCGGGCAGAGGCGCTCCTCACATCCCAG	40	600
	followed by 3× ATGGGGCGACTGGCCGGGCGGGGGGCTGACCCCCCCCACCTCCCTCCCGG	50	150
VII	CTGGTCATTGCGGGGGG	17	390
VIII	CCCACACACCCCTGCCCAC	19	300
IX	CGGGTGGCACGCCTCTGCGAATATACTAAAGCGGGGAGTTGTTTTTGGGGGTGCTG	56	1145

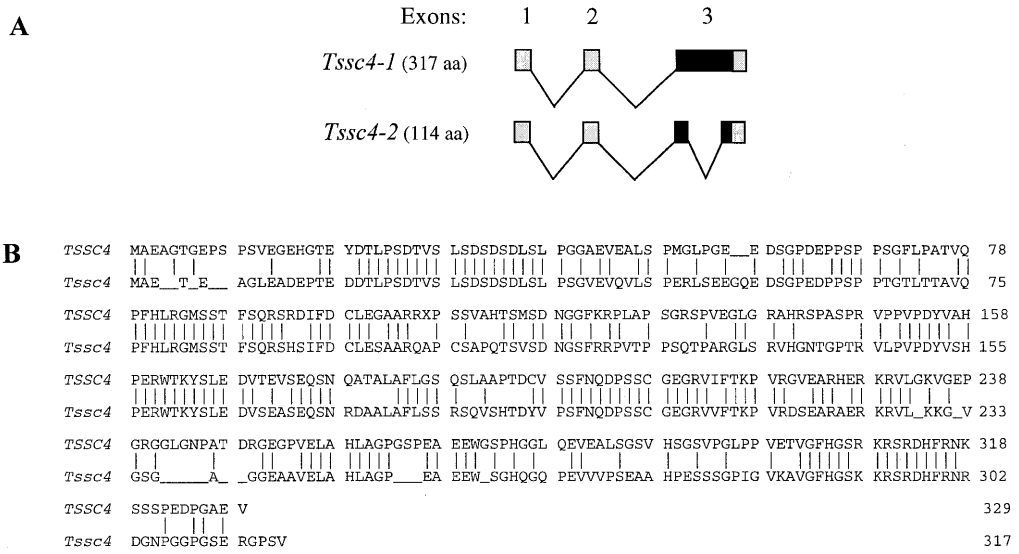

Figure 3. Structure of mouse *Tssc4*. (**A**) The two different splice variants found (*Tssc4-1* and *Tssc4-2*) are shown with coding regions labelled in black. (**B**) Sequence alignment of the human TSSC4 and the mouse Tssc4-1 deduced protein sequences.

The complete cDNA sequence was verified by sequencing one of the ESTs (GenBank accession no. AA268143) and RT–PCR products. A putative polyadenylation signal is present 20 bp upstream of the 3′-end. The human orthologue of *Tssc4* has previously been identified by a subtractive cDNA screen and is regarded as a potential tumour suppressor gene (52). Both mouse *Tssc4* and human *TSSC4* are very compact genes. Although the mouse *Tssc4* gene is composed of three exons spanning 1.7 kb of genomic sequence, the published cDNA sequence of the human orthologue consists of only two exons (52). However, an additional exon at the 5′-end is indicated by an overlapping EST sequence (GenBank accession no. AW249431). Based on this finding, the finally assembled human *TSSC4* gene contains three exons spanning 3131 bp of genomic DNA and the cDNA has a length of 1503 bp. In the human only the first exon is located in a CpG island, whereas in the mouse the first and the second exons are positioned in a 1.3 kb long CpG island. Sequence similarity (61% sequence identity) is confined to exon 3 of *Tssc4*/*TSSC4*. This exon harbours an open reading frame (ORF) of 951 bp that potentially encodes a peptide of 317 amino acids. Although syntenically placed and actively transcribed, *Tssc4*/*TSSC4* show a significantly lower level of sequence conservation between human and mouse compared with the flanking genes *Cd81*/*CD81* and *Ltrpc5*/*LTRPC5*. Compared with the deduced human *TSSC4* peptide (329 amino acids) the overall similarity value of 67.8% results from only six highly conserved stretches of 15–20 amino acids (Fig. 3). This patterned homology suggests a conservation of important functional domains. A motif search using these homologous regions against sequences in public databases did not reveal significant similarities to known proteins. RT–PCR revealed the existence of an alternative *Tssc4* transcript (*Tssc4-2*) which is 498 bp shorter as a result of an internal splicing event in exon 3. This alternative splicing generates a translational shift and a truncated peptide of 114 amino acids (Fig. 3A).

Tssc6. The mouse homologue of the human *TSSC6* gene, which was isolated as a transcript with potential tumour suppressor activity (52), is located between *Cd81* and *Mash2* (Fig. 1) and transcribed in the same direction as *Cd81* and *Tssc4*. The *Tssc6* gene is split into eight exons covering 13.8 kb of genomic sequence. Based on two clusters of ESTs matching to the 3′- and 5′-ends of the gene, respectively, a 1657 bp long cDNA sequence (GenBank accession no. AJ279791) was generated by combining these EST sequences with sequences of overlapping RT–PCR products. The cDNA sequence encompasses an ORF of 678 bp potentially coding for a peptide of 226 amino acids. Similar to Tssc4, the primary sequence of *Tssc6* does not show any significant similarities to other proteins in public databases except to the human TSSC6 peptide and some weak similarity (29%) to the peptide of the neighbouring *Cd81* gene, a transmembrane protein. The human- and mouse-derived peptide sequences are only 62.8% identical (Fig. 4). In contrast to the Tssc4/TSSC4 orthologues, the homologies between Tssc6/TSSC6 are equally distributed along the entire length of the sequence. However, the mouse Tssc6 peptide is 68 amino acids shorter than the human and this dissimilarity is also reflected in the DNA sequences. Whereas the sequences of the first seven exons of both mouse and human *Tssc6*/*TSSC6* genes are quite similar (68–81% identity), *Tssc6* exon 8 does not show any similarity to the published human *TSSC6* cDNA and genomic sequences.

Analysis of RT–PCR products between exons 1 and 8 from RNA isolated from tissues (i.e. lung and spleen tissues) of newborn mice detected four additional *Tssc6* splice variants (*Tssc6-2*–*Tssc6-5*) (GenBank accession nos AJ279792–AJ279795). The shorter splice variants lack various internal exons (Fig. 4). Consequently, the alternative splicing shortens the ORFs or leads to translational frameshifts and truncated ORFs with a stop codon shortly after exon 5. The putative peptides have the following sizes: Tssc6-2, 198 amino acids;

