C15orf2 and a novel noncoding transcript from the Prader–Willi/Angelman syndrome region show monoallelic expression in fetal brain

Karin Buiting a,⁎, Hülya Nazlicana a, Danuta Galetzka b, Michaela Wawrzik a, Stephanie Groß a, Bernhard Horsthemke a

a Institut für Humangenetik, Universitätsklinikum Essen, Hufelandstrasse 55, 45122 Essen, Germany
b Institut für Humangenetik, Johann-Gutenberg-Universität, Langenbeckstrasse 1, 55131 Mainz, Germany

Received 20 June 2006
Available online 6 March 2007

Abstract

The Prader–Willi syndrome (PWS) region contains several genes transcribed from the paternal chromosome only. We have previously identified a testis-specific gene, C15orf2, which maps between NDN and SNURF-SNRPN and is expressed from both alleles. Here we report on two novel genes (prader-willi region non-protein-coding RNA 1 and 2) located between NDN and C15orf2. By database search we found five partially duplicated copies, of which only one of each appears to be active. PWRN2 is expressed only in testis and is biallelic. PWRN1 is biallelically expressed in testis and kidney, but monoallelically in fetal brain. Methylation analysis of a CpG island 15 kb upstream of exon 1 showed absence of methylation in spermatozoa, but methylated and unmethylated alleles in fetal brain. Reinvestigation of C15orf2 revealed that this gene is also expressed in fetal brain and that expression in this tissue is monoallelic. We conclude that PWRN1 and C15orf2 may play a role in PWS.

© 2007 Elsevier Inc. All rights reserved.

Keywords: Prader–Willi syndrome; imprinting; DNA methylation

Prader–Willi syndrome (PWS) is a neurogenetic disorder that results from the loss of paternal contribution of a 1.5-Mb imprinted region on the proximal long arm of chromosome 15. PWS is characterized by neonatal muscular hypotonia and failure to thrive, hyperphagia and obesity starting in early childhood, and hypogonadism, short stature, small hands and feet, sleep apnea, behavioral problems, and mild to moderate mental retardation. The PWS phenotype results from a 3- to 4-Mb deletion affecting the entire imprinted domain on the paternal chromosome, maternal uniparental disomy, or an imprinting defect leading to maternal imprint on the paternal chromosome. All three lesions lead to the lack of expression of imprinted genes that are active on the paternal chromosome only. So far four protein coding genes that are transcribed from the paternal chromosome only (MKRN3, MAGEL2, NDN, and SNURF-SNRPN) have been identified. Noncoding alternative transcripts of the SNURF-SNRPN gene have been shown to serve as a host for several snoRNA genes. Most of the paternally expressed genes are associated with a differentially methylated region, which is methylated on the maternal chromosome. The paternally expressed snoRNA genes located in introns of SNURF-SNRPN lack a direct methylation imprint. These snoRNAs are expressed from the paternal allele only, because they are processed from the paternally expressed SNURF-SNRPN transcript [1]. Thus, imprinted expression of the snoRNAs is regulated indirectly through SNURF-SNRPN methylation. The contribution to the PWS phenotype of any of these genes is unknown, and it is still a matter of debate whether PWS is caused by the loss of function of a single gene or of several genes. Also it is not clear whether all genes in the PWS/Angelman syndrome (AS) region have been detected.

In 2000 we identified an intronless gene (C15orf2) between SNRPN and NDN, which encodes an 1156-amino-acid protein of unknown function [2]. By Northern blot analysis we had found that C15orf2 is exclusively expressed in testis. Biallelic expression of C15orf2 in adult testis correlated with the absence of methylation of a 250-bp Cpg island. Here we report the
identification of two novel, non-protein-coding genes between NDN and C15orf2. Expression analysis of C15orf2 and one of the novel transcripts suggests that these two genes may be imprinted in fetal brain.

