UM 9(5)h and UM 9(5)p, human and porcine noncoding transcripts with preferential expression in the cerebellum

U. Michel, B. Kallmann, P. Rieckmann and D. Isbrandt

RNA 2002 8: 1538-1547

RNA (2002), *8*:1538–1547. Cambridge University Press. Printed in the USA.

DOI: 10.1017/S1355838202028042

UM 9(5)h and UM 9(5)p, human and porcine noncoding transcripts with preferential expression in the cerebellum

UWE MICHEL,[1] BORIS KALLMANN,[2] PETER RIECKMANN,[2] and DIRK ISBRANDT[3]

[1]Neurologische Klinik, 37075 Göttingen, Germany
[2]Neurologische Klinik im Kopfklinikum, Julius-Maximilians-University, 97080 Würzburg, Germany
[3]Zentrum für Molekulare Neurobiologie Hamburg, Institut für Neurale Signalverarbeitung, 20246 Hamburg, Germany

ABSTRACT

We compared the gene expression patterns of fetal and adult porcine brains and identified a sequence tag that was more abundant in adult than in fetal brain. The RNA corresponding to the sequence tag has the highest expression level in adult cerebellum. Lower expression levels of the transcript were found in adult cerebrum, pituitary, and uterus, as well as in fetal brain, heart, intestine, kidney, and liver. The sequence tag was used to screen a cDNA library from adult porcine brain. Two independent clones of 2,273 nt and 1,701 nt were isolated. The shorter cDNA is a 5′-truncated form of the longer clone, and both clones have almost identical sequences with multiple start and stop codons in all three reading frames. Screening of two different human brain cDNA libraries with porcine cDNA probes resulted in four overlapping cDNA fragments, which were assembled to one contig of 2,336 nt in length. Like noncoding RNAs, the porcine and human sequences have no common conserved open reading frame and share stretches of high homology interrupted by stretches with almost no homology. The human and porcine RNAs were named UM 9(5)h and UM 9(5)p, respectively. They are part of larger transcripts, which are transcribed from single-copy genes, they have very similar tissue distributions, and their sequences are colinear with the respective genomic fragment.

Keywords: brain; gene expression patterns; noncoding RNA

INTRODUCTION

Gene expression usually involves transcription and translation, but although translation is necessary for the expression of most genes, it is not mandatory for all of them. A large amount of RNA consists of untranslated transcripts generated by polymerase I (Pol I) and polymerase III (Pol III), such as ribosomal (rRNAs) or transfer RNAs (tRNAs), respectively, and only approximately 2.5% of the cellular RNA content are transcripts generated by RNA polymerase II (Pol II), that is, messenger RNAs (mRNAs). But even Pol II transcripts can be noncoding RNAs. In this case, the transcripts lack a translated open reading frame (ORF) and display their functions solely as RNA molecules. They are therefore sometimes referred to as "riboregulators" (Rastinejad & Blau, 1993).

Noncoding RNAs are found in all phyla, and in eukaryotic cells they can be nuclear, cytoplasmic, or localized to the mitochondrion (Michel, 2002). Their sizes differ from a few nucleotides (Lee et al., 1993) by up to more than 100 kb (Lyle et al., 2000), and most of them have neither a common sequence motif nor a common structure. Work within the last decade revealed that noncoding RNAs fulfill diverse functions. They were shown to control the expression of proteins (Lee et al., 1993), to be responsible for the subcellular localization of RNAs (Kloc et al., 1993), and to be involved in the dosage compensation (Brockdorff et al., 1992; Brown et al., 1992; Amrein & Axel, 1997; Meller et al., 1997) and imprinting of genes (Lyle et al., 2000). Furthermore, they can be essential for the control of meiotic cell division (Watanabe & Yamamoto, 1994), the enzymatic processing of RNAs (Chang & Clayton, 1989; Topper & Clayton, 1990; Chu et al., 1994), and the activation of steroid receptor-dependent gene expression (Lanz et al., 1999), and they can be hosts for small nucleolar RNAs (snoRNAs; Tycowski et al., 1996). Some noncoding RNAs such as DD3, SCA8, WT1, and DISC2 are associated with human diseases (Bussemakers et al., 1999; Koob et al., 1999; Malik et al., 2000; Millar et al., 2000). However, for most noncoding RNAs the function, mode of action, and even the complete se-

Reprint requests to: Uwe Michel, Neurologische Klinik, Robert-Koch-Str. 40, 37075 Göttingen, Germany; e-mail: umichel@gwdg.de.

quence and structure are still elusive (summarized in Michel, 2002).

Currently, there is no systematic search for noncoding RNAs and, therefore, most have been identified serendipitously as, for example, in screens of differential gene expression. We compared the gene expression patterns of fetal and adult porcine brains and identified a transcript without significant ORF that is preferentially expressed in adult cerebellum. Isolation and characterization of the human homolog revealed that the human and the porcine transcript have no common conserved ORF. Furthermore, the human and porcine transcripts are derived from single-copy genes, have the same tissue distribution, and are colinear to their respective genomic sequence.

RESULTS

We used mRNA differential display (DD) to compare the gene expression patterns of fetal and adult porcine brains to find genes regulated during the ontogeny of the brain. Some 17,500 bands of each tissue were examined. Those primer combinations that resulted in differentially expressed PCR products were used in two further rounds of DD with cDNA from independent RNA preparations. Forty PCR clones showed a reproducible differential expression in all three rounds of DD. The PCR products were cloned in vectors, and labeled cRNA probes derived from these clones were used for the measurement of RNA steady levels with S1 nuclease protection assays. Only five out of the 40 clones revealed a differential RNA expression between fetal and adult porcine brains. One clone was derived from a mitochondrial antisense transcript that is more abundant in fetal brain (Michel et al., 2000), two clones have yet to be characterized, and two clones were from the same transcript that is preferentially expressed in adult porcine brain. This latter transcript was named UM 9(5)p, and Figure 1 shows a comparison of the RNA expression levels of UM 9(5)p in fetal and adult porcine brains.

The DD clone had a size of 105 nt, with no significant sequence homology to sequences deposited in databases. Therefore, we constructed a cDNA library from adult porcine brain and screened 10^6 clones with a random-primed probe derived from the DD clone. Two independent clones of 1,701 and 2,273 nt in length were isolated. Sequencing revealed that the shorter clone is a 5′ truncated form of the larger one and that both have almost identical sequences. Figure 2 shows the sequence of the UM 9(5)p clones and the original DD product. Sequencing revealed that the DD product was amplified from the middle of the cDNA and not from the 3′ end. Again, no significant matches between the isolated clones and the sequences deposited in databases were found. Furthermore, none of the three theoretical reading frames of the UM 9(5)p clones contained a significant ORF. The expression of a possible