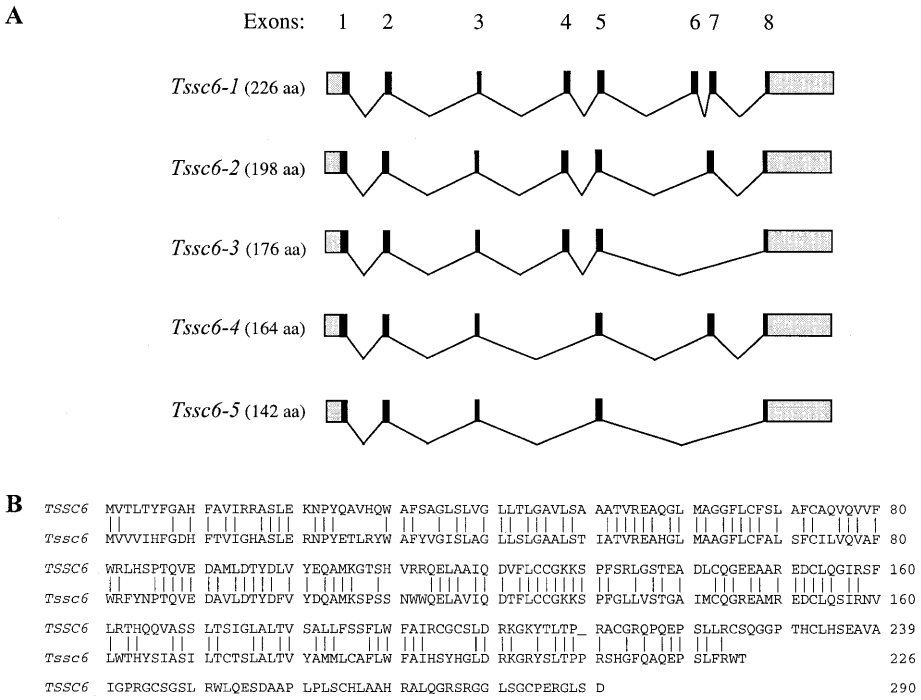

Figure 4. Structure of mouse *Tssc6*. (**A**) The five alternative transcripts of mouse *Tssc6* as found by RT–PCR; coding regions are labelled in black. (**B**) Alignment of the human TSSC6 and mouse Tssc6-1 deduced protein sequences.

Tssc6-3, 176 amino acids; Tss6-4, 164 amino acids; Tssc6-5, 142 amino acids.

Ltrpc5. *Ltrpc5* is located 17.8 kb upstream of the first exon of *Kcnq1* and encompasses at least 17.8 kb of genomic sequence. The human orthologue of this gene was identified on the basis of sequence similarity and one human EST (GenBank accession no. AA577486) described in the database. Expression of the human *LTRPC5* (*MTR1*) gene has recently been described (55). The genes should be named *Ltrpc5*/*LTRPC5* (large transient receptor potential-related channel 5) in accordance with the unifying nomenclature (59).

Ltrpc5 is transcribed in the opposite direction to the *Kcnq1*, *Tssc4*, *Cd81* and *Tssc6* genes. As no matching mouse ESTs (and only one human) were found in the databases our initial discovery of *Ltrpc5* was based on exon predictions and sequence similarities between human and mouse genomic sequences. The similarity within the predicted 24 exons ranges from 86 to 92% identity on the DNA level. Transcription of the mouse *Ltrpc5* gene was examined by RT–PCR between exons 1 and 24 and an almost full-length cDNA sequence (GenBank accession no. AJ271092) was assembled. The 5′-end of the first exon is predicted by sequence similarity to the human cDNA. The 3′-end, containing one potential polyadenylation signal, was identified by 3′-RACE. Alternative splicing resulting in an additional transcript was observed within the last two exons of the gene (Fig. 5A). No alternative splicing involving exon 18 was detected as described for the human *LTRPC5* gene (55). The start of the ORF was predicted by its sequence similarity to the human and the proposed translation start point matches perfectly to a Kozak consensus sequence. The two *Ltrpc5* transcripts comprise ORFs potentially coding for peptides of 1148 and 1138 amino acids length, respectively. Both isoforms contain the typical TRPC topology of six transmembrane domains. The peptide sequences deduced from the long mouse *Ltrpc5* transcript and human *LTRPC5* are 82% identical (Fig. 5). Within the transmembrane domains the similarity rises to 98% (Fig. 5). Database searches revealed that the Ltrpc5 protein shows highest similarities to human melastatin (30% identity) and LTRPC7 (37% identity) as well as to a hypothetical *Caenorhabditis elegans* TRPC-like protein (GenBank accession no. CAB05572) (26% identity). These similarities suggest that *Ltrpc5* codes for a potential membrane-spanning protein regulating Ca^{2+} entry into cells. The predicted structure of the Ltrpc5 protein assigns it to a subfamily of long transient receptor potential proteins (LTRPCs) (59).

Imprinting analysis of *Tssc4*, *Tssc6* and *Ltrpc5*

Kcnq1, *Cd81* and *Mash2* have previously been shown to exhibit imprinted expression (17–19,26). Imprinting of these genes appears to be restricted to specific tissues and developmental stages. Here we describe the imprinting status of the remaining three novel mouse genes within the sequenced region. To monitor possible developmental and tissue-specific differences the studies were performed on RNA preparations from whole embryos and placentae at two developmental stages (12.5 and 16.5 d.p.c.) and a variety of organs isolated from embryos (16.5 d.p.c.) and newborn mice. A *Tssc4* cDNA probe (GenBank accession no. AA646792) hybridized on a northern blot to various transcripts of ~1.7–1.9 kb length (data not shown). We were able to amplify RT–PCR products for all three genes from all tissues analysed, suggesting a ubiquitous