Results

Identification of two noncoding alternatively spliced transcripts

To identify novel transcripts in the 1-Mb region between NDN and C15orf2 (Fig. 1A) we performed a computer-based search for expressed sequences using the NCBI database (http://www.ncbi.nlm.nih.gov/) and the Ensembl Human Genome Server (http://www.ensembl.org). By this we identified two EST clusters. Both clusters contain spliced transcripts, and each cluster represents one gene. By sequence analysis of 15 overlapping EST and 17 RT-PCR clones from one gene (PWRN1, prader-willi region non-protein-coding RNA 1) we identified 26 exons (Supplementary Fig. 1). This gene spans a genomic region of 160 kb. The 26 exons are located on the overlapping PAC clones AC139362, AC139147, and AC100720 with the most 3' end 20 kb upstream of C15orf2. The EST cluster representing the second gene (PWRN2, prader-willi region non-protein-coding RNA 2) consists of 17 overlapping ESTs, which span 7.2 kb of genomic DNA located on PAC clone AC087474 (Supplementary Fig. 2). By analyzing the EST sequences and one RT-PCR product from human testis
RNA we found three exons. The PWRN1 and PWRN2 transcripts are subject to alternative splicing and polyadenylation and have no protein-coding potential.

**Complex pattern of duplication**

Detailed BLAST searches (NCBI) of exonic and intronic sequences revealed a complex pattern of duplications of both genes. In addition to the PWRN1 copy represented by the EST clones and RT-PCR products mentioned above, we found five partial duplications of PWRN1 with sequence similarities between 93 and 97% containing exons 10–20, exons 2–20 (lacking exons 14–18), exons 2–13 (lacking exon 7), exons 1–20 (lacking exons 7–9), and exons 15, 16, 17, and 20 in a 700-kb genomic region upstream of C15orf2 (see Fig. 1B). Several copies of single PWRN1 exons, especially four single copies of exon 1, were also found.

PWRN2 was found to occur in five copies (sequence similarity of 93–97%). All copies are located between exons 12 and 13 of PWRN1 on the opposite strand. Two of the copies lack exon 3 and the most centromeric copy contains only exon 1 in close proximity to exon 13 of PWRN1. Although the duplicated copies share a high degree of sequence similarity, there are several diagnostic sequence differences between them. By employing these differences, we could assign all of the EST and RT-PCR clones to the PWRN1 and PWRN2 copies indicated in Fig. 1B. Thus, only these copies appear to be expressed, although we cannot exclude expression of the other copies at a very low level.

The duplicated region also contains several copies of two polymorphic loci (IR4-3R and D15S817) [3,4]. By Southern blot analysis of genomic DNA and YAC clones we had previously obtained evidence for five copies of IR4-3R [3]. By sequence analysis we now know that it occurs in a total of seven copies (Fig. 1B). Some of these are located between exons 13 and 14 of PWRN1, others are close to a single copy of exon 1. D15S817 appears to occur in four copies.

**Tissue-specific expression**

The fact that the majority of EST clones found for PWRN1 and PWRN2 are derived from testis-specific cDNA libraries suggested that these transcripts are abundant in this tissue and low in other tissues, although we obtained several RT-PCR clones for PWRN1 from fetal brain (Supplementary Fig. 1). For more detailed analysis of the tissue-specific expression patterns we first performed RT-PCR for the most downstream exons of PWRN1, which are not subject to duplication. We used RNA from human blood, fibroblasts, and testis and additional 24 RNA samples contained in a commercially available RNA panel. By RT-PCR we obtained a PCR product in human testis only. However, after reamplification a PCR product in some tissues (prostate, fetal brain, heart, kidney, liver, lung, skeletal muscle, trachea, and spinal cord) was observed (Fig. 2). For PWRN2 also we obtained an RT-PCR product in testis RNA only. In contrast to PWRN1, reamplification gave no PCR product in any other tissue tested.