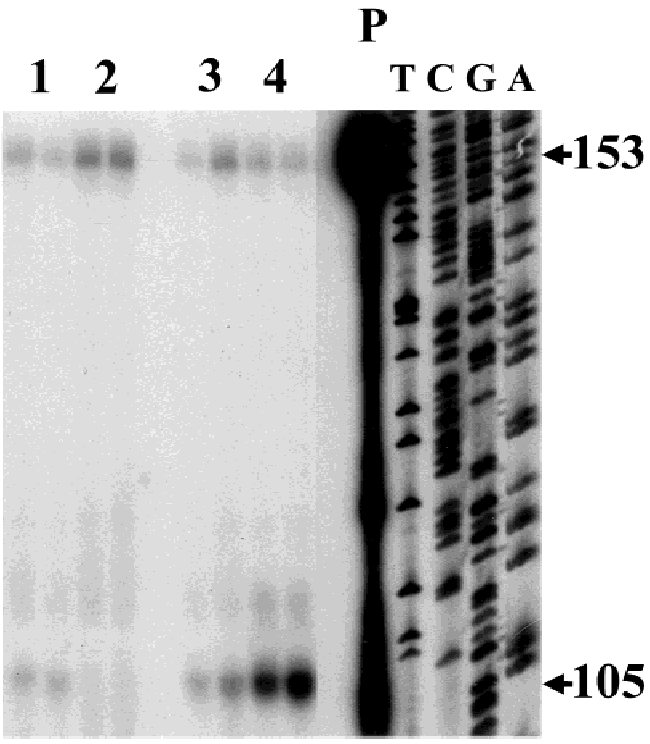

FIGURE 1. X-ray autoradiograph of a comparison of steady-state mRNA levels of UM 9(5)p in porcine brain. A cRNA probe derived from the original DD clone (P) was used to analyze UM 9(5)p RNA expression in fetal porcine brain (lane 1), yeast tRNA as negative control (lane 2), adult porcine cerebrum (lane 3), and cerebellum (lane 4). From each tissue, two independent samples of 20 μg of total RNA were measured by S1 nuclease protection assay as described in Materials and Methods. The lane marked with P depicts the undigested probe, and TCGA shows a sequencing reaction as size marker. The arrows with numbers indicate the sizes in nucleotides of the undigested probe (153) and of the fragments of the digested samples (105).

antisense RNA from the opposite strand was excluded by an S1 nuclease protection assay with an antisense-specific probe. Neither Northern blots with total RNA nor with poly(A)$^+$ RNA resulted in the detection of bands with defined sizes. Only diffuse signals ranging from larger than 10 kb to smaller than 1 kb could be observed. Primer extension experiments with different primers from the 5′ region of the larger cDNA clone also showed no distinct-sized bands, but signals in the slot of the gel and a faint smear over the whole range of size (data not shown). Therefore, the size of the UM 9(5)p RNA had to be determined by other means (see below).

We then performed a Southern blot to see whether the sequence of UM 9(5)p is derived from a single- or multicopy gene. Southern blotting with a 1,051 bp *Hin*cII probe clearly showed that the UM 9(5) clones were derived from a single-copy gene and also excluded that the UM 9(5)p transcripts were derived from an expressed repetitive sequence or an expressed pseudogene with a closely related functional gene (Fig. 3). The same probe was used to screen 10^6 plaque-forming units (pfu) of a porcine genomic library. Two overlap-

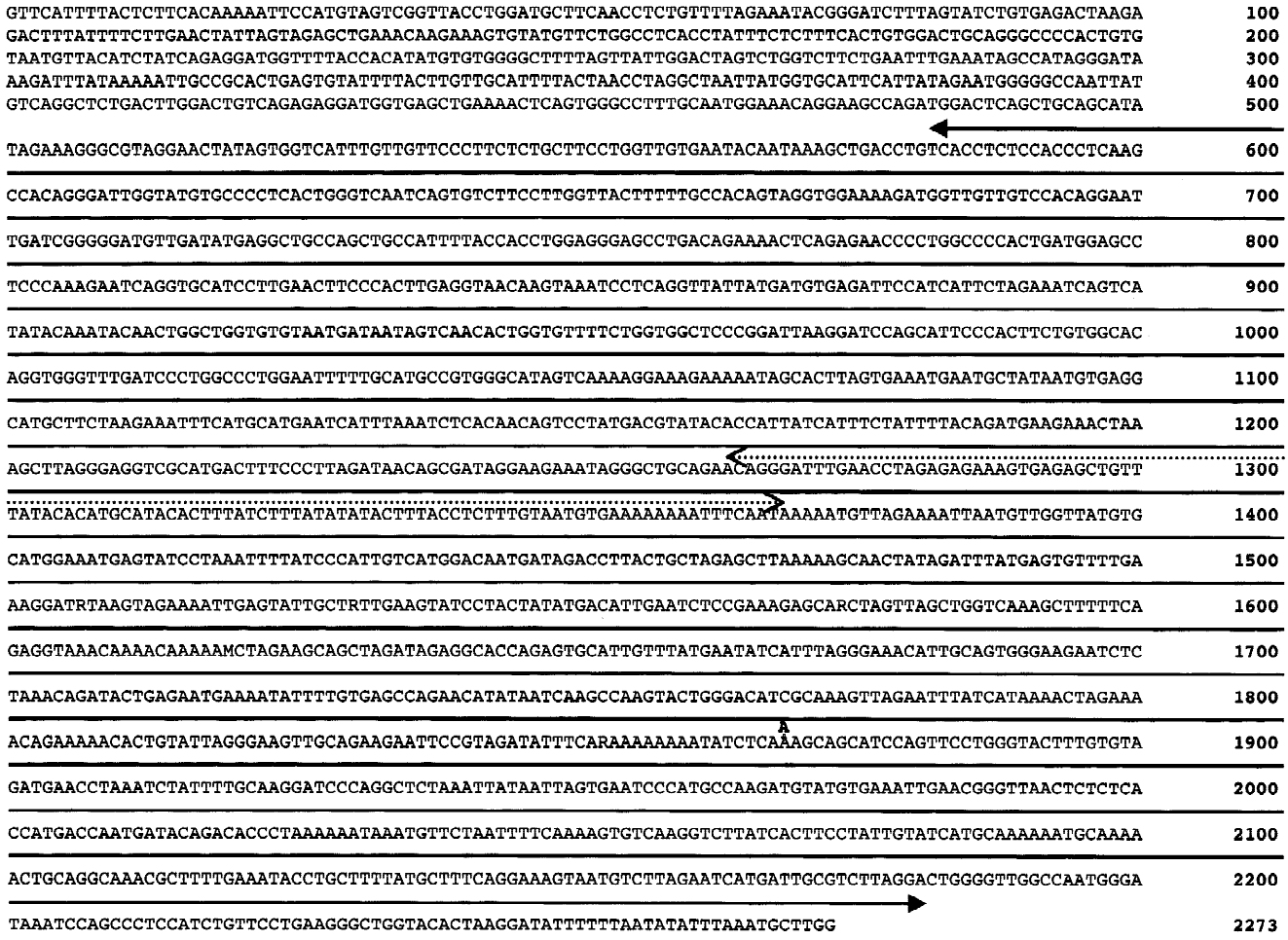

FIGURE 2. Sequence of the long and short UM 9(5)p cDNA and the original differential display clone. The double-sided arrow marks the short UM 9(5)p cDNA clone of 1,701 nt and the double-sided dotted arrow marks the original differential display clone. The bold A with arrow above the lane between nt 1861 and nt 1862 indicates the additional A residue found only in the smaller UM 9(5)p cDNA clone. Nucleotides that differed between the two cDNA clones were designated as follows: Y = C/T, R = A/G, K = G/T, M = A/C.

ping clones of 14,664 nt and 16,415 nt, which covered 18,544 nt of genomic sequence, were isolated. Sequencing of the genomic clones revealed that the cDNA clones were colinear to the genomic sequence and did not span intron–exon boundaries. The genomic sequence information was used to construct probes overlapping the 5′ and 3′ ends of the longer cDNA clone. S1 nuclease protection analysis with these probes and with RNA from porcine cerebellum mapped the transcription start and stop sites of UM 9(5)p to the same regions as the 5′ and 3′ ends of the larger clone isolated from the cDNA library (Fig. 4). No CAAT and TATA boxes in close vicinity to the 5′ end and no consensus sequence for polyadenylation at the 3′ end of the cDNA clones were found in the genomic sequence.