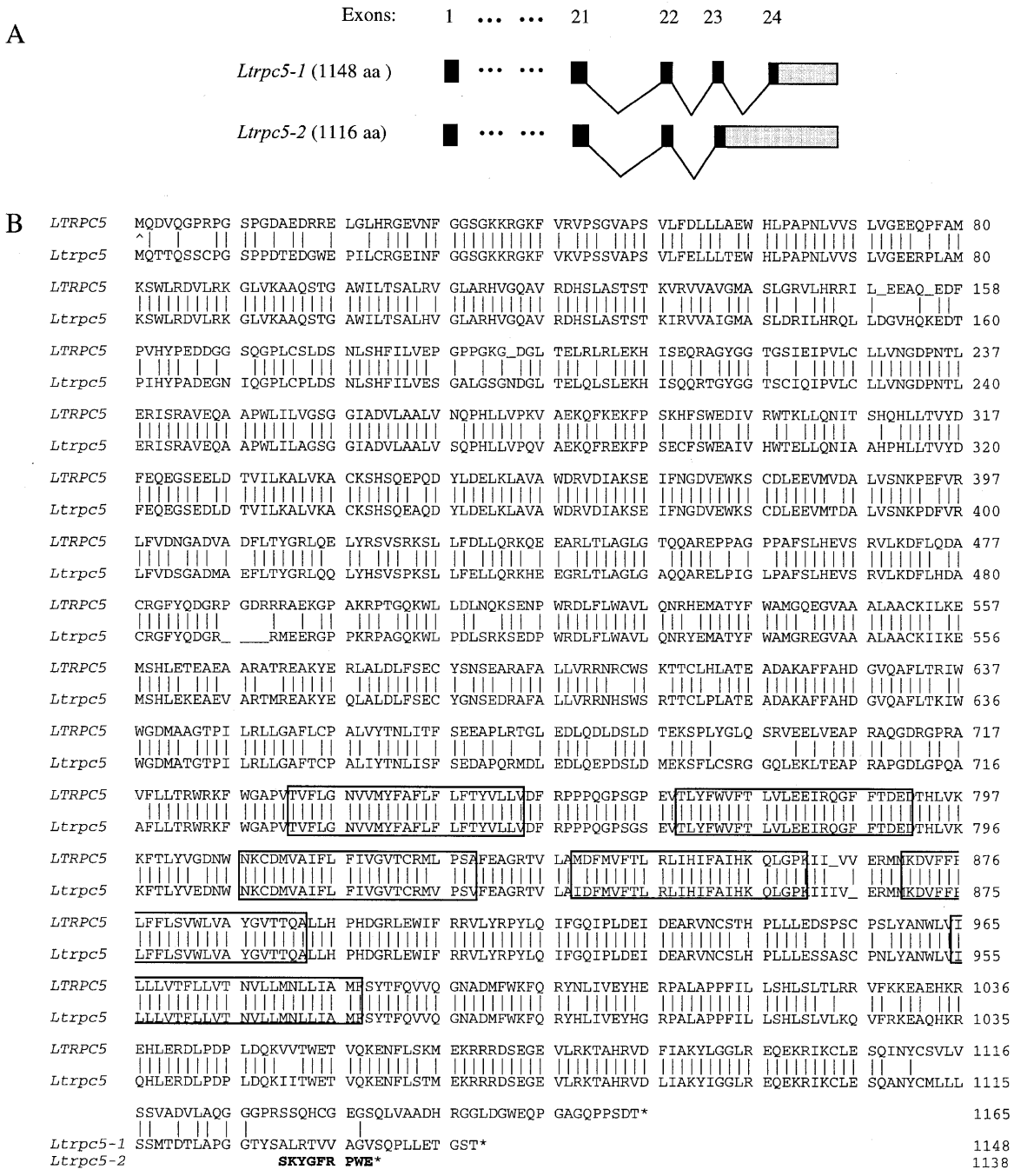

Figure 5. Structure of mouse *Ltrpc5*. (**A**) The two alternative splice variants of *Ltrpc5*; coding regions are labelled in black. Both transcripts are identical between exons 1 and 22. (**B**) Alignment of the human and mouse LTRPC5/Ltrpc5 deduced protein sequences. The proposed transmembrane domains are boxed. The peptides encoded by the two transcripts [see (A)] differ only in the very C-terminal sequences due to a frameshift in *Ltrpc5-2*. These amino acids are highlighted in bold.

expression pattern of *Tssc4*, *Tssc6* and *Ltrpc5*, with the exception of *Ltrpc5* transcripts in placentae (12.5 d.p.c.).

To perform an allele-specific expression analysis we used RNA derived from progeny of C57BL/6 × SD7 crosses as well as from reciprocal crosses. SD7 is a congenic mouse strain which carries a defined segment of distal chromosome 7 of *Mus spretus* (SEG) on an *M.musculus* background (60). In all three genes we found DNA polymorphisms in restriction sites that allowed us to determine the allelic origin of the transcripts. RT–PCR products were cut with the relevant restriction endonucleases and the relative intensity ratios of the allele-specific DNA fragments were visualized by agarose gel electrophoresis and staining. At all three developmental stages, i.e. 12.5 d.p.c., 16.5 d.p.c. and newborn mice, and in all different tissues we detected RT–PCR products from both parental alleles for *Tssc6* and *Ltrpc5*, indicating biallelic expression (Fig. 6). A

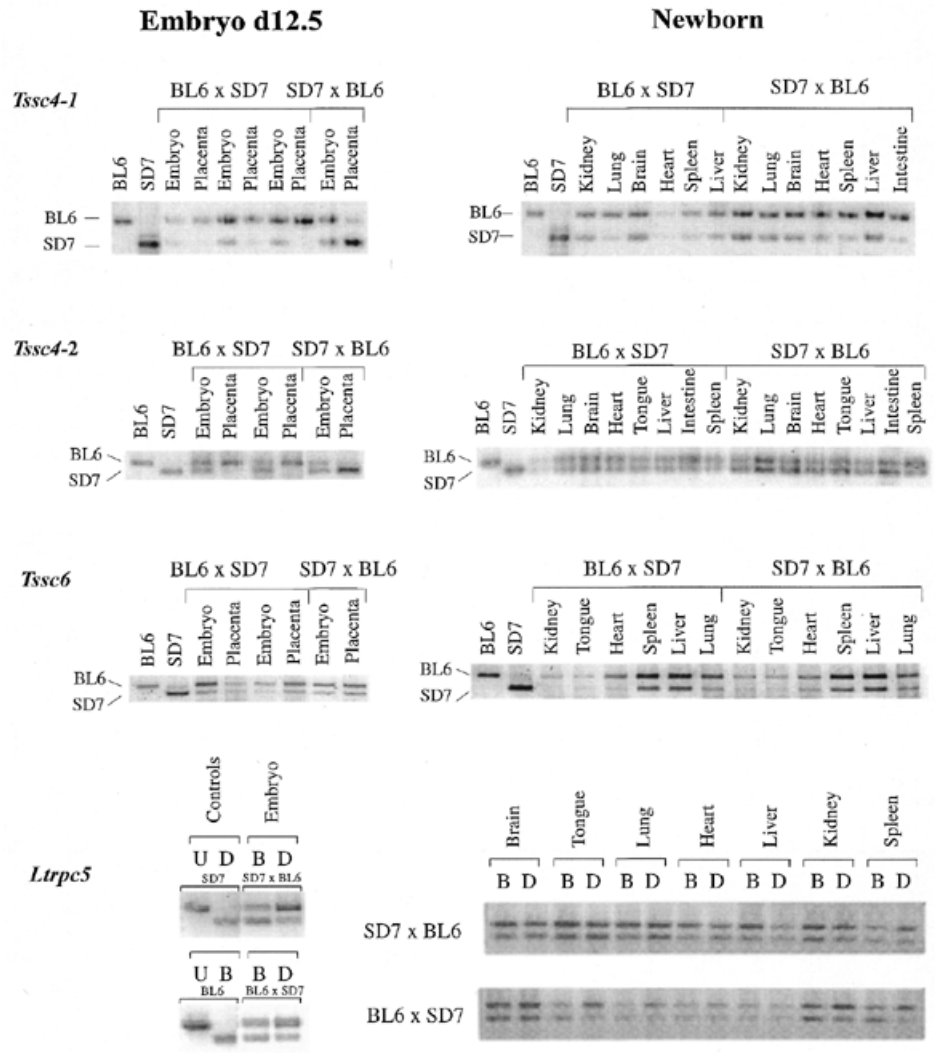

Figure 6. Allele-specific expression analysis of *Tssc4*, *Tssc6* and *Ltrpc5* transcripts. RT–PCR was performed on total RNAs isolated from whole embryos, placentae (12.5 d.p.c.; 16.5 d.p.c. data not shown) and organs of newborn mice. Results are shown for both reciprocal crosses C57BL/6 × SD7 and SD7 × C57BL/6. RT–PCR on *Tssc4* (Fig. 3) was performed using primer pairs specific for the individual transcripts (see Materials and Methods). The obtained RT–PCR products were cut with *Tai*I (*Tssc4-1*) and *Alu*I (*Tssc4-2*), respectively. Both endonucleases only cut RT–PCR products derived from the SD7 allele. For *Tssc6* the DNA polymorphism used is located in exon 7. *Tssc6* was therefore analysed using primers between exons 7 and 8 that are only present in splice variants *Tssc6-1*, *Tssc6-2* and *Tssc6-4* (Fig. 4A). PCR products were cut with *Tai*I which only cleaves the SD7 allele. RT–PCR for *Ltrpc5* was performed between exons 16 and 19. The obtained products were cut at a polymorphic *Bam*HI/*Dde*I site. *Bam*HI (B) cuts specifically in C57BL/6-derived RT–PCR products whereas *Dde*I (D) cuts only products obtained from the SD7 allele. Control digestions of RT–PCR products derived from the parental strains (SD7, upper lane, labelled D; C57BL/6, lower lane, labelled B; U, undigested RT–PCR product) are shown on the left. No RT–PCR products were obtained from placentae (12.5 d.p.c.). Restriction fragments were separated on 2% agarose gels, stained with Sybr Green I (FMC BioProducts, Vallensbaek Strand, Denmark) and scanned on a Fluoroimager.

biased, i.e. strain-specific, amplification was ruled out by control experiments using a 1:1 mixture of cDNAs of both parental strains (data not shown). *Tssc4* was also biallelically expressed in embryos and neonatal tissues. However, we observed a significant bias towards the maternal allele in placentae at 12.5 and 16.5 d.p.c. for both *Tssc4* transcripts (Fig. 6). This indicates that *Tssc4* is imprinted in the placenta.