**DNA methylation analysis**

By sequence analysis using the genome annotation package NIX (http://www.hgmp.mrc.ac.uk/Registered/Webapp/nix/) and the EMBASS CpG plot software (http://www.ebi.ac.uk/emboss/cpgplot/) we identified four CpG islands that were 339 bp (CpG1, AC139362, nucleotides 74804–75142), 227 bp (CpG2, AC139362, nucleotides 125334–125560), 336 bp (CpG3, AC087463, nucleotides 121473–121808), and 337 bp (CpG4, AC087474, nucleotides 92508–92844) in size (see Fig. 1B). These CpG islands show 89.6 to 95.6% sequence identity, indicating that they represent duplicated sequences (Supplementary Fig. 3). One CpG island, CpG1, is located 15 kb upstream of the first exon of PWRN1 and the remaining three CpG islands seem to be duplicated together with one copy of PWRN1 exon 1 (Fig. 3). To find out whether the tissue-specific expression of PWRN1 is associated with methylation of CpG1, we performed bisulfite sequencing in different tissues. Because of the high sequence identity of the four CpG islands we were not able to investigate each CpG island separately. Thus, we established a common methylation-specific PCR (MS-PCR) for all four CpG islands. MS-PCR was carried out on bisulfite-treated DNA of human blood, fetal brain, and spermatozoa in two PCRs for all four CpG islands, one with a primer pair specific for the methylated DNA and another using a primer pair specific for the unmethylated DNA. PCR products for methylated and unmethylated DNA were obtained from DNA...
derived from blood and fetal brain. In contrast, from DNA from spermatozoa only a PCR product for unmethylated DNA was obtained. Amplification products were subcloned and a total of 197 clones (59 for spermatozoa, 67 for blood, and 71 for fetal brain) were analyzed by sequencing. Sequence differences between the four CpG islands were used to assign the clones to the different islands (Fig. 3). CpG1 is unmethylated in spermatozoa and blood. In fetal brain, equal proportions of methylated and unmethylated sequences were found. CpG2, CpG3, and CpG4 are unmethylated in spermatozoa, but methylated at various degrees in blood and fetal brain.

Allelic expression analysis of PWRN1 and PWRN2

To investigate the imprinting status of the novel transcripts we performed RT-PCR with the same primer pairs as for the tissue-specific expression studies. To distinguish between parental alleles we made use of expressed single nucleotide polymorphisms (SNPs). For PWRN1 we analyzed an A to T exchange in exon 26 (position 34585 in AC100720, rs12908526) and for PWRN2 a G to T exchange in exon 2 (position 25618 in AC087474). For PWRN1 we performed RT-PCR on RNA from three different human testis samples and one kidney sample obtained from different individuals. DNA samples were not available. Sequence analysis of the PWRN1 RT-PCR products revealed that two of the testis RNAs contained the A allele only. In the third testis sample and in the one kidney sample we detected both the A and the T allele, indicating that PWRN1 is biallelically expressed in these tissues (data not shown). The first two individuals may be homozygous, although we cannot exclude monoallelic expression. For fetal brain we had DNA and RNA samples from eight different individuals. Two individuals (fb112 and fb156) were heterozygous at the DNA level. Sequence analysis of the RT-PCR products of these two samples revealed the presence of the T allele only in the fb112 RNA and the presence of the A allele only in the fb156 RNA, indicating monoallelic expression of the transcript in fetal brain (Fig. 4). Allele-specific expression was confirmed by cloning and sequencing of the RT-PCR products. For RNA sample fb112 we obtained 38 clones, all containing the T allele, and for RNA sample 156 we obtained 38 clones, 35 of them containing the A allele and 3 containing the T allele. Since parental DNA was not available for further studies we could not distinguish if expression in fetal brain is restricted to the maternal or paternal allele.

For PWRN2, for which expression was found only in testis RNA, two testis RNA samples were investigated for the G/T polymorphism in exon 2. One sample contained only the G allele. The second sample contained both, the G and the T allele, indicating that this transcript was expressed from both parental alleles.

Monoallelic expression of C15orf2 in fetal brain

By Northern blot analysis we had found that C15orf2 is expressed in testis only [2]. Since PWRN1, which maps 25 kb upstream of C15orf2, showed most abundant expression in testis, but low-level expression in other tissues also, we investigated by RT-PCR whether low-level C15orf2 expression can be detected in other tissues, especially in fetal brain. We
performed RT-PCR on RNA from fetal brain using two different primer pairs, one for the coding region and the other for the 3′ untranslated region (3′ UTR) of the gene. Three of the six DNA samples studied (fb55, fb88, and fb154) were found to be heterozygous for one tetranucleotide repeat inside the 3′ UTR (position nt 5898, GenBank Accession No.AF179681). In all three corresponding RNAs we could detect only one allele, indicating that C15orf2 is monoallelically expressed in fetal brain, similar to PWRN1 (Fig. 5B).