Next, we asked whether related species had a homolog transcript with similar tissue distribution as UM 9(5)p in the pig. We therefore constructed probes from different regions of the long UM 9(5)p cDNA clone and screened 10^6 pfu from cDNA libraries of human cortex, human cerebellum, and mouse brain. No positive clones could be isolated from the mouse cDNA library, whereas screening of the human libraries resulted in the isolation of four overlapping clones (AY072608, AY072609, AY072610, and AY072611), which could be assembled to one contig of 2,336 bp (accession number AY072607). The chromosomal localization of UM 9(5)h was determined with the GeneBridge 4 whole-genome radiation hybrid panel to chromosome 4p15.1 with the maximal LOD score of 21. This confirmed that UM 9(5)h, like UM 9(5)p, is also derived from a single-copy gene. The chromosomal localization assigned later by a BLAST search against the published human genome sequence is 4p15.31. This search localized UM 9(5)h to the sequence segment 10000–12335 nt of fragment NT 006138 on chromosome 4, and showed that the human transcript, UM 9(5)h, is also colinear to its genomic sequence.

A sequence comparison between the human and porcine transcripts with a pustell matrix shows stretches of

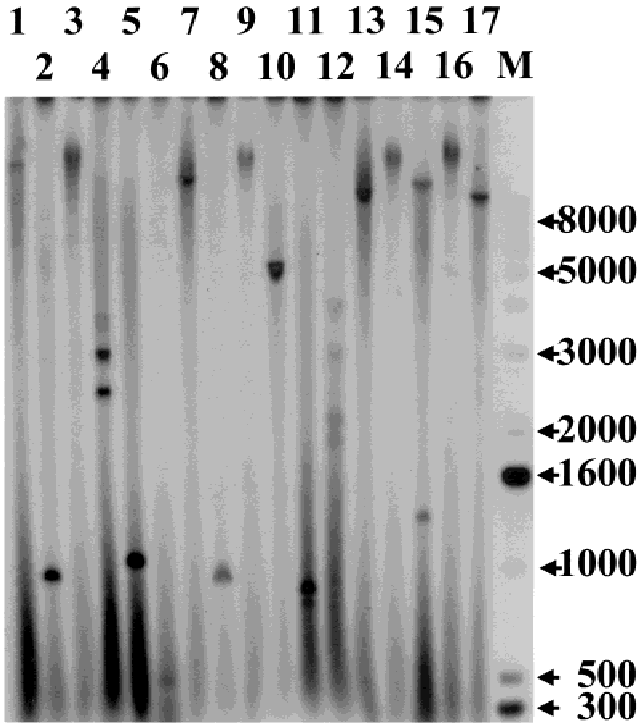

FIGURE 3. Genomic Southern blot of UM 9(5)p. Twenty micrograms of porcine DNA were digested with *Apa*I (lane 1), *Bam*HI (lane 2), *Cfo*I (lane 3), *Eco*RI (lane 4), *Hin*dII (lane 5), *Hin*fI (lane 6), *Kpn*I (lane 7) *Mbo*I (lane 8), *Nar*I (lane 9), *Nco*I (lane 10), *Pst*I (lane 11), *Pvu*II (lane 12), *Sac*I (lane 13), *Sac*II (lane 14), *Spe*I (lane 15), *Sty*I (lane 16), and *Xho*I (lane 17) and analyzed by Southern blotting as described in Materials and Methods. A 1-kb ladder was run along as size marker (M). The arrows with numbers indicate the sizes in nucleotides.

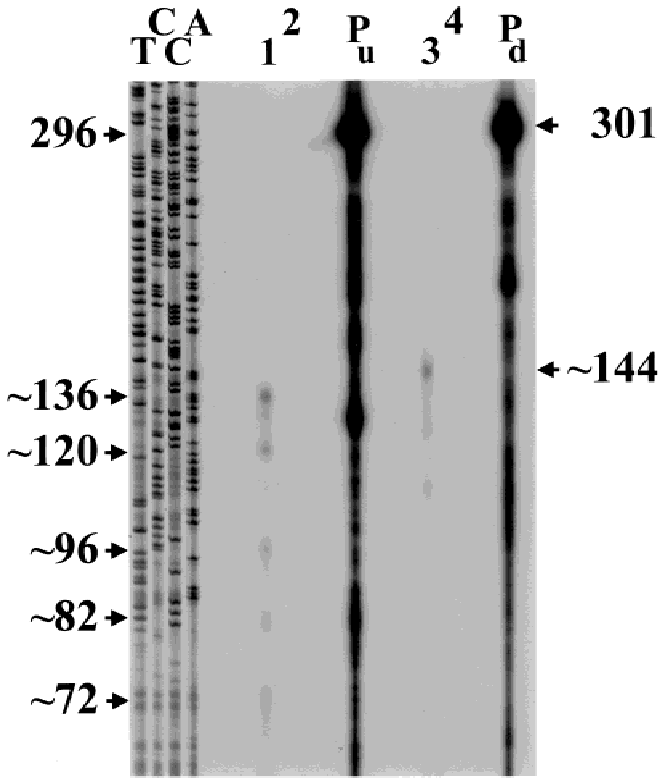

FIGURE 4. Mapping of the transcription start and stop sites of UM 9(5)p. Twenty micrograms of total RNA from adult porcine cerebellum were hybridized to a 296-nt-long upstream probe (P_u) or to a 301-nt-long downstream probe (P_d). The resulting hybrids were analyzed with the S1 nuclease protection assay as described in Materials and Methods. The upstream probe contained an insert of 188 nt, which covered 118 nt from the 5′ end of the long UM 9(5)p clone plus 70 nt from the neighboring upstream genomic sequence; the downstream probe contained an insert of 193 nt, which covered 120 nt from the 3′ end of the long UM 9(5)p clone plus 73 nt from the neighboring downstream genomic sequence. Numbers with arrows indicate the probe sizes, whereas numbers with a tilde in front of them indicate the approximate size of the protected fragments. Two major (~136 nt and ~120 nt) and three minor (~96 nt, ~82 nt, ~72 nt) fragments were detected with the upstream probe, and one fragment of ~144 nt was detected with the downstream probe. The sequence of the original cDNA predicted a size of 118 nt for the protected upstream fragment and a size of 120 nt for the protected downstream fragment. A sequencing reaction (TCGA) was run along with the S1 digestion products as size marker.

high homology between both sequences, which are interrupted by parts sharing almost no homology (Fig. 5A). This pattern of homology is typical of noncoding RNAs as exemplified by a comparison to the mouse and human XIST pustell matrix (Fig. 5B). The sequence homologies between UM 9(5)p and UM 9(5)h depicted in the pustell matrix are distributed along a regression line, and therefore do not occur by chance. The Southern blot with DNA from different species shows that the gene for UM 9(5) is not only conserved in pig (*Sus scrofa*) and man (*Homo sapiens sapiens*), but also in mini pig (*Sus scrofa domestica min.*) and cow (*Bos taurus*; Fig. 6). Only diffuse signals were detected with tissue from old (*Callithrix juccus*) and new world (*Macaca mulatta*) apes, cat (*Felis catus*), fish (*Salmo gairderi*), rat (*Rattus norvegicus*), and mouse (*Mus musculus*; data not shown).