Imprinted genes are associated with CpG islands

It was remarkable that the non-imprinted *Tssc6*/*TSSC6* genes and the non-imprinted mouse *Ltrpc5* gene did not have CpG islands, whereas the imprinted human *LTRPC5* did have an island and so did all other imprinted genes in the cluster. We therefore decided to compare CpG island density in the cluster with the reported averages in the mouse and human genomes (61), to the density in other non-imprinted gene-rich genomic regions (53,54) and to the frequency of CpG islands in all imprinted genes (http://www.mgu.har.mrc.ac.uk/imprinting/imptables.html#impgene).

First, the CpG island per gene ratio in the cluster (86% in human and 66% in mouse) was significantly above average (56% in human and 47% in mouse; P = <0.01 and 0.001, respectively). Second, when comparing two randomly selected gene-rich regions, one on human chromosome 12p13/mouse

chromosome 6 (12p/6), the other the X-linked BTK locus (BTK), the CpG island per gene ratios were found to be even lower than the genome averages, with 53 (12p/6) and 25% (BTK) for human and 41 (12p/6) and 25% (BTK) for mouse, respectively.

The CpG dinucleotide content in the imprinted and non-imprinted regions was, however, not significantly different (imprinting cluster, human 2.4%, mouse 1.8%; 12p/6 region, human 2.1%, mouse 1.7%). Third, of all imprinted genes in mouse (both those which are in clusters and those outside clusters) for which information was available, 88% had CpG islands. These results demonstrate that one of the key features of imprinted genes is their association with CpG islands.

DISCUSSION

The determination of a 250 kb sequence between the imprinted *Kcnq1* and *Mash2* genes in the central part of the BWS region in the mouse offered the opportunity to compare the structural organization of six genes within the BWS gene cluster in two mammalian species, as the human DNA sequence was already available in the GenBank database. The purpose of the comparative sequencing effort was to obtain information on sequence conservation and on the conservation of gene regulation and imprinting. This enhances our understanding of the genetic and epigenetic control mechanisms in the whole region and of pathological mechanisms leading to BWS in the human.

The key findings of this study are, first, that the organization of the human and mouse regions is highly conserved except for two large insertions in the human sequence comprising repetitive sequences and a homologue of the ribosomal *RPL26* gene. Second, whereas some genes in this region such as *Tssc6*/*TSSC6* apparently escape imprinting in mouse and human, others like *Tssc4*/*TSSC4* are either imprinted in mouse placenta but not in the human or, conversely, like *Ltrpc5*/*LTRPC5* are not imprinted in the mouse but are in the human. Hence the central region of the cluster shows considerable variability in imprinting patterns between the two species. Lastly, imprinted genes in the cluster and elsewhere in the genome have a CpG island frequency which is remarkably above genome average, indicating that CpG islands may be necessary for imprinted gene silencing.

The chromosomal organization of the region in both species exhibits nearly complete conservation as all orthologues (*Kcnq1*/*KCNQ1*, *Ltrpc5*/*LTRPC5*, *Tssc4*/*TSSC4*, *Cd81*/*CD81*, *Tssc6*/*TSSC6* and *Mash2*/*ASCL2*) have the same relative order and identical transcriptional orientation in mouse and human. The only major difference is the presence of an additional copy of the *RPL26* gene that codes for a full-length ribosomal-like protein in the human and resides between *TSSC6* and *CD81* in the middle of a large insertion of repetitive elements. It remains unclear whether the unique presence of this gene in the human sequence reflects a recent evolutionary event of a *de novo* insertion or whether *RPL26* has been deleted in the mouse. Its presence within a cluster of repetitive sequences favours the first explanation.

With the exception of the two large insertions in the human sequence the distribution of repetitive elements, like LINE/L1, SINE/B1 and ALU/Alu elements and microsatellites, is similar in both organisms (Fig. 1). Major clusters of these repetitive elements are present 5′ of *Cd81*/*CD81* and *Mash2*/*ASCL2* in human and mouse, minor clusters in the upstream region of *Kcnq1*/*KCNQ1* and 3′ of *Mash2*/*ASCL2*. This conserved distribution of major clusters of repetitive elements and the presence of strong predicted MARs in the 5′ region of *Kcnq1*/*KCNQ1* and around *Tssc6*/*TSSC6* may be a reflection of a structural subdivision of the region (Fig. 1). Thus, there appear to be two peripheral domains with stably maintained and conserved imprinting patterns. The first domain is demarcated by *Tssc3*/*TSSC3* and *Kcnq1*/*KCNQ1* and the second by *Mash2*/*ASCL2* and *H19*. In contrast, the central domain is characterized by variable, relaxed or lack of imprinting and is located between *Ltrpc5*/*LTRPC5* and *Tssc6*/*TSSC6*. The functional analysis of human and mouse genes in this region also supports this notion (19,52,55). Examples of relaxation of imprinting are the mouse *Cd81* gene (19) and the *Tssc6* and *Ltrpc5* genes identified here, which were found to be biallelically expressed during mid-gestation and late-stage embryonic development as well as after birth. It remains open whether *Tssc6* and *Ltrpc5* are imprinted at earlier developmental stages (i.e. before 12.5 d.p.c.), as observed for *Cd81* (19). In contrast, all other genes in the cluster are either not subject to relaxation of imprinting, such as *Cdkn1c*, *Mash2*, *Igf2* and *H19* (22,25,26,28), or, as in the cases of *Tssc5* and *Kcnq1*, the onset of relaxation is much later, around the time of birth (17,19).

Examples of species-specific variation in imprinting are *Tssc4*/*TSSC4* and *Ltrpc5*/*LTRPC5*. Whereas *Tssc4* was biallelically transcribed in fetal tissues, it was maternally expressed in placenta. These results indicate that *Tssc4*/*TSSC4* is imprinted in mouse but not in human (52). We note that imprinting of mouse *Tssc3* and *Tssc5* is also most pronounced in extraembryonic tissues (14,15). The other genes showing species-specific differences in imprinted gene expression are *Ltrpc5*/*LTRPC5*. *Ltrpc5* was not imprinted in the mouse but human *LTRPC5* showed exclusive paternal expression in monochromosomal hybrid cell lines (55). Imprinting in normal somatic tissues, however, as well as during embryonic stages, has not been analysed. Lack of conservation of imprinting between human and mouse has also been described for the *Igf2r*/*IGF2R* gene (62), which is imprinted in mouse but not consistently imprinted in the human.

Because of the relaxation in the centre of the imprinting cluster, previously postulated boundaries of the imprinting cluster need to be treated with caution as these are based only on the biallelic expression of *Nap1l4* (17), upstream of *Tssc3*, and *Nctc1* and *Rpl23l*, downstream of *H19* (41,42). It cannot be excluded that more imprinted genes are present in the vicinity of the cluster. Furthermore, although most of our own and previous results (52,55) suggest that the whole imprinting cluster can be separated into (at least) three subdomains with different characteristics, such a division does not exclude the possibility that some functional aspects, like the germline establishment or maintenance of imprinting, are controlled at the level of the cluster as a whole. This view is supported by the observation that translocations in *KVLQT1* in BWS patients are associated with loss of imprinting in *IGF2* (39,51) and that at least some BWS patients show loss of imprinting of both *KCNQ1OT1* (*LIT1*) and *IGF2* (37,39).