Discussion

We have identified two novel genes between NDN and C15orf2. The genes occur in five copies in a 700-kb region (Fig. 1B), but only one copy each appears to be expressed. The genes have no protein-coding potential and are subject to alternative splicing and polyadenylation. In view of the presence of snoRNA genes within the noncoding 3′ part of the SNURF-SNRPN transcription unit [5] we searched the 160-kb PWRN1 locus for possible candidates for C/D box snoRNAs using a computer-aided analysis, but we did not obtain any sequences showing evidence for the presence of all sequences (C-, C′-, D′-, and D-boxes) and structural motifs (short inverted repeats at their 5′ and 3′ ends) of bona fide C/D box snoRNAs.

The expression of PWRN1 is most abundant in testis, but also present at low levels in several other tissues, including fetal brain. SNP analysis revealed biallelic expression of this transcript in human testis and kidney, whereas in two independent human fetal brain samples we detected monoallelic expression of three different isoforms, two of which occur in fetal brain only. Since parental DNA samples were not available, we could not distinguish if expression is restricted to the paternal or maternal allele.

We also found four duplicated copies of a CpG island, one of which (termed CpG1) is located 15 kb upstream of exon 1 of PWRN1. Methylation analysis of CpG1 showed complete absence of methylation in DNA from human spermatozoa, but an equal proportion of methylated and unmethylated alleles in DNA from human fetal brain. Due to the absence of any polymorphism close to or inside CpG1 we were not able to find out whether methylation in fetal brain is allele specific. However, monoallelic expression in this tissue suggests that expression is regulated by differential DNA methylation. In RNA from peripheral blood we could not detect any expression of PWRN1, although CpG1 is unmethylated. A possible explanation for these findings is that blood cells lack a transcription factor for PWRN1.

As reported previously by us [2], C15orf2 encodes an 1156-amino-acid protein of unknown function, which is present in primates only. In contrast to other genes in the PWS/AS region, C15orf2 is conserved in other primates but not in mice. Based on these findings we had suggested that C15orf2 may play a role in primate spermatogenesis. This notion has recently been substantiated by the finding that C15orf2 harbors several genes for PIWI-interacting RNAs (piRNAs), which are believed to regulate spermatogenesis [6–9]. The data reported here suggest that C15orf2 may also have a role in fetal brain. By comparing genes from humans with their chimpanzee orthologues, Nielsen et al. [10] obtained

Fig. 4. Monoallelic expression of PWRN1. Sequence analysis of DNA samples fb112 and fb156 from human fetal brain around a single nucleotide polymorphism in exon 26 (rs12908526) revealed presence of an A and a T allele. In RNA of fb112 only the T allele and in RNA of fb156 only the A allele is present.
evidence that \textit{C15orf2} is under positive selection. Among the 13,731 annotated genes studied, \textit{C15orf2} had the sixth highest score. Since we have now found that \textit{C15orf2} is expressed in fetal brain and may be subject to genomic imprinting, we cannot exclude a role for this gene in PWS.

\textbf{Material and methods}

\textit{Patients and tissues}

Fetal brain samples were obtained from aborted fetuses, which underwent pathological examination and cytogenetic analysis in the Pediatric Pathology Department at Mainz University School of Medicine. Use of anonymized “excess” tissue materials for scientific analyses was approved by the local ethics committee [Aerztekammer Rheinland-Pfalz, Decision No. 837.103.04 (4261)]. Gestational age of each fetus was determined by foot length measurements (in mm) and last menstrual period [11]. Only fetuses without detectable abnormal development and with normal karyotypes were selected for this study. Fetal tissues were dissected within 24 h after abortion and stored at −80 °C until further analysis.

