J Mol Evol (2006) 62:73–88
DOI: 10.1007/s00239-005-0041-3

JOURNAL OF MOLECULAR EVOLUTION

Primate-Specific Endogenous Cis-Antisense Transcription in the Human 5q31 Protocadherin Gene Cluster

Leonard Lipovich, Ravi Raj Vanisri, Say Li Kong, Chin-Yo Lin, Edison T. Liu

Genome Institute of Singapore, 60 Biopolis Street #02-01, Singapore 138672, Singapore

Received: 22 February 2005 / Accepted: 31 July 2005 [*Reviewing Editor:* Dr. Manyuan Long]

Abstract. Protocadherins (PCDH), localized to synaptic junctions, contribute to the formation of neuronal networks during brain development; thus, it is speculated that protocadherins may play a role in evolution of neuronal complexity. While protocadherin genes are highly conserved in vertebrates, EST evidence from the locus suggests apparently species-specific cis-antisense transcripts. Novel cis-antisense transcripts, which partially overlap the PCDHα12 variable exon, PCDHβ3 single-exon gene, and PCDHψ5 unprocessed pseudogene in the human 5q31 PCDHα/β/γ gene cluster and which are coexpressed with sense-strand transcripts in fetal and adult brain, were identified computationally and validated by gene-specific strand-specific reverse transcriptase PCR (SSRTPCR) and sequencing. Absence of antisense transcripts arising from equivalent genomic locations in mouse indicates that the antisense transcripts originated in the primates after the primate-rodent divergence. Furthermore, not all expected orthologues of human sense and antisense PCDH transcripts were detected in rhesus macaque brain, implying that protocadherin expression patterns differ between primate species. RT followed by quantitative real-time PCR (QPCR) analysis of the three genes in the brain of all three species, and of the PCDHβ15 gene paralogous to PCDHψ5 in human and rhesus, revealed that the presence of antisense transcripts was significantly associated with lower sense expression levels across all orthologues. This inverse relationship, along with the pattern of sense and antisense coexpression in the brain, is consistent with a regulatory role for the primate-specific PCDH cis-antisense transcripts, which may represent recent evolutionary inventions modulating the activity of this conserved gene cluster.

Key words: Protocadherin — Antisense — Primate-specific — Gene birth — Noncoding RNA

Correspondence to: Leonard Lipovich; *email:* LL@gis.a-star.edu.sg

Introduction

Protocadherins: Localization and Function

Protocadherins (PCDHs), the largest subgroup of the cadherin superfamily of calcium-dependent cell-cell adhesion glycoproteins (Frank and Kemler 2002), are major structural and functional components of synapses and are expressed on the surfaces of neurons at synaptic junctions (Noonan et al. 2003). Each PCDH displayed on the surface of a given cell potentially facilitates homophilic adhesion formation with adjacent cells displaying the identical PCDH (Vanhalst et al. 2001). Evidence for heterophilic interactions of Pcdh α and γ proteins in the brain also exists (Murata et al. 2004). Individual neurons are capable of expressing distinct but overlapping subsets of PCDHs, which may provide a combinatorial molecular code for neuron-to-neuron connections (Wang et al. 2002). By determining which neurons interact with which other neurons in the vicinity, PCDHs likely account

for some of the combinatorial complexity in neuronal networks, affecting brain development and possibly memory (Noonan et al. 2003) through neuronal morphogenesis, synaptic connection formation, and synaptic transmission regulation (Frank and Kemler 2002).

Structure of the 5q31 Protocadherin Gene Cluster

The majority of PCDH genes are found in three adjacent clusters, mapping in human (*Homo sapiens*) to 5q31. The three clusters—PCDHα, PCDHβ, and PCDHγ—occur sequentially to one another, with PCDHα closest to the centromere and PCDHγ closest to the 5q telomere. Together, the three human 5q31 PCDH clusters span ~750 kb and contain putative regulatory elements upstream of each variable exon. For a detailed diagrammatic representation of PCDH gene cluster structure, see Wu et al. (2001).