One of the structural features investigated, the distribution and quality of CpG islands, seems to integrate the division into subdomains and gene-specific effects with the general aspects of imprinting in the cluster. The comparison of imprinting

patterns in the cluster with the distribution and characteristics of CpG islands thus revealed intriguing associations. *Cdkn1*/*CDNK1C* and *Mash2*/*ASCL2* (26–30) are tightly imprinted and are linked to the most pronounced CpG islands in this region. In the region between *Kcnq1* and *Mash2* there seems to be a connection between the quality (the length) of CpG islands and the timing of relaxation or the extent of imprinting. *Kcnq1* is biallelically expressed after birth and possesses a more pronounced CpG island than *Tssc4*, which is already biallelically expressed at 12.5 d.p.c. in the embryo and only imprinted in the mouse placenta. The non-imprinted *Tssc6* and *Ltrpc5* do not have CpG islands, whereas the imprinted human *LTRPC5* gene (55) has one. The only exception to this rather consistent picture is the human *TSSC4* gene which, while still maintaining a CpG island, is not imprinted. Hence, although CpG islands and their quality may be linked to the extent of imprinted expression, their presence seems to be necessary but not sufficient to establish/maintain imprinting. However, the functional association of CpG islands with imprinted expression is further strengthened by the observation that regions that have a similar gene density but are not imprinted have a significantly lower density of CpG islands. Hence we suggest that establishment and stable maintenance of imprinting patterns in mammalian genes requires the presence of CpG islands and is possibly influenced by their quality.

The conservation of major parts of the genomic sequence and the pronounced relaxation of imprinting in this region in human and mouse indicate that the mouse will serve as a very useful model to study general mechanisms of imprinting and the aetiology of BWS. Systematic comparative analysis should now be extended to the whole cluster and to the study of epigenetic modifications throughout the cluster region using new functional genomics technologies.

MATERIALS AND METHODS

DNA sequencing

DNA of BAC clone 300P2 (CITB Library, Research Genetics, http://resgen.com ; B. Birren *et al.*, unpublished data) was prepared according to a standard protocol (63) which included two CsCl gradient ultracentrifugations. Aliquots of 10 µg DNA were sonicated for 15 s using a cell disruptor B-30 (output control setting 5; Branson Sonic Power, Danbury, CT). After treatment with Klenow and T4 DNA polymerases (MBI Fermentas, Vilnius, Lithuania) 1.0–1.5 kb long DNA fragments were purified from 1% agarose gels (1× TBE) using the QiaexII kit (Qiagen, Hilden, Germany), ligated into *Sma*I-digested pUC19 (MBI Fermentas) and introduced into *Escherichia coli* SURE cells (Stratagene, La Jolla, CA) according to standard protocols.

In order to obtain on average a 5- to 6-fold sequence coverage of the entire BAC insert, the inserts of 3042 clones were unidirectionally sequenced on ABI 377 automatic sequencers using standard protocols. The sequences were assembled using Gap4 software (Sanger Centre, http://sanger.ac.uk/Software). Sequence gaps were filled by reverse sequencing of individual clones or by sequencing PCR products obtained by amplification with oligonucleotides situated on both sides of the gaps. Finally, 4205 sequences were assembled; per consensus character 6.6 average reading characters were obtained.

Sequence analysis

The human and the mouse sequences were compared by dot plot analysis using PipMaker (Pennsylvania State University, http://globin.cse.psu.edu/cgi-bin/pipmaker) and the software package MacMolly (match length 35, mismatches 12, gap penalty 3, mismatch penalty 1). Exon prediction was performed with GRAIL (Oak Ridge National Laboratory, http://compbio.ornl.gov), Genscan (Massachusetts Institute of Technology, http://CCR-081.mit.edu/GENSCAN.html) and Genefinder (Sanger Centre, http://genomic.sanger.ac.uk/gf/gf.shtml) software. Searches for homologous sequences were performed in the GenBank databases for non-redundant sequences and ESTs (National Center for Biotechnology Information, http://www.ncbi.nlm.nih.gov) and in the TIGR database for mouse EST clusters (Institute for Genomic Research, http://www.tigr.org). Repetitive elements were detected and masked using Censor software (http://charon.girinst.org/~server/censor.html). CpG and G+C distribution were obtained using the 'window' function of the GCG package (window size 500, shift increment 3; Genetics Computer Group, Madison, WI) (64). Additionally, CpG islands were detected with the 'CpG islands' search (input score threshold 20; http://www2.ebi.ac.uk/cpg/). The base content of sequences was determined using 'composition' of the GCG package and tandem repeats with motifs longer than 6 bp were detected with the 'compare' function of the same software package. The searches for MARs were performed using MarFinder (http://www.ncgr.org/MarFinder).

RNA preparations

Total RNAs of various mouse tissues and developmental stages were prepared according to a standard protocol (65). Total RNA was randomly primed and reverse transcribed using reverse transcriptase (Promega, Madison, WI) according to the manufacturer's protocol.

RT–PCR analysis

RT–PCRs were carried out on randomly primed cDNA on Perkin Elmer Thermocyclers (Perkin Elmer, Norwalk, CT) under standardized conditions (0.5 µM each primer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs). A standard amplification protocol was as follows: denaturation at 95°C for 3 min followed by 35 cycles of 94°C at 30 s, 30 s annealing at 53–68°C (for temperatures see Table 2) and 1 min at 75°C and a final elongation at 72°C for 5 min.

Transcription of both *Tssc4* transcripts (GenBank accession no. AJ279796) was verified by RT–PCR using primers 1 and 2. For *Tssc6* overlapping RT–PCR products were generated using the primer pairs 3 + 4 (exons 1–4), 5 + 6 (exons 4–7) and 7 + 8 (exons 7–8).

Overlapping RT–PCR products of *Ltrpc5* were generated using the following primer combinations: 9 + 10 (exons 1–2), 11 + 12 (exons 2–4), 13 + 14 (exons 4–8), 15 + 16 (exons 8–9), 17 + 18 (exons 9–13), 19 + 20 (exons 13–16), 21 + 22 (exons 16–19), 23 + 24 (exons 19–21) and 25 + 26 (exons 21–22).