\textit{Preparation of RNA and DNA}

Total RNA from human blood and fibroblasts was isolated with the QIAamp RNA Blood Mini kit (Qiagen). RNA and DNA from human fetal brain samples were extracted using the NucleoSpin RNA II kit (Macherey and Nagel, Cat. No. 740955) in combination with the NucleoSpin RNA/DNA buffer set (Macherey and Nagel, Cat. No. 740944). Furthermore, 24 RNA samples from the total RNA master panel (BD Bioscience) were used.
Genotyping

Eight DNA samples from human fetal brain were analyzed for heterozygosity of an expressed SNP (rs12908526) in exon 26 of PWRN1 using primers Ex26SNP1, 5′-GAGAACAGTGTATCACAACAGA-3′, and Ex26SNP2, 5′-TGTCAACAGGAGGAAGGAA-3′ (annealing 58 °C). Six DNA samples were studied for heterozygosity of four polymorphic nucleotide repeats in the 3′ untranslated region of C15orf2 as described by Färber et al. [2].

cDNA analysis

PWRN1 and PWRN2 transcripts were isolated from Marathon-Ready human testis and human fetal brain cDNA (BD Biosciences). PCR was carried out with 0.5 ng cDNA in a total volume of 100 μl for 35 cycles of denaturation (95 °C for 15 s), annealing (30 s; for annealing temperature see primer differences), and extension (72 °C for 30 s). Reamplification was performed with 5 μl of PCR products in a total volume of 50 μl using the same PCR conditions. PCR products were verified by sequencing with PCR primers. In the case of multiple PCR products, the products were cut from the gel, purified with the MinElute Gel Extraction kit (Qiagen), and sequenced directly or sequenced with vector-specific primers after subcloning into the pGEM-T Easy vector (Promega). RT-PCR primer pairs used for PWRN1 were PWRN2-Exon7F4, 5′-AATCCTGGGACAATTTGACATC-3′ (annealing 56 °C (RT-1)); PWRN1-Exon19F, 5′-GCAAATCTGGAATCCGGTCT-3′, and Ex26R, 5′-ACCTGGGCAAACTTGATCAGA-3′ (annealing 58 °C (RT-2)); AK081747-F2, 5′-AACACGGGAGATGCATTGACA-3′, and AL704599R2, 5′-TCTGTTTCCCGAAGGTAGTCT-3′ (annealing 60 °C (RT-3)); PWRN1-Exon2F, 5′-GATTCCCCAGCT-GATCCTCG-3′ and PWRN1-Exon8R, 5′-AGAAGGGTGGTCAATGCTAATGACAG-3′ (annealing 63 °C (RT-4-RT-15)); Ex23F, 5′-GCTCCTTTTTAGTGGAGAAATGTGTTT-3′ and Ex26R, 5′-ACCTGGGCAA-CAATTGACTC-3′ (RT-16 and RT-17).

For PWRN2 the following primers were used: B156178F1, 5′-GCCCTTTTATGTGGAGACATGCTG-3′ and CDS57156R2, 5′-AACATGGTGAAGGGTGGGCA-3′ (annealing 60 °C).

Tissue-specific expression analysis.

Total RNA (1 μg) was reverse transcribed with random hexamers. For PWRN1 and PWRN2 RT-PCR was performed with the GeneAmp RNA PCR kit (Perkin-Elmer). For PWRN1 we used primers for exon 23 (Ex23F, 5′-GCTCCTTTTATGTGGAGACATGCTG-3′) and exon 26 (Ex26R, 5′-ACCTGGGCAAACTTGATCAGA-3′). For PWRN2 we used primers for exon 1 (B156178F1, 5′-GCCCTTTTATGTGGAGACATGCTG-3′) and exon 2 (CDS57156R2, 5′-AACATGGTGAAGGGTGGGCA-3′). PCR on cDNA products was carried out in a total volume of 50 μl for 35 cycles of denaturation (95 °C for 15 s), annealing (60 °C for PWRN1 and 60 °C for PWRN2 for 30 s), and extension (72 °C for 30 s). Reamplification was performed on 1 μl of the original PCR sample using the same PCR conditions.