The sequence information from the human genome project allowed us to PCR-amplify fragments from human genomic DNA, which spanned the 5′ and 3′ ends of the human UM 9(5) contig. Probes derived from the cloned PCR products were used to map the transcription start and stop sites of UM 9(5)h, as we did for the porcine RNA. The results shown in Figure 8 pointed to a transcript size very similar to the size expected from the cDNA contig. The results summarized in Figure 4 and Figure 7 suggested that the size of human and porcine cDNA clones represent the complete or the almost complete transcripts of UM 9(5)h and UM 9(5)p, respectively.

To test whether the sequences 5′ and 3′ adjacent to the porcine and human cDNAs are not transcribed, we made probes with sequences located up- and downstream from the mapped 5′ and 3′ ends of the cDNA clones. To our surprise, S1 nuclease protection assays revealed that these regions are also transcribed in human and porcine cerebellum. We therefore continued to examine the regions further 5′ and 3′ from the orig-

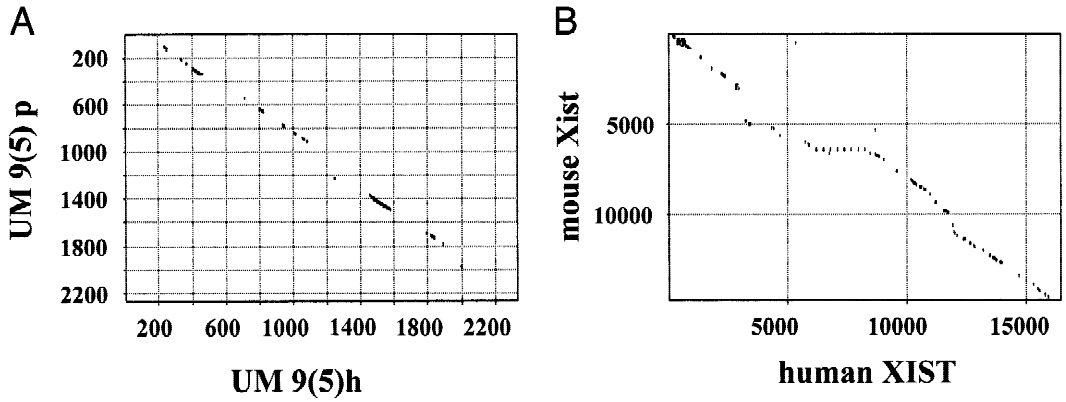

FIGURE 5. Pustell matrix analysis of human and porcine UM 9(5). In **A**, the sequence homology of UM 9(5)h (*x*-axis) and UM 9(5)p (*y*-axis) were compared with a pustell matrix analysis. The settings for the analysis were: window size 30, hash value 6, and the minimum percent score for the window was 65%. **B** shows a comparable pustell matrix of a homology comparison of the human and mouse XIST. Please note that in both matrices, the sequence homologies are not distributed by chance and that homolog parts are interrupted by sequences with almost no homology.

inal cDNAs for RNA expression. As pointed out in Figure 8, transcription in human cerebellum was found up to 5,492 nt further upstream from the 5′ end of the cDNA contig, but no obvious transcription was found 5′ from this point. The transcribed 5′ region in porcine cerebellum extended at least up to 12,371 nt further upstream from the 5′ end of the long cDNA clone, whereas 5′ from this sequence, S1 analysis revealed no clear signals with RNA from porcine cerebellum. Expression of the 3′ region extended at least up to 2 kb downstream from the original cDNAs in human and porcine cerebellum. Therefore, the originally mapped 2,273 nt of the porcine and the 2,336 nt of the human transcript are part of larger transcript units.

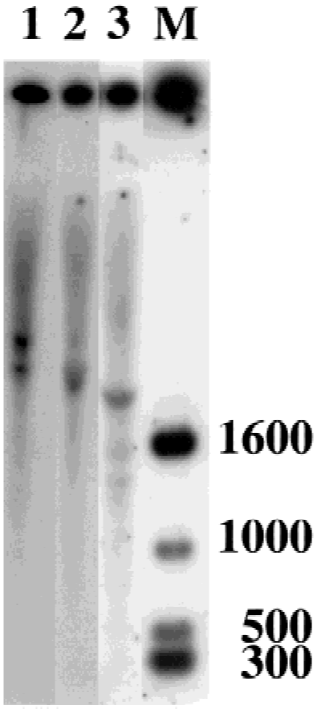

FIGURE 6. Genomic Southern blot of UM 9(5) with DNA from different species. Ten micrograms of genomic DNA from different species were digested with *Eco*RI and analyzed by Southern blotting as described in Materials and Methods. DNA from the following species showed defined bands: pig (*Sus scorfa*; lane 1), minipig (*Sus scorfa domestica min.*; lane 2), and cow (*Bos taurus*; lane 3). Please note that different exposure times are shown in the figure to optimize the signal-to-background ratio for each lane. The following exposure times were used: lanes 1 and 2 were 1 day, lane 3 was 4 days. A 1-kb ladder was run along as size marker (M). The numbers indicate the sizes in nucleotides.

As there are noncoding transcripts of extreme size that are colinear to their genomic sequences (Saitoh et al., 1996; Lee et al., 1999; Mise et al., 1999; Mitsuya et al., 1999; Lyle et al., 2000) and that can cause the imprinted expression of a neighboring gene as antisense RNA (Lyle et al., 2000), we examined the RNA expression from the sense and antisense strand of Slit2, which is the next known gene 3′ from the location of the UM 9(5)h transcript with an opposite orientation. A probe spanning the last four 3′ exons of Slit2 detected low levels of Slit2 sense but no antisense transcription in RNA from human cerebellum, whereas a sense and antisense probe from the last intron and the antisense probe from the last four exons did not reveal any transcription (data not shown). The 9,848 nt of transcribed human sequence and the 16,429 nt of transcribed porcine sequence were used in a BLASTN search for est sequences. The mapped transcribed human region overlaps with 14 nt (nt 4976–4989) of a theoretical protein on LOCI 133009, for which no sequence homologies were found in SwissProt, and it matched with the 131 nt (nt 3737–3870) of an est clone (accession number BF363265). No significant matches with the porcine sequence were found.

Next, we analyzed the expression of UM 9(5)h and UM 9(5)p in different human and porcine tissues. Although not all tissues could be analyzed in both species, Table 1 shows that the tissue distribution of the