Table 2. Sequences of primers used for (RT–)PCR amplifications and 3′-RAC

Primer	Sequence	Annealing temperature (°C)
1	AAC CCG AGC GAG ACT GGA CAG GAC C	60
2	GCT TCT TGC TGC CAT GGA AAC CAA CA	60
3	CCT AAG AAG CCC AGA CCC TA	55
4	CTG CA CCA GTT GCT AGA TG	56
5	AGG ATG CGG TGT TGG ACA C	56
6	GGT GTC AGG GAG TAT CTG CC	53
7	GTC ACA GTC CAG CAT CTC CC	55
8	TAT GAT GCT CTG TGC CTT CC	56
9	AGG GTC TGG GAA GAA GCG A	58
10	GAC ATC CCG AAG CCA CGA CT	58
11	TTT TTG AAC TCC TGC TCA CCG A	58
12	TCG TTC CCA CTC CCA AGG GCG CCT GAC	58
13	ACA GCA ATC TCT CCC ACT TGA TCC	65
14	AGA TTT CAC TCT TGG CAA TGT C	65
15	ACT ACC TAG ATG AGC TCA AGT TAG CAG	58
16	TCC TGG TAG AAG CCA CGG CA	58
17	TCT TGA CCT ATG GGC GGC TG	65
18	ATG AGG TTT GTG TAG ATG AGG G	65
19	ACC CAT CCT ACG GCT TCT G	58
20	GTT CCA GTT GTC TTC CAC ATA C	58
21	TCC TGT TCA TTG TGG GAG TCA C	60
22	TGG CGA TCA GAA GGT TCA TG	60
23	TGC TGG AAA GCT CGG CTT CCT GCC CT	65
24	TCG CTG TCC CTC CTC CGT TTC TCC A	65
25	TTG GAC CAG AAG ATC ATT ACC TGG GAA	63
26	CTA ATG CTA CCC ACC AGG ATG CTG GTA CAG	63
27	GGC CAA GTC CTA CAT AGG ACA CCA TCT GCC	68
28	AAC CCG AGC GAG ACT GGA CAG GAC C	58
29	GCT TCT TGC TGC CAT GGA AAC CAA CA	58
30	CCC AAG TCA GAC TCC AGC AAG GG	55
31	AAA GCC ACA CAA TTC AAG AA	55
32	CTG TGC GCG CTA GGA TAC CCT TTA A	58
33	AGG TCC AGC GAG ATG GGC TAG CTC C	58
34	GCT CCC CAA ACC AGT GCC CC	60
35	AAA GGC CTT CGA GGT CCC CTG	60
36	GTC ACA GTC CAG CAT CTC CC	56

(RT–)PCR was carried out as described in Materials and Methods.

The obtained RT–PCR products were cloned into vector pGEM-T (Promega) and sequenced using standard protocols. The 3′-RACE of *Ltrpc5* was performed with primer 27 using the SMART RACE cDNA amplification kit from Clontech (Heidelberg, Germany). The products were cloned and sequenced.

Imprinting analysis

Strain-specific sequence polymorphisms were detected by sequencing PCR products derived from *Mus mus domesticus* C57BL/6 and *M.spretus* DNA templates, respectively. For *Tssc4* PCRs were performed on *M.m.domesticus* C57BL/6 and *M.spretus* DNAs using the primer pairs 28 + 29 and 30 + 31. For *Tssc6* (GenBank accession no. AJ279791) and *Ltrpc5* (GenBank accession no. AJ271092) RT–PCR products were amplified and sequenced as described above using *M.m.domesticus* C57BL/6 and SD7 cDNAs.

For the subsequent imprinting analysis of *Tssc4-1* (GenBank accession no. AJ279796) RT–PCR was performed using primers 32 and 33. The 811 bp long RT–PCR products obtained were purified from a 1.5 % agarose gel (1× TBE) using the QiaexII kit and cut with *Tai*I. In SD7-derived RT–PCR products a *Tai*I site is present at nucleotide 126 932 (GenBank accession no. AJ251835) generated by an A→G transition, therefore *Tai*I digestion results in 298 and 513 bp DNA fragments specific for SD7.

The imprinting status of *Tssc4-2* (GenBank accession no. AJ279797) was analysed using primer pair 34 + 35. The 268 bp long RT–PCR products were gel purified from a 1.5% agarose gel (1× TBE) and cut with *Alu*I, at an SD7-specific restriction site generated by a G→T transition at nucleotide 126 351 (GenBank accession no. AJ251835). The obtained SD7-specific restriction fragments were 215 and 53 bp long.

For *Tssc6* RT–PCR was performed with primer pair 36 + 37. The 756 bp long product obtained was gel purified and cut with *Tai*I at an SD7-specific site, generated by an A→G exchange at nucleotide 178 113 (GenBank accession no. AJ251835). This resulted in 128 and 628 bp long DNA fragments specific for the SD7 allele.

Ltrpc5 RT–PCR products were amplified with primer pair 21 + 22. The imprinting status of *Ltrpc5* was analysed using a strain-specific polymorphism at nucleotide 117 174 (GenBank accession no. AJ251835) which results in a *Bam*HI restriction site in *M.m.domesticus* C57BL/6 and a *Dde*I restriction site in SD7.

ACKNOWLEDGEMENTS

We would like to thank the members of the MPI-MG sequencing department, in particular Sven Klages, Katja Heitmann and Katja Borzym, for their technical support, Juliane Ramser and Steffen Hennig for help with database searches and technical tips, C. Harteneck for helpful comments on the LTRPC family, Angelika Daser for her support, and Patricia Ruiz and Lucy Bowden for technical help. We also gratefully acknowledge Prof. T.A. Trautners support for this work and many helpful discussions. Part of the work was supported by the Deutsche Forschungsgemeinschaft (WA1031-2), HFSP (RG0088/99), BBSRC, MRC and CRC.

REFERENCES

1. Reik, W. and Walter, J. (1998) Imprinting mechanisms in mammals. *Curr. Opin. Genet. Dev.*, **8**, 154–164.
2. Feil, R. and Khosla, S. (1999) Genomic imprinting in mammals: an interplay between chromatin and DNA methylation? *Trends Genet.*, **15**, 431–435.
3. Brannan, C.I. and Bartolomei, M.S. (1999) Mechanisms of genomic imprinting. *Curr. Opin. Genet. Dev.*, **9**, 164–170.