For reverse transcription of C15orf2, the reaction mixture containing 30 μg of total RNA, 4 μl of random primer (Roche), and 4 μl of 5× reverse transcriptase buffer (Invitrogen) was incubated for 10 min at 70 °C and then for 15 min at room temperature. After addition of 2 μl of dNTPs (5 mM each), 2 μl of 0.1 M DTT, 0.5 μl of RNase inhibitor and 1 μl of reverse transcriptase (Invitrogen), the mixture was incubated for 2 h at 37 °C. Then 1 μl RNase H was added and incubated for 20 min at 37 °C. The resulting cDNA was purified with a QIAquick PCR purification kit (Qiagen). Two sets of primers were used for PCR: Pcr1en47, 5′-TATGGGAAAGACTCCCTTGC-3′, and pGRTr2, 5′-CTGGGCTAAGGGCCCTGTC-3′ (annealing 58 °C), for the 3′ untranslated region and pen2c, 5′-AGGCTGATCCGCGTCA-3′, and pen2mr, 5′-AATA- CAGAAAGGCTCGTAC-3′ (annealing 56 °C), for the coding region of the gene [2]. PCR was performed on 450 ng cDNA in a total volume of 50 μl for 35 cycles of denaturation (95 °C for 15 s), annealing (58 or 56 °C for 15 s), and extension (72 °C for 30 s). In control experiments, reverse transcriptase or cDNA was omitted from the reactions. To check the integrity of the RNA and to exclude contamination with genomic DNA we used primers for β-actin [12].

Allelic expression analysis

To investigate allelic expression of PWRN1 and PWRN2, we used an expressed SNP in exon 26 of PWRN1 and exon 2 of PWRN2. RT-PCR conditions were as described for the tissue-specific expression analysis. The RT-PCR products from human testis and kidney were analyzed by sequencing.

Allelic expression of PWRN1 in human fetal brain was investigated in two heterozygous fetuses. Reverse transcription of the RNA was performed as described above for C15orf2 (see previous paragraph). PCR products were analyzed by direct sequencing. For both RNA samples, rb12 and rb15 the PCR products were cloned into the plasmid vector pGEM-T Easy (Promega) and clones were sequenced using a vector-specific primer.

Allelic expression of C15orf2 in human fetal brain was studied in three fetuses heterozygous for an intragenic tetranucleotide repeat (nt 5898, GenBank Accession No. AF179681) [2]. PCR was carried out with primers pc313b, 5′-ATGTTTCTTTAAGCTACCCAGTC-3′, and pc314rFM, 5′-GCACTATGGCTG- TAAAGGA-3′ (annealing 56 °C). Fluorescence-tagged PCR products were analyzed using an ABI 3100 automatic capillary genetic analyzer and GeneScan and Genotyper software (Applied Biosystems).

Analysis of DNA methylation

Genomic DNA from human testis, spermatozoa, blood, and fibroblasts was treated with sodium bisulfite according to standard methods. A 114-bp PCR product was amplified in a common methylation-specific PCR for all four CpG islands using primers Cpg-all-FM, 5′-AATCTAGAAAATCTCAAAACGAA-3′, and Cpg-all-RM, 5′-AATCTACAAAAAATTCCAAAAAGGAA-3′, for the methylated DNA and Cpg-all-FU, 5′-GGATGATGATATAAGTTAGTTGATT-3′, and Cpg-all-RU, 5′-AATCTACAAAAAATTCCAAAAAGGAA-3′, for the unmethylated DNA (annealing temperature 58 °C). PCR products were gel-purified using the MinElute Gel Extraction kit (Qiagen) and cloned into the pGEM-T Easy vector (Promega). Plasmid clones obtained were sequenced with the SP6 vector-specific primer.

Sequence analysis

Sequencing reactions were performed with fluorescence-tagged dideoxynucleotides (BIG Dye kit) and the Taq cycle sequencing procedure (ABI). Sequences were analyzed on an ABI 3100 DNA sequencer.

Acknowledgments

We thank Jérôme Cavaillé and Herve Seitz for computer-assisted analysis to detect canonical and conserved C/D RNA genes; Thomas Haaf, Xavier Estivill, Miroslava Ogorelkova, and Celia Cerrato Rivera for helpful discussions; and Christina Lich for expert technical assistance. Part of this work was supported by the Deutsche Forschungsgemeinschaft (BU907/1-3).

Appendix A. Supplementary data


References