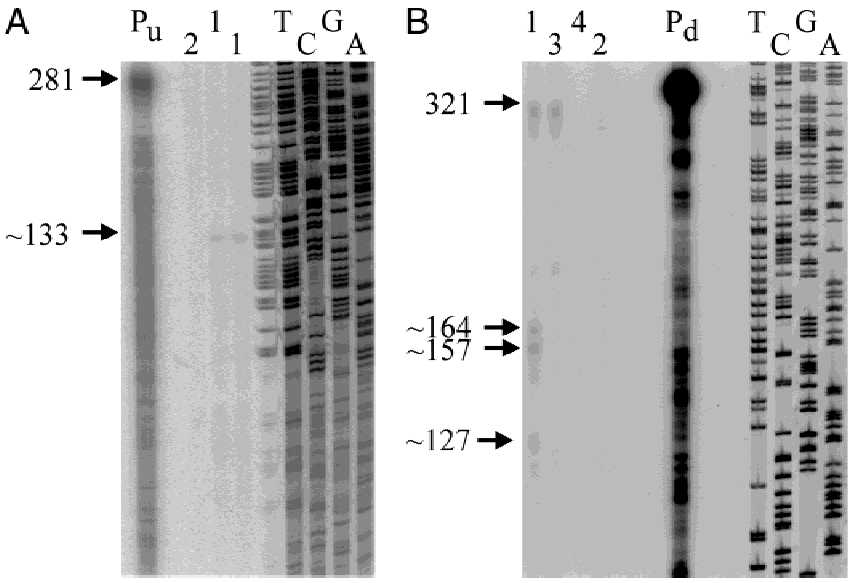

FIGURE 7. Mapping of the transcription start and stop sites of UM 9(5)h. In **A**, 20 μg of total RNA from cerebellum (lane 1) or 20 μg of RNA from yeast (lane 2) were hybridized to a 281-nt-long upstream probe (P_u). In **B**, 20 μg of total RNA from adult human cerebellum (lane 1), 20 μg of yeast total RNA (lane 2), 20 μg of human liver RNA (lane 3), and 4 μg of human genomic DNA (lane 4) were hybridized to a 321-nt-long downstream probe (P_d). In both cases, the resulting hybrids were analyzed with the S1 nuclease protection assay as described in Materials and Methods. The upstream probe contained an insert of 173 nt, which covered 133 nt from the 5′ end of the UM 9(5) cDNA contig plus 40 nt from the neighboring upstream genomic sequence; the downstream probe contained an insert of 205 nt, which covered 141 nt from the 3′ end of the UM 9(5)p cDNA contig plus 64 nt from the neighboring downstream genomic sequence. Numbers with arrows indicate the probe sizes, whereas numbers with a tilde in front of them indicate the approximate size of the protected fragments. One fragment of ~133 nt was detected with the upstream probe, and two major fragments of ~164 nt and ~157 nt plus a minor fragment of ~127 nt were detected with the downstream probe. The sequence of the original cDNA contig predicted a size of 133 nt for the protected upstream fragment and a size of 141 nt for the protected downstream fragment. A sequencing reaction (TCGA) was run along with the S1 digestion products as size marker.

UM 9(5) RNA is very similar in humans and pig, and that the highest expression is found in the cerebellum of both species. The only obvious discrepancy in the expression levels of porcine and human UM 9(5) was seen in the uterus, where strong signals were found in RNA from uteri of two women aged 61 and 79. However, only faint signals were detected in the uteri of cycling pigs. As noncoding RNAs can be either nuclear or cytoplasmic, polyadenylated or without poly(A)$^+$-tract, we examined whether UM 9(5)p is polyadenylated or not and whether the RNA is nuclear or cytoplasmic. S1 analysis of fractionated RNA showed that UM 9(5)p is a nuclear RNA, which is mainly detected in the nonpolyadenylated but rarely in the polyadenylated fraction. In contrast to most coding transcripts, but similar to many other noncoding RNAs, the porcine and human UM 9(5) transcripts contain repeat structures such as CA repeats, a short interspersed nucleotide element (SINE), two long interspersed nucleotide elements in the human and one in the porcine transcript, respectively, one long terminal repeat, and AT-rich regions.

DISCUSSION

In this study, we describe a noncoding transcript that is expressed in different porcine and human tissues and that has the highest steady-state levels in the adult cerebellum of both species. Similar to other noncoding transcripts (Velleca et al., 1994; Ninomiya et al., 1996; Liu et al., 1997; Rougeulle et al., 1998; Brosius, 1999; Mitsuya et al., 1999; Lyle et al., 2000; Millar et al., 2000), the human and porcine UM 9(5) are unspliced and colinear transcripts of their respective genomic regions and, similar to AIR, DISC2, NTT, and BORG (Liu et al., 1997; Takeda et al., 1998; Lyle et al., 2000; Millar et al., 2000), contain different kinds of repeat elements.

As for many other noncoding RNAs from higher eukaryotes (Kloc et al., 1993; Saitoh et al., 1996; Liu et al., 1997; Rougeulle et al., 1998; Jong et al., 1999; Koob et al., 1999; Mitsuya et al., 1999; Millar et al., 2000), the complete sequences of UM 9(5) transcripts could not yet be determined, but indirect evidence suggests that the human and porcine UM 9(5) transcripts are noncoding RNAs. The transcripts of both species do not contain a conserved ORF with a Kozak consensus sequence (Kozak, 1987). The possibility that the currently mapped regions of the UM 9(5) RNAs are 5′ untranslated regions of coding transcripts is also unlikely. The currently mapped size of 9,845 nt of the human and of 16,429 nt of the porcine transcript (Fig. 8) would be very unusual for a 5′ UTR, as the presence of too many start codons would prevent an efficient translation of a potential peptide-encoding ORF (Kozak, 1989,