4. Tilghman, S.M. (1999) The sins of the fathers and mothers: genomic imprinting in mammalian development. *Cell*, **96**, 185–193.
5. Barlow, D.P. (1997) Competition—a common motif for the imprinting mechanism? *EMBO J.*, **16**, 6899–6905.
6. Surani, M.A. (1998) Imprinting and the initiation of gene silencing in the germ line. *Cell*, **93**, 309–312.
7. Constancia, M., Pickard, B., Kelsey, G. and Reik, W. (1998) Imprinting mechanisms. *Genome Res.*, **8**, 881–900.
8. Nicholls, R.D., Saitoh, S. and Horsthemke, B. (1998) Imprinting in Prader-Willi and Angelman syndromes. *Trends Genet.*, **14**, 194–200.
9. Reik, W. and Maher, E.R. (1997) Imprinting in clusters: lessons from Beckwith-Wiedemann syndrome. *Trends Genet.*, **13**, 330–334.
10. Kitsberg, D., Selig, S., Brandeis, M., Simon, I., Keshet, I., Driscoll, D.J., Nicholls, R.D. and Cedar, H. (1993) Allele-specific replication timing of imprinted regions. *Nature*, **364**, 459–463.
11. Knoll, J.H.M., Cheng, S.D. and Lalande, M. (1994) Allele specificity of DNA replication timing in Angelman/Prader-Willi syndrome imprinted chromosomal region. *Nature Genet.*, **6**, 41–46.
12. Pàldi, A., Gyapay, G. and Jami, J. (1995) Imprinted chromosomal regions of the human genome display sex-specific meiotic recombination frequencies. *Curr. Biol.*, **5**, 1030–1035.
13. Robinson, W.P. and Lalande, M. (1995) Sex-specific meiotic recombination in the Prader-Willi/Angelman syndrome imprinted region. *Hum. Mol. Genet.*, **4**, 801–806.
14. Qian, N., Frank, D., O'Keefe, D., Dao, D., Zhao, L., Yuan, L., Wang, Q., Keating, M., Walsh, C. and Tycko, B. (1997) The *IPL* gene on chromosome 11p15.5 is imprinted in humans and mice and is similar to *TDAG51*, implicated in Fas expression and apoptosis. *Hum. Mol. Genet.*, **6**, 2021–2029.
15. Dao, D., Frank, D., Qian, N., O'Keefe, D., Vosatka, R.J., Walsh, C.P. and Tycko, B. (1998) *IMPT1*, an imprinted gene similar to polyspecific transporter and multi-drug genes. *Hum. Mol. Genet.*, **7**, 597–608.
16. Cooper, P.R., Smilinich, N.J., Day, C.D., Nowak, N.J., Reid, L.H., Pearsall, R.S., Reece, M., Prawitt, D., Landers, J., Housman, D.E. *et al.* (1998) Divergently transcribed overlapping genes expressed in liver and kidney and located in the 11p15.5 imprinted domain. *Genomics*, **49**, 38–51.
17. Paulsen, M., Davies, K.R., Bowden, L.M., Villar, A.J., Franck, O., Fuermann, M., Dean, W., Moore, T.F., Rodrigues, N., Davies, K.E. *et al.* (1998) Syntenic organization of the mouse distal chromosome 7 imprinting cluster and the Beckwith-Wiedemann syndrome region in chromosome 11p15.5. *Hum. Mol. Genet.*, **7**, 1149–1159.
18. Gould, T.D. and Pfeifer, K. (1998) Imprinting of mouse *Kvlqt1* is developmentally regulated. *Hum Mol. Genet.*, **7**, 483–487.
19. Caspary, T., Cleary, M.A., Baker, C.C., Guan, X.-J. and Tilghman, S.M. (1998) Multiple mechanisms regulate imprinting of the mouse distal chromosome 7 gene cluster. *Mol. Cell. Biol.*, **18**, 3466–3474.
20. Ohlsson, R., Nystrom, A., Pfeifer-Ohlsson, S., Tohonon, V., Hedbourg, F., Schofield, P., Flam, F. and Ekström, T.J. (1993) *IGF2* is parentally imprinted during human embryogenesis and in the Beckwith-Wiedemann syndrome. *Nature Genet.*, **4**, 94–97.
21. Giannoukakis, N., Deal, C., Paquette, J., Goodyer, C.G. and Polychronakis, C. (1993) Parental genomic imprinting of the human *IGF2* gene. *Nature Genet.*, **4**, 98–101.
22. Ferguson-Smith, A.C., Cattanach, B.M., Barton, S.C., Beechey, C.V. and Surani, M.A. (1991) Embryological and molecular investigations of parental imprinting on mouse chromosome 7. *Nature*, **351**, 667–670.
23. Deltour, L., Montagutelli, X., Guenet, J.L., Jami, J. and Pàldi, A. (1995) Tissue- and developmental stage-specific imprinting of the mouse proinsulin gene, *Ins2*. *Dev. Biol.*, **168**, 686–688.
24. Rainier, S., Johnson, L.A., Dobry, C.J., Ping, A.J., Grundy, P.E. and Feinberg, A.P. (1993) Relaxation of the imprinted genes in human cancer. *Nature*, **362**, 747–749.
25. Bartolomei, M.S., Zemel, S. and Tilghman, S.M. (1991) Parental imprinting of the mouse H19 gene. *Nature*, **351**, 153–155.
26. Guillemot, F., Caspary, T., Tilghman, S.M., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Anderson, D.J., Joyner, A.L., Rossant, J. and Nagy, A. (1995) Genomic imprinting of *Mash2*, a mouse gene required for trophoblast development. *Nature Genet.*, **9**, 235–242.
27. Alders, M., Hodges, M., Hadjantonakis, A.-K., Postmus, J., van Wijk, I., Bliek, J., de Meulemeester, M., Westerveld, A., Guillemot, F., Oudejans, C. *et al*. (1997) The human *Achaete-Scute* homologue 2 (*ASCL2, HASH2*) maps to chromosome 11p15.5, close to *IGF2* and is expressed in extravillus trophoblasts. *Hum. Mol. Genet.*, **6**, 859–867.
28. Hatada, I. and Mukai, T. (1995) Genomic imprinting of $p57^{KI}$P2, a cyclin-dependent kinase inhibitor, in mouse. *Nature Genet.*, **11**, 204–206.
29. Hatada, I., Ohashi, H., Fukushima, Y., Kaneko, Y., Inoue, M., Komoto, Y., Okada, A., Ohishi, S., Nabetani, A., Morisaki, H. *et al*. (1996) An imprinted gene $p57^{KIP2}$ is mutated in Beckwith-Wiedemann syndrome. *Nature Genet.*, **14**, 171–173.
30. Matsuoka, S., Thompson, J.S., Edwards, M.C., Barletta, 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, $p57^{KI}$P2, on chromosome 11p15. *Proc. Natl Acad. Sci. USA*, **93**, 3026–3030.
31. Hoovers, J.M.N., Kalkin, L.M., Johnson, L.A., Alders, M., Redeker, B., Law, D.J., Bliek, J., Steenman, M., Benedict, M., Wiegant, J. *et al*. (1995) Multiple genetic loci within 11p15 defined by Beckwith-Wiedemann syndrome rearrangement breakpoints and subchromosomal transferable fragments. *Proc. Natl Acad. Sci. USA*, **92**, 12456–12460.
32. Lee, M.P., Hu, R.-J., Johnson, L.A. and Feinberg, A.P. (1997) Human *KVLQT1* gene shows tissue-specific imprinting and encompasses Beckwith-Wiedemann syndrome chromosomal rearrangements. *Nature Genet.*, **15**, 181–185.
33. Reid, L.H., Davies, C., Cooper, P.R., Crider-Miller, S.J., Sait, S.N.J., Nowak, N.J., Evans, G., Stanbridge, E.J., deJong, P., Shows, T.B. *et al*. (1997) A 1-Mb physical map and PAC contig of the imprinted domain in 11p15.5 that contains TAPA1 and the BWSCR1/WT2 region. *Genomics*, **43**, 366–375.
34. Andria, M.L., Hsieh, C.-L., Oren, R., Francke, U. and Levy, S. (1991) Genomic organization and chromosomal localization of the *TAPA-1* gene. *J. Immunol.*, **147**, 1030–1036.
35. Moore, T., Constancia, M., Zubair, M., Bailleul, B., Feil, R., Sasaki, H. and Reik, W. (1997) Multiple imprinted sense and antisense transcripts, differential methylation and tandem repeats in a putative imprinting control region upstream of mouse *Igf2*. *Proc. Natl Acad. Sci. USA*, **94**, 12509–12514.
36. Crider-Miller, S.J., Reid, L.H., Higgins, M.J., Nowak, N.J., Shows, T.B., Futreal, P.A. and Weissman, B.E. (1997) Novel transcribed sequences within the BWS/WT2 region in 11p15.5: tissue-specific expression correlates with cancer type. *Genomics*, **46**, 355–363.
37. Lee, M.P., DeBaun, M.R., Mitsuya, K., Galonek, H.L., Brandenburg, S., Oshimura, M. and Feinberg, A.P. (1999) Loss of imprinting of a paternally expressed transcript, with antisense orientation to *KvLQT1*, occurs frequently in Beckwith-Wiedemann syndrome and is independent of insulin-like growth factor II imprinting. *Proc. Natl Acad. Sci. USA*, **96**, 5203–5208.
38. Mitsuya, K., Meguro, M., Lee, M.P., Katoh, M., Schulz, T.C., Kugoh, H., Yoshida, M.A., Nikawa, N., Feinberg, A.P. and Oshimura, M. (1999) *LIT1*, an imprinted antisense RNA in the human *KvLQT1* locus identified by screening for differentially expressed transcripts using monochromosomal hybrids. *Hum. Mol. Genet.*, **8**, 1209–1217.
39. Smilinich, N.J., Day, C.D., Fitzpatrick, G.V., Caldwell, G.M., Lossie, A.C., Cooper, P.R., Smallwood, A.C., Joyce, J.A., Schofield, P.N., Reik, W. *et al*. (1999) A maternally methylated CpG island in *KvLQT1* is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc. Natl Acad. Sci. USA*, **96**, 8064–8069.
40. Hu, R.-J., Lee, M.P., Johnson, L.A. and Feinberg, A.P. (1996) A novel human homologue of yeast nucleosome assembly protein, 65 kb centromeric to the $p57^{KIP2}$ gene, is biallelically expressed in fetal and adult tissues. *Hum. Mol. Genet.*, **5**, 1743–1748.
41. Tsang, P., Gilles, F., Yuan, L., Kuo, Y.-H., Lupu, F., Samara, G., Moosikasuwan, J., Goye, A., Zelenetz, A.D., Selleri, L. and Tycko, B. (1995) A novel L23-related gene 40 kb downstream of the imprinted *H19* gene is biallelically expressed in mid-fetal and adult tissues. *Hum. Mol. Genet.*, **4**, 1499–1507.
42. Ishihara, K., Kato, R., Furuumi, H., Zubair, M. and Sasaki, H. (1998) Sequence of a 42-kb mouse region containing the imprinted *H19* locus: identification of a novel muscle-specific transcription unit showing biallelic expression. *Mamm. Genome*, **9**, 775–777.
43. Leighton, P.A., Ingram, R.S., Eggenschwiler, J., Efstratiadis, A. and Tilghman, S.M. (1995) Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature*, **375**, 34–39.
44. Steenman, M.J., Rainier, S., Dobry, C.J., Grundy, P., Horon, I.L. and Feinberg, A.P. (1994) Loss of imprinting of IGF2 is linked to reduced expression and abnormal methylation of H19 in Wilms' tumour. *Nature Genet.*, **7**, 433–439.
45. Moulton, T., Crenshaw, T., Hao, Y., Moosikasuwan, J., Lin, N., Dembitzer, F., Hensle, T., Weiss, L., McMorrow, L., Loew, T. *et al.*