The PCDHα cluster consists of a tandem array of multiple alternative first exons, followed by three constitutive exons. Any one of the multiple alternative first exons can be spliced to the downstream constant-exon cassette. The PCDHγ cluster is organized in the same fashion. Each alternative PCDHα and PCDHγ first exon encodes an entire extracellular domain, the transmembrane segment, and a part of the intracellular C-terminal domain. The PCDHβ cluster, which maps between the PCDHα and PCDHγ clusters, consists exclusively of single-exon genes.

Most human PCDHα first exons, PCDHβ genes, and PCDHγ first exons have one-to-one mouse (*Mus musculus*) orthologues. However, some are the products of duplications that postdated the human/mouse divergence and lack true one-to-one orthologues in the mouse. Some other PCDH exons and genes are functional in one species but pseudogenic in another (Vanhalst et al. 2001). PCDHα and PCDHγ expression regulation is in agreement with the "alternative promoter choice and cis-splicing" model. However, specific mechanisms of PCDH expression remain uncharacterized (Wang et al. 2002).

Cis-Antisense: Definition, Incidence, and Significance

A cis-antisense gene pair is operationally defined as a pair of genes which reside on opposite strands in the same locus in such a configuration that at least one exon of one gene overlaps at least one exon of the other. Cis-antisense has been detected in prokaryotes (Vanhee-Brossolet and Vaquero 1998), *Arabidopsis* (Yamada et al. 2003), and *Drosophila* (Misra et al. 2002). Up to 22% of human genes may participate in cis-antisense pairs (Chen et al. 2004).

Cis-antisense is a gene expression regulatory mechanism which functions at both transcriptional and post-transcriptional levels. At the transcriptional level, competitive transcriptional interference (Prescott and Proudfoot 2002) and sense-strand silencing by antisense-mediated promoter methylation (Tufarelli et al. 2003) have been demonstrated. Posttranscriptionally, cis-antisense transcripts can regulate alternative splicing (Hastings et al. 1997), may be involved in RNA editing (Lavorgna et al. 2004), and have been shown to form double-stranded RNA duplexes with their sense counterparts. Although the duplexes may be targeted for degradation by cellular RNAses, translation attenuation through formation of stable undegraded RNA duplexes may occur as well (Podlowski et al. 2002).

Confirmed functions of cis-antisense are diverse. Antisense transcripts are associated with autosomal imprinting and X-inactivation (Shibata and Lee 2004) and may function by allele-specific silencing (Verona et al. 2003). Transcript abundance ratios of certain key developmental regulators and their noncoding cis-antisense partners are inversely correlated and change during cell differentiation, suggesting a function for cis-antisense in modulating sense levels (Blin-Wakkach et al. 2001). Mutations in a noncoding cis-antisense transcript are sufficient for pathogenesis of a neurodegenerative disorder whose mechanism depends on the sense-encoded protein (Nemes et al. 2000). Downregulation of sense-encoded protein expression by an endogenous trans-antisense transcript has been demonstrated in a mammalian cell line, although the antisense in that case involved an interspersed repetitive element in trans (Stuart et al. 2000). Cis-antisense–mediated downregulation of sense expression, albeit in an in vitro system, has been confirmed as well (Thenie et al. 2001).

Early samplings of the mammalian cis-antisense subtranscriptome suggested that only a minority of the cis-antisense pairs (27% in mouse [Kiyosawa et al. 2003]) involve solely protein-coding genes. Noncoding cis-antisense transcripts in the remainder of the dataset differ drastically from other types of regulatory RNAs, i.e., microRNAs, which have received more attention in recent years. Unlike microRNAs, cis-antisense noncoding RNAs are mRNA-like in all respects, except that they do not code for protein. Cis-antisense noncoding RNAs are 5′-capped (Imamura et al. 2004; Kiyosawa et al. 2003), mostly canonically spliced (Chen et al. 2004) and polyadenylated, longer than microRNAs and snRNAs, pol(II)-promoted, encoded by the same locus as the target (Lavorgna et al. 2004), independent of cytoplasmic Dicer (Tran et al. 2004), and complementary to coding targets both within and

outside of 3′ UTRs (Chen et al. 2004; Yelin et al. 2003).