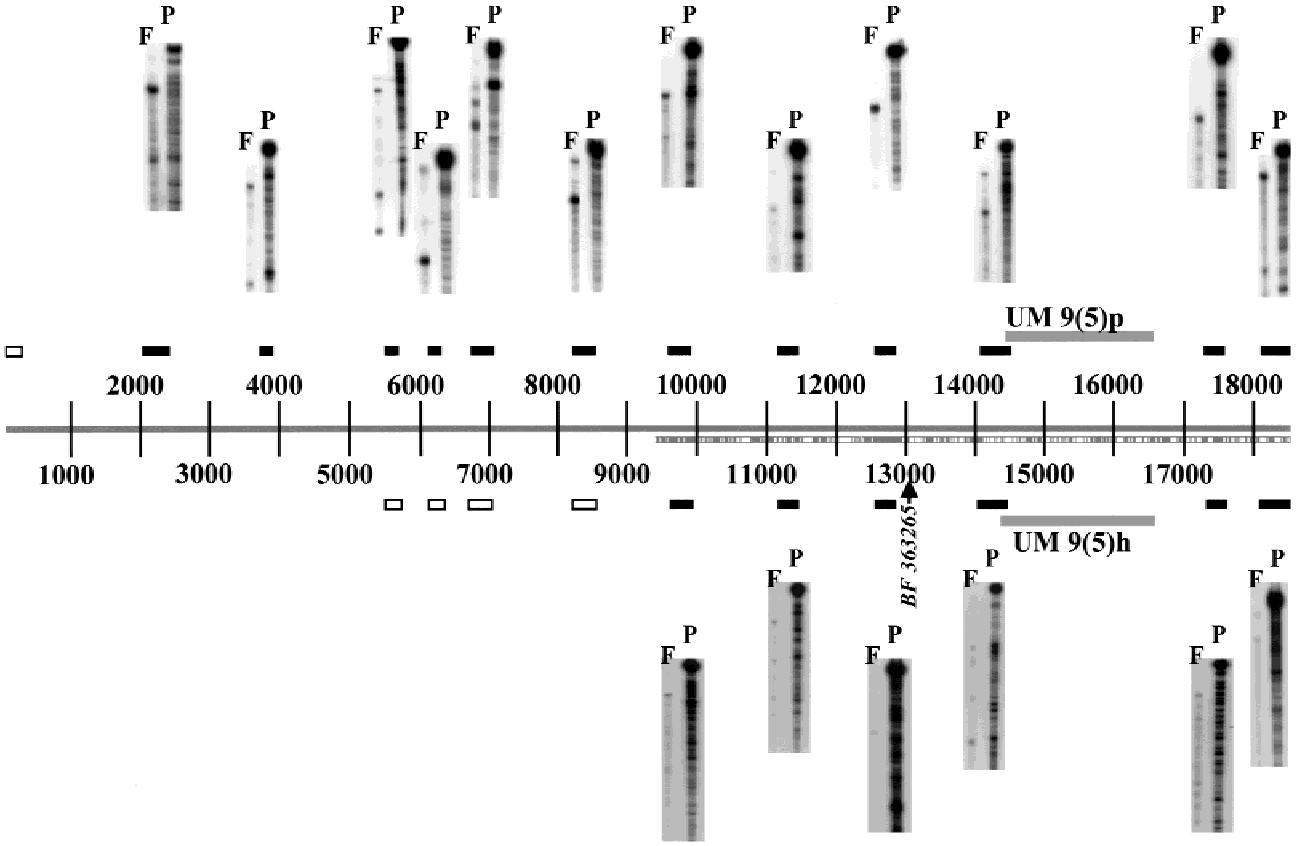

FIGURE 8. Analysis of the extension of UM 9(5)p and UM 9(5)h transcription. The bars in the middle of the figure show the genomic sequence of the porcine (upper, black bar) and the human (lower, striped bar) UM 9(5). The numbers above or below the bars indicate the sequence numbers of the porcine genomic fragment. Small black rectangles above (porcine) and below (human) the bars show the probes used to analyze the extension of UM 9(5) transcription. Filled bars represent probes that detect transcription, whereas S1 nuclease analysis with probes represented by the unfilled bars did not reveal clear signals. Inserts above (porcine) and below the filled bars show the protected fragment(s) (F) and the undigested probe (P) of the respective S1 nuclease analysis. For the sake of simplicity, analyses of yeast tRNA, genomic DNA, and RNA from other tissues were omitted in the inserted pictures. The filled bars with the name above or below the bar show the long porcine cDNA clone and the human contig, respectively. The arrow and the italic type mark the position of the identified human est clone (*BF363265*).

1991). It is also unlikely that the transcribed regions of the UM 9(5) RNAs are parts of noncoding 3′ exons of protein-encoding genes, because the largest 3′ UTR described so far is just 9,280 nt in length and the average length of 3′ UTRs is 2,131 nt ± 1,368 nt (Millar et al., 2000). Two more features of the UM 9(5) transcripts are mainly found in noncoding RNAs; this is the interruption of the sequence homologies between UM 9(5)p and UM 9(5)h by stretches of almost no homology (see Fig. 6), and the presence of SINEs (and other repeats), which are frequently found in noncoding RNAs (Michel, 2002) but rarely in protein-coding transcripts (Yulug et al., 1995). Furthermore, the UM 9(5) transcripts seem to be mainly nuclear and not polyadenylated; these are also characteristics usually not associated with polypeptide-encoding RNAs. We therefore conclude that the 9,848 nt from the human and the 16,429 nt from the porcine transcripts are the 5′ ends of noncoding RNAs.

Although the mapping of the 5′ and 3′ ends of UM 9(5)p and UM 9(5)h revealed distinct transcripts that correspond well to the large porcine cDNA clone and the human cDNA contig (see Figs. 4 and 8), respectively, other experiments, such as Northern blots and primer extensions, did not result in the detection of defined-sized RNAs. This seeming discrepancy is also seen in other noncoding RNAs as, for example, in the early bdx transcripts (Lipshitz et al., 1987) or the PAR-1 RNA (Bielinska et al., 2000) and recently it was questioned whether large noncoding transcripts can correspond to a single RNA species (Heard et al., 1999). Isolation of the two distinct porcine cDNA clones with the same 3′ end and the mapping experiments suggest that at least certain parts of the transcripts seem to be more stable than others. As proposed earlier for the synapse-associated, noncoding 7H4 RNA (Velleca et al., 1994), the partial transcripts of UM 9(5) are probably also not generated from internal transcription start sites, but rather through nucleolytic cleavage from precursor transcripts.

Many noncoding RNAs are tissue specific (Michel, 2002), and some, such as 7H4 (Velleca et al., 1994),

TABLE 1. Comparison of the tissue distribution of UM 9(5)p and UM 9(5)h.[a]

Tissue	Human	Porcine
Aorta endothelial cell culture	n.d.	—
BMV endothelial cell culture	—	—
Brain microvessels	n.d.	+++
Cerebellum	+++	+++
Cerebrum	+	++
Heart	—	—
Testis	n.d.	—
Kidney	—	—
Liver	—	—
Lung	—	—
Muscle	n.d.	—
Meninges	—	—
Ovary	n.d.	—
Pancreas	n.d.	—
Pituitary	(+)	+
Placenta	—	—
Plexus	n.d.	—
Small intestine	—	n.d.
Spleen	—	n.d.
Thyroid gland	—	n.d.
Tongue	—	n.d.
Uterus[b]	++	(+)
Embryonic brain	n.d.	(+)
Embryonic heart	n.d.	+
Embryonic intestine	n.d.	(+)
Embryonic kidney	n.d.	(+)
Embryonic liver	n.d.	—
Embryonic lung	n.d.	—
Genomic DNA	—	—

[a]The expression of UM 9(5) was analyzed with the S1 nuclease protection assay as described in Materials and Methods. The expression was rated high [+++], medium [++], low [+], very low [(+)], or undetectable [—]; some tissues were not available for RNA extraction and, therefore, the RNA expression levels of these tissues could not be determined [n.d.].

[b]Please note that for human uterus only postmenopausal tissue was available, whereas the RNA from the porcine uteri was from cycling pigs.

Bsr (Komine et al., 1999), BC1, BC2, BC 200 (Brosius, 1999), SCA8 (Koob et al., 1999), UBE3A antisense (Rougeulle et al., 1998), the PAR transcripts (Bielinska et al., 2000), and the ZNF antisense RNA (Jong et al., 1999), are mainly or exclusively expressed in neuronal or brain tissue. Although UM 9(5) transcripts are not tissue specific, their predominant expression is also in the cerebellum. Noncoding RNAs can be associated with human diseases (Michel, 2002), but currently no uncharacterized human genetic disorders are linked to chromosome 4p15 in the OMIM database. The chromosomal region of the human UM 9(5) locus has not yet been completely characterized, and data from the porcine genome are not available. Therefore, the possibility cannot be excluded that the UM 9(5) transcripts are, such as AIR (Lyle et al., 2000), antisense to a protein-encoding transcription unit on the opposite DNA strand. Nevertheless, our experiments at least exclude an overlap of UM 9(5)h with Slit2, the next known gene 3′ from the UM 9(5)h location.

Our current knowledge about noncoding RNAs is mainly descriptive, and the function and mode of action have been deduced for some examples only (Michel, 2002). Our work adds another, hitherto unknown noncoding transcript to the growing list of untranslated RNAs. Currently, we cannot assign any obvious function to UM 9(5) and maybe only the transcription as such is important to maintain a certain chromatin status. To find out more about the physiological role of UM 9(5), we are now trying to localize the cellular sources of UM 9(5) transcripts in UM 9(5)-positive tissues. This should allow us the establishment of a cell culture system in which the regulation of UM 9(5) transcription can be studied. In addition, it should facilitate the complete mapping of the UM 9(5) transcription unit.

MATERIAL AND METHODS

Tissues sources

Porcine tissues from pregnant animals and their 61- to 63-day-old fetuses were generously provided by Drs. S. Wallenhorst and E. Dietrich (Institut für Tierzucht und Haustiergenetik; University of Göttingen; Wallenhorst & Holtz, 1999). Other porcine tissues were from the local abattoir. Human post mortem tissues and specimens from surgeries were from the local Departments of Pathology, Neuropathology, and Gynecology; the monkey tissues were from the German Primate Center of Göttingen.