(1994) Epigenetic lesions at the H19 locus in Wilms' tumour patients. *Nature Genet.*, **7**, 440–447.
46. Reik, W., Brown, K.W., Schneid, H., Le Bouc, Y., Bickmore, W. and Maher, E.R. (1995) Imprinting mutations in the Beckwith-Wiedemann syndrome suggested by altered imprinting pattern in the IGF2-H19 domain. *Hum. Mol. Genet.*, **4**, 2379–2385.
47. O'Keefe, D., Dao, D., Zhao, L., Sanderson, R., Warburton, D., Weiss, L., Anyane-Yeboa, K. and Tycko, B. (1997) Coding mutations in $p57^{KIP2}$ are present in some cases of Beckwith-Wiedemann syndrome but are rare or absent in Wilms tumours. *Am. J. Hum. Genet.*, **61**, 295–303.
48. Lee, M.P., DeBaun, M., Randhawa, G., Reichard, B.A., Elledge, S.J. and Feinberg, A.P. (1997) Low frequency of p57(KIP2) mutation in Beckwith-Wiedemann syndrome. *Am. J. Hum. Genet.*, **61**, 304–309.
49. Lam, W.W., Hatada, I., Ohishi, S., Mukai, T., Joyce, J.A., Cole, T.R., Donnai, D., Reik, W., Schofield, P.N. and Maher, E.R. (1999) Analysis of germline CDKN1C (p57KIP2) mutations in familial and sporadic Beckwith-Wiedemann syndrome (BWS) provides a novel genotype-phenotype correlation. *J. Med. Genet.*, **36**, 518–523.
50. Weksberg, R., Shen, D.R., Fei, Y.L., Song, Q.L. and Squire, J. (1993) Disruption of insulin-like growth factor 2 imprinting in Beckwith-Wiedemann syndrome. *Nature Genet.*, **5**, 143–150.
51. Brown, K.W., Villar, A.J., Bickmore, W., Clayton-Smith, J., Catchpoole, D., Maher, E.R. and Reik, W. (1996) Imprinting mutation in the Beckwith-Wiedemann syndrome leads to biallelic *IGF2* expression through an *H19*-independent pathway. *Hum. Mol. Genet.*, **5**, 2027–2032.
52. Lee, M.P., Brandenburg, S., Landes, G.M., Adams, M., Miller, G. and Feinberg, A.P. (1999) Two novel genes in the center of the 11p15.5 imprinted domain escape genomic imprinting. *Hum. Mol. Genet.*, **8**, 683–690.
53. Oeltjen, J.C., Malley, T.M., Muzny, D.M., Miller, W., Gibbs, R.A. and Belmont, J.W. (1997) Large-scale comparative sequence analysis of the human and murine Bruton's tyrosine kinase loci reveals conserved regulatory domains. *Genome Res.*, **7**, 315–329.
54. Ansari-Lari, M.A., Oeltjen, J.C., Schwartz, S., Zhang, Z., Muzny, D.M., Lu, J., Gorrell, J.H., Chinault, A.C., Belmont, J.W., Miller, W. and Gibbs, R.A. (1998) Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6. *Genome Res.*, **8**, 29–40.
55. Prawitt, D., Enklaar, T., Klemm, G., Gärtner, B., Spangenberg, C., Winterpacht, A., Higgins, M., Pelletier, J. and Zabel, B. (2000) Identification and characterization of *MTR1*, a novel gene with homology to melastatin (*MLSN1*) and the *trp* gene family located in the BWS-WT2 critical region on chromosome 11p15.5 and showing allele-specific expression. *Hum. Mol. Genet.*, **9**, 203–216.
56. Kenmochi, N., Kawaguchi, T., Rozen, S., Davis, E., Goodman, N., Hudson, T.J., Tanaka, T. and Page, D.C. (1998) A map of 75 human ribosomal protein genes. *Genome Res.*, **8**, 509–523.
57. Neumann, B., Kubicka, P. and Barlow, D.P. (1995) Characteristics of imprinted genes. *Nature Genet.*, **9**, 12–13.
58. Tanaka, M., Puchyr, M., Gertsenstein, M., Harpal, K., Jaenisch, R., Rossant, J. and Nagy, A. (1999) Parental origin-specific expression of *Mash2* is established at the time of implantation with its imprinting mechanism highly resistant to genome-wide demethylation. *Mech. Dev.*, **87**, 129–142.
59. Harteneck, C., Plant, T.D. and Schultz, G. (2000) From worm to man: three subfamilies of TRP channels. *Trends Neurosci.*, **23**, 159–166.
60. Hemberger, M., Redies, C., Krause, R., Oswald, J., Walter, J. and Fundele, R.H. (1998) *H19* and *Igf2* are expressed and differentially imprinted in neuroectoderm-derived cells in the mouse brain. *Dev. Genes Evol.*, **208**, 393–402.
61. Antequera, F. and Bird, A. (1993) Number of CpG islands and genes in human and mouse. *Proc. Natl Acad. Sci. USA*, **90**, 11995–11999.
62. 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. *Nature Genet.*, **5**, 74–78.
63. Sambrook, J., Fritsch, E. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
64. Devereux, J., Haeberli, P. and Smithies, O. (1984) A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.*, **12**, 387–395.
65. Chomczynski, P. and Sacchi, N. (1987) Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.*, **162**, 156–159.