Origins of New Genes and Primate-Specific Functions

The genomic basis of phenotypic distinctions between closely related species remains uncertain. Existing explanations for the drastic differences in phenotypes between closely related species such as chimpanzees and humans invoke regulatory element differences responsible for distinct expression profiles of homologous genes (King and Wilson 1975) or lineage-specific phenotypes related to the loss of function of particular genes during evolution (Olson and Varki 2003). However, it is conceivable that some phenotypic differences might be due to a gain of function in one lineage, encoded by a gene absent in the other. In fact, origin of new genes, which enable novel functions and contribute to genetic diversity, is recognized as a fundamental biological process. In view of the obvious and pronounced differences in mental ability, immune response, and reproductive biology between primates and nonprimates—as well as within primates—novel functions encoded by primate-specific new genes are important to study. Nevertheless, the exact mechanisms giving rise to new genes remain to be elucidated, although several case studies suggest that one pathway by which new genes are created is the shuffling of existing coding-gene exons, which generates both coding and noncoding new genes and is often facilitated by retrotransposition (Long et al. 2003). For example, the primate-specific genes PMCHL1 and PMCHL2 have formed through a complex combination of cis-antisense transcription, retrotransposition, novel splice site recruitment, and block duplication during primate evolution (Courseaux and Nahon 2001). Nonconservation of multiple genes in mammalian antisense pairs (Shendure and Church 2002; Veeramachaneni et al. 2004) makes it plausible that certain cis-antisense transcripts have recent evolutionary origins. If such transcripts exist at the PCDH gene cluster, they can represent attractive functional candidates underlying the genomic basis of mammalian interspecies differences in neuronal and thus behavioral complexity.

Although the structure of the human 5q31 PCDH gene cluster is known in exquisite detail, no mention of endogenous cis-antisense transcription in this locus could be found in the literature. We report in silico discovery and experimental validation of novel cis-antisense transcripts in the 5q31 PCDH gene cluster, followed by a qualitative and quantitative multispecies analysis of sense and antisense transcription.

Materials and Methods

Sequence Analysis

Identification of Human Cis-Antisense Transcription

Finished and HTGS-draft human genomic clones encompassing the 5q31 PCDH cluster (tiling path, 5cen to 5qter: AC005609.1, AC010223.6, AC025436.2, AC005754.1, AC074130.3, AC005752.1, AC005618.1, AC005366.1) were visualized using Seqhelp software (Lee et al. 1998). All visual clusters of ESTs partially overlapping sense-strand PCDH exons but having a discordant genomic footprint (distinct genomic locations of transcription start and end sites, and of splice donor/acceptor sequences for spliced ESTs) relative to the sense exons were noted. Orientation of representative ESTs from each cluster was determined by BL2SEQ (Tatusova and Madden 1999) and Spidey (Wheelan et al. 2001) pairwise alignments to genomic sequence. For plus/minus HSPs with 3′ ESTs, strandedness was inverted, because by convention the first nucleotide of 3′ EST sequences in GenBank represents the 3′ end of the corresponding transcript.

Identification of Sequences Orthologous to Human PCDH Regions

Putatively orthologous chimpanzee sequences were identified by a BLAT search (Kent 2002) of the November 2003 chimpanzee WGS assembly at the UCSC Genome Browser portal (Kent et al. 2002) with human queries. Putatively orthologous rhesus monkey (*Macaca mulatta*) sequences were identified by a TraceDB Mega-BLAST BLASTN search (http://www.ncbi.nlm.nih.gov/blast/mmtrace.shtml) of the "*Macaca mulatta*—WGS" and "*Macaca mulatta*—other" databases with human queries. Mouse orthologues and best homologues were defined as the gene hits with simultaneously highest BLASTN and BLASTP scores relative to the given human query. All orthologues were verified by reciprocal BLAST against the appropriate human databases, and nonhuman sequences whose top-scoring human matches differed from the original human query were discarded. Precomputed global human/chimpanzee and human/mouse BLASTZ outputs (Schwartz et al. 2003), underlying the "chained BLASTZ alignments" track of the UCSC portal, were utilized to obtain pairwise alignments of orthologous regions.

Experimental Protocols

Note: PCR conditions, along with all primer and probe sequences, are cited in the supplementary information file.

Nonquantitative PCR (SSRTPCR)

Strand-Specific cDNA Synthesis. Human adult brain total RNA (Clontech; one donor; 43-year-old male Caucasian; no pathology noted), human fetal brain total RNA (Clontech; pooled from 59 spontaneous abortuses, Caucasian, male and female, 20–33 weeks), *Macaca mulatta* adult brain total RNA (BioChain Institute, Inc.; one donor; no pathology), and mouse pooled RNA from a male and a 12.5-day-pregnant female (a gift from Sai-Kiang Lim, GIS, Singapore) were used for reverse transcription (RT) reactions to make strand-specific cDNAs.

Mouse pooled RNA was treated with DNase I before RT reaction. One microgram of total RNA was incubated with 1 μl of DNase I (2 U/μl;Ambion,USA) at 37°C for 60 min and the reaction was inactivated at 95°C for 5 min.

Two different types of RT reactions were performed, with SuperScriptII reverse transcriptase and ThermoScript reverse transcriptase, using gene-specific untagged primers and gene-specific tagged primers, respectively. For human and rhesus PCDHβ15, as well as human PCDHψ5, ThermoScript was used to detect transcription in both sense and antisense orientations, whereas for other amplicons solely SuperScriptII was used.

For SuperScriptII reverse transcriptase (Invitrogen) reaction, 200 ng total RNA, 20 μM gene-specific primers, and 10 mM dNTP (Invitrogen) were incubated at 65°C for 5 min and the tubes were immediately placed on the ice. The contents were colleted by brief centrifugation. Then 5× first-strand buffer, 0.1 M DTT, 1 μl RNaseOut, and 1 μl SuperScript 11 (200 U) were added to the tube for a final volume of 20 μl. RT was carried out for 60 min at 42°C and the reverse transcriptase activity was inactivated at 70°C for 15 min.

For ThermoScript reverse transcriptase (Invitrogen) RT, 200 ng total RNA, 20 μM gene-specific tagged primer, and 20 μM dNTP were incubated for 5 min at 70°C. The tubes were immediately placed on ice. Then 5× cDNA synthesis buffer, 0.1 M DTT, 1 μl RNaseOut, and 1 μl TheromoScript reverse transcriptase (15 U/μl) were added. The temperature was reduced for primer annealing for 2 min and then returned to 70°C for a further 30 min. Reverse transcriptase activity was inactivated at 98°C for 15 min. Three negative controls accompanied each RT reaction: exclusion of template, exclusion of enzyme, and both.

For ThermoScript reverse transcriptase RT, Exonuclease 1 (10 U; Amersham International) was added to 10 μl of cDNA to degrade unincorporated primers upon completion of RT and incubated at 37°C for 45 min, followed by inactivation at 98°C for 15 min.

Nested PCR Amplification. After optimization of amplicons on genomic DNA (details available upon request), 2 μl of cDNA was used in the first-round PCR and 2 μl of the first-round product was used in the second-round PCR. PCR products were analyzed by 2% agarose gel electrophoresis.

Both genomic PCR products and cDNA PCR products were purified using a QIAquick Gel Extraction kit and 25 to 50 ng of purified PCR products was used for cycle sequencing reactions with 3.2 pmol of forward or reverse primers and 4 μl of sequencing Premix (Big Dye terminator), to confirm all amplicon identities at the sequence level.

Quantitative PCR (QPCR)

We quantified the sense expression level of PCDH using a real-time fluorescence detection method. Human adult and fetal brain total RNAs (Clontech), *Macaca mulatta* adult brain total RNA (Bio-Chain Institute, Inc.), mouse adult brain total RNA (Clontech), and mouse embryonic brain total RNA (E17; Zyagen Laboratories; catalog no. MR-201-E17) were used in a nested RT-PCR. Single-stranded cDNAs were generated using Superscript 11 reverse transcriptase (Invitrogen) and random primers according to the manufacturer's protocol and then 1 μl of serial 100-, 50-, 25-, and 12.5-fold dilutions of cDNA, a 5 μM concentration of each primer, and 2 μl of LightCycler FastStart DNA Master PLUS SYBR Green1 (Roche Diagnostics Asia Pacific Pte. Ltd.) were used in a 10-μl total volume. The relative sense-RNA transcript abundance was obtained by calculating the ratio of the fluorescent intensity (cross-point value). Each sample was normalized on the basis of its ß-actin content. Real-time quantitative PCR experiments were performed with a Roche Lightcycler instrument.

The cycling conditions were as follows: 95°C for 10 min, 95°C for 10 s, annealing dependent on the primer temperature, and 72°C for 10 s. Melting curve analysis was performed depending on the primer annealing temperature using the Lightcycler software supplied with the instrument.

PCR products were visualized on a 2% agarose gel to confirm amplicon sizes prior to quantification.

Northern Blot Analysis

We used human adult and fetal multiple tissue Northern (MTN) blots from Clontech (catalog nos. 636818 and 636803) in an attempt to detect PCDH antisense-strand transcription. 5'-End-labeled 50-mer oligonucleotide probes were used in hybridization. Thirty picomoles of antisense oligonucleotide, 7 μl of [γ-^{32}P]ATP (6000 Ci/mmol; Amersham Biosciences Ltd.), and one tube of Ready-To-Go T4 PNK (Amersham Biosciences Ltd.) were used in final volume of 50 μl. The reaction was incubated for 60 min at 37°C, then stopped by the addition of 5 μl of 250 mM EDTA. The labeled probe was separated from unincorporated nucleotides through Sephadex G25 Quick Spin Columns (Roche Diagnostics Asia Pacific Pte. Ltd.). Specific activity of the probe was quantified by scintillation counting.

Membranes were prehybridized in ExpressHyb Solution at 42°C for 2 h. Denatured probes were added to ExpressHyb Solution (0.73 × 107 cpm/ml) and the membranes were hybridized overnight at 42°C. Membranes were rinsed a few times in 2× SSC and 0.05% SDS and washed two times at room temperature, followed by two washings in 0.1× SSC and 0.1% SDS at 42°C. Blots were exposed for 3 days against a PhosphorImaging screen and visualized using a PhosphorImager System (Typhoon 9410; Amersham Biosciences). A human adult MTN blot was stripped by hot water containing 0.5% SDS for 10 min. Then it was used for a control hybridization with a human ß-actin probe. The blot was hybridized at 0.8 × 107 cpm/ml ExpressHyb solution overnight. After washing, PhosphorImager signal detection was performed as above.

Results

In Silico Discovery of Putative Novel Cis-Antisense Transcripts in the Human PCDH Gene Cluster

During manual annotation of the 5q31 PCDH gene cluster, we encountered 12 novel transcriptional units (TUs) supported by EST and/or flcDNA evidence (partial listing; Table 1). Their genomic locations partially overlapped those of PCDH exons, but their genomic footprints (transcriptional unit boundaries and splice junction locations) were distinct from those of the sense-strand PCDH exons which they overlapped. In all 12 cases, this distinction was due to the cis-antisense orientation of the novel transcripts relative to the PCDH cluster.

Three lines of evidence supported the antisense orientation of the novel TUs relative to PCDH genes. First, their canonical polyadenylation signals (e.g., AATAAA) and canonical splice donor and acceptor sites (GT-AG) resided on the strand opposite to that encoding the PCDH exons. Figure 1 highlights the 3′ end of a representative antisense transcript (anti-PCDHβ3), including the antisense-strand polyadenylation signal and the genomic footprint difference with sense, as visualized in SeqHelp (Lee et al. 1998).

Table 1. Characterization of flcDNA