Differential mRNA display, cloning, and sequencing of the PCR-amplified cDNAs

RNA was extracted from brain tissue of adult and embryonic pigs with GTC and a CsCl-gradient according to standard methods (Sambrook et al., 1989). Fifty micrograms of each RNA were digested with DNAse, and phenol/$CHCl_3$- and $CHCl_3$-extracted, precipitated, and redissolved in RNAse-free water. DD, cloning, and sequencing were essentially performed as recently described (Michel et al., 2000).

Northern blots, Southern blots, S1 nuclease protection assays, rapid amplification of cDNA ends, and primer extension

Northern blots with 20–30 μg of total RNA or 5–10 μg poly(A)$^+$ RNA were done with denaturing 1% formaldehyde gels according to standard methods (Sambrook et al., 1989). Southern blots were performed with 10 to 15 μg restriction enzyme-digested genomic DNA according to standard methods. ^{32}P-labeled probes were prepared either by random-primed synthesis using a commercially available kit (Megaprime, Amersham, Braunschweig, FRG), or by in vitro transcription with a commercial kit (transcription in vitro systems, Promega, Mannheim, FRG), according to the manufacturers' instructions. S1 nuclease analysis with total, poly(A)$^+$, cytoplasmic, and nuclear RNA was performed as recently described (Michel et al., 2000). The RACE was done with a Marathon cDNA Amplification kit from Clontech (Palo

Alto, California) according to the instructions of the supplier. Resulting PCR products were cloned into the pGEMT vector and sequenced. The primer extension experiments were done with the AMV reverse transcriptase system from Promega (Promega, Mannheim, FRG) according to the instructions.

Subcellular fractionation and poly(A)$^+$ extraction

Poly(A)$^+$ RNA was extracted from total RNA with the PolyATract mRNA isolation system IV from Promega. For nuclear RNA-enriched and cytoplasmic RNA-enriched preparations, porcine cerebellum was cut with a razor blade in pieces of ~2–3 mm^3 and 10 g of the minced tissue were added to 40 mL of STE buffer (0.32 M sucrose, 10 mM TRIS, pH 7.4, and 30 mM EDTA). The tissue was homogenized on ice with a homogenizer from Braun (Melsungen, FRG), and the resulting suspension was centrifuged at 1,000 × *g* for 10 min at 4 °C. The supernatant was discarded and the pellet was resuspended in STE buffer at a final volume of 50 mL for a second homogenization. The homogenization procedure was repeated twice, and the supernatants were transferred to a new tube to pellet the cellular debris and the residual nuclei at 50,000 × *g* for 45 min at 4 °C. The cytoplasmic RNA was extracted from the resulting supernatant with a CsCl gradient according to standard methods (Sambrook et al., 1989). The nuclear RNA-enriched preparation was prepared from the pellet of the 1,000 × *g* centrifugation. The pellet was again intensively homogenized and 5 mL of the homogenate was layered on top of 15 mL 0.34 M sucrose. The gradient was centrifuged for 10 min at 700 × *g* at 4 °C and the procedure was repeated once with the pellet. The RNA from the resulting pellet was also extracted with a CsCl gradient. The cytoplasmic and nuclear RNAs were run on agarose gels to check for the integrity and the presence (nuclear) or absence (cytoplasmic) of the 45S RNA, respectively.

Construction and screening of genomic and cDNA libraries and chromosomal localization

Five micrograms of poly(A)$^+$ RNA from adult porcine brain were used to construct a cDNA library in λ ZAP XR vector, with the Gigapack II Gold Cloning kit from Stratagene (Heidelberg, FRG). The original library had a titer of 2.2 × 10^6 pfu per microgram phage, and the amplified stock library had a titer of 3 × 10^9 pfu/mL. The premade human cDNA libraries were a λ ZAP library from human cerebellum of a two-year old female (Stratagene) and a 5′-STRETCH™ cDNA library in λgt10 from human cortex (Clontech, Heidelberg, FRG). The mouse brain cDNA library was a commercial library in λTriplEx™ from Clontech. The porcine genomic DNA library was in Lambda FIX II replacement vector (Stratagene), which was a generous gift from Dr. Dr. Bertram Brenig from the Department of Veterinary Medicine in Göttingen. The porcine genomic library was screened with a random-primed labeled 1,051-bp *Hin*cII fragment from the porcine 2,271-bp cDNA clone. Positive clones were isolated and phage DNA was prepared by conventional methods. The purified phage DNA was cut with *Sal*I and cloned in a *Sal*I-cut pGEM4Z for sequencing. All other libraries were also screened with random-primed probes and a hybridization puffer containing 500 mM $NaHPO_4/NaH_2PO_4$, pH7.2, 1% BSA, 1 mM EDTA, and 7% SDS. The GeneBridge 4 whole-genome radiation hybrid panel (Research Genetics, USA) was screened by PCR as described (Isbrandt et al., 2000). In brief, UM 9(5)h-specific primers were: upstream primer: GTCTCCATTTCACAGGAA GAAACA and downstream primer: GCTAACTCAGTCTCT TACTGAGA. PCR conditions were as follows. The first cycle was 5 min at 95 °C; the next 30 cycles were 45 s 95 °C, 1 min 61 °C, 1 min 72 °C; the last cycle was 72 °C for 10 min. The reaction mixture contained 10% DMSO and the Taq polymerase was from Gibco BRL (Karlsruhe, FRG). Genomic DNA from hamster served as control.

Accession numbers

The following GenBank accession numbers were assigned by the NCBI: AY072604 for the genomic porcine sequence and the mapped transcribed region, AY072605 for the long porcine cDNA clone, AY072606 for the short cDNA clone, AY072607 for the human cDNA contig, AY072608 for hc9(5)h-2-18, AY072609 for hc9(5)h-2-1/4, AY072610 for ph9(5)h-1-33, and AY072611 for ph9(5)-2-15/16. The mapped transcribed human sequence is available in the Third Party Annotation Section of the DDBJ/EMBL/GenBank database under the accession number TPA: BK000166.

ACKNOWLEDGMENTS

The authors wish to thank Dr. A. Bayas for early input to the work and Dr. S. Wallenhorst and Dr. E. Dietrich for providing the porcine embryonic tissues. The work was supported by grants of the DFG to P.R. (Ri 516/2-2) and D.I. (DFG-SFB545), and the Verein zur Förderung von Forschung, Wissenschaft und Lehre an der Neurologischen Klinik in Göttingen (U.M.).

Received June 27, 2002; returned for revision July 24, 2002; revised manuscript received September 13, 2002

REFERENCES

Amrein H, Axel R. 1997. Genes expressed in neurons of adult male *Drosophila*. *Cell 88*:459–469.

Bielinska B, Blaydes SM, Buiting K, Yang T, Krajewska-Walasek M, Horsthemke B, Brannan CI. 2000. De novo deletions of SNRPN exon 1 in early human and mouse embryos result in a paternal to maternal imprint switch. *Nat Genet 25*:74–78.

Brockdorff N, Ashworth A, Kay GF, McCabe VM, Norris DP, Cooper PJ, Swift S, Rastan S. 1992. The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell 71*:515–526.

Brosius J. 1999. RNAs from all categories generate retrosequences that may be exapted as novel genes or regulatory elements. *Gene 238*:115–134.

Brown CJ, Hendrich BD, Rupert JL, Lafreniere RG, Xing Y, Lawrence J, Willard HF. 1992. The human XIST gene: Analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell 71*:527–542.

Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, Debruyne FM, Ru N, Isaacs WB. 1999. DD3: A new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res 59*:5975–5979.

Chang DD, Clayton DA. 1989. Mouse RNAase MRP RNA is encoded by a nuclear gene and contains a decamer sequence complementary to a conserved region of mitochondrial RNA substrate. *Cell 56*:131–139.

Chu S, Archer RH, Zengel JM, Lindahl L. 1994. The RNA of RNase MRP is required for normal processing of ribosomal RNA. *Proc Natl Acad Sci USA 91*:659–663.

Heard E, Lovell-Badge R, Avner P. 1999. Anti-Xistentialism. *Nat Genet 21*:343–344.

Isbrandt D, Leicher T, Waldschutz R, Zhu X, Luhmann U, Michel U, Sauter K, Pongs O. 2000. Gene structures and expression profiles of three human KCND (Kv4) potassium channels mediating A-type currents I(TO) and I(SA). *Genomics 64*:144–154.

Jong MT, Carey AH, Caldwell KA, Lau MH, Handel MA, Driscoll DJ, Stewart CL, Rinchik EM, Nicholls RD. 1999. Imprinting of a RING zinc-finger encoding gene in the mouse chromosome region homologous to the Prader–Willi syndrome genetic region. *Hum Mol Genet 8*:795–803.

Kloc M, Spohr G, Etkin LD. 1993. Translocation of repetitive RNA sequences with the germ plasm in *Xenopus* oocytes. *Science 262*:1712–1714.

Komine Y, Tanaka NK, Yano R, Takai S, Yuasa S, Shiroishi T, Tsuchiya K, Yamamori T. 1999. A novel type of non-coding RNA expressed in the rat brain. *Brain Res Mol Brain Res 66*:1–13.

Koob MD, Moseley ML, Schut LJ, Benzow KA, Bird TD, Day JW, Ranum LP. 1999. An untranslated CTG expansion causes a novel form of spinocerebellar ataxia (SCA8). *Nat Genet 21*:379–384.

Kozak M. 1987. An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res 15*:8125–8148.

Kozak M. 1989. Circumstances and mechanisms of inhibition of translation by secondary structure in eucaryotic mRNAs. *Mol Cell Biol 9*:5134–5142.

Kozak M. 1991. An analysis of vertebrate mRNA sequences: Intimations of translational control. *J Cell Biol 115*:887–903.

Lanz RB, McKenna NJ, Onate SA, Albrecht U, Wong J, Tsai SY, Tsai MJ, O'Malley BW. 1999. A steroid receptor coactivator, SRA, functions as an RNA and is present in an SRC-1 complex. *Cell 97*:17–27.

Lee JT, Davidow LS, Warshawsky D. 1999. Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet 21*:400–404.

Lee RC, Feinbaum RL, Ambros V. 1993. The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell 75*:843–854.

Lipshitz HD, Peattie DA, Hogness DS. 1987. Novel transcripts from the Ultrabithorax domain of the bithorax complex. *Genes & Dev 1*:307–322.

Liu AY, Torchia BS, Migeon BR, Siliciano RF. 1997. The human NTT gene: Identification of a novel 17-kb noncoding nuclear RNA expressed in activated CD4+ T cells. *Genomics 39*:171–184.

Lyle R, Watanabe D, te Vruchte D, Lerchner W, Smrzka OW, Wutz A, Schageman J, Hahner L, Davies C, Barlow DP. 2000. The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nat Genet 25*:19–21.

Malik K, Salpekar A, Hancock A, Moorwood K, Jackson S, Charles A, Brown KW. 2000. Identification of differential methylation of the WT1 antisense regulatory region and relaxation of imprinting in Wilms' tumor. *Cancer Res 60*:2356–2360.

Meller VH, Wu KH, Roman G, Kuroda MI, Davis RL. 1997. roX1 RNA paints the X chromosome of male *Drosophila* and is regulated by the dosage compensation system. *Cell 88*:445–457.

Michel U. 2002. Non-coding ribonucleic acids—A class of their own? *Int Rev Cytol 218*:143–219.

Michel U, Stringaris AK, Nau R, Rieckmann P. 2000. Differential expression of sense and antisense transcripts of the mitochondrial DNA region coding for ATPase 6 in fetal and adult porcine brain: Identification of novel unusually assembled mitochondrial RNAs. *Biochem Biophys Res Commun 271*:170–180.

Millar JK, Wilson-Annan JC, Anderson S, Christie S, Taylor MS, Semple CA, Devon RS, Clair DM, Muir WJ, Blackwood DH, Porteous DJ. 2000. Disruption of two novel genes by a translocation cosegregating with schizophrenia. *Hum Mol Genet 9*:1415–1423.

Mise N, Goto Y, Nakajima N, Takagi N. 1999. Molecular cloning of antisense transcripts of the mouse Xist gene. *Biochem Biophys Res Commun 258*:537–541.

Mitsuya K, Meguro M, Lee MP, Katoh M, Schulz TC, Kugoh H, Yoshida MA, Niikawa N, Feinberg AP, 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.

Ninomiya S, Isomura M, Narahara K, Seino Y, Nakamura Y. 1996. Isolation of a testis-specific cDNA on chromosome 17q from a region adjacent to the breakpoint of t(12;17) observed in a patient with acampomelic campomelic dysplasia and sex reversal. *Hum Mol Genet 5*:69–72.

Rastinejad F, Blau HM. 1993. Genetic complementation reveals a novel regulatory role for 3′ untranslated regions in growth and differentiation. *Cell 72*:903–917.

Rougeulle C, Cardoso C, Fontes M, Colleaux L, Lalande M. 1998. An imprinted antisense RNA overlaps UBE3A and a second maternally expressed transcript. *Nat Genet 19*:15–16.

Saitoh S, Buiting K, Rogan PK, Buxton JL, Driscoll DJ, Arnemann J, Konig R, Malcolm S, Horsthemke B, Nicholls RD. 1996. Minimal definition of the imprinting center and fixation of chromosome 15q11-q13 epigenotype by imprinting mutations. *Proc Natl Acad Sci USA 93*:7811–7815.

Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.

Takeda K, Ichijo H, Fujii M, Mochida Y, Saitoh M, Nishitoh H, Sampath TK, Miyazono K. 1998. Identification of a novel bone morphogenetic protein-responsive gene that may function as a noncoding RNA. *J Biol Chem 273*:17079–17085.

Topper JN, Clayton DA. 1990. Characterization of human MRP/Th RNA and its nuclear gene: Full length MRP/Th RNA is an active endoribonuclease when assembled as an RNP. *Nucleic Acids Res 18*:793–799.

Tycowski KT, Shu MD, Steitz JA. 1996. A mammalian gene with introns instead of exons generating stable RNA products. *Nature 379*:464–466.

Velleca MA, Wallace MC, Merlie JP. 1994. A novel synapse-associated noncoding RNA. *Mol Cell Biol 14*:7095–7104.

Wallenhorst S, Holtz W. 1999. Transfer of pig embryos to different uterine sites. *J Anim Sci 77*:2327–2329.

Watanabe Y, Yamamoto M. 1994. *S. pombe* mei2+ encodes an RNA-binding protein essential for premeiotic DNA synthesis and meiosis I, which cooperates with a novel RNA species meiRNA. *Cell 78*:487–498.

Yulug IG, Yulug A, Fisher EM. 1995. The frequency and position of Alu repeats in cDNAs, as determined by database searching. *Genomics 27*:544–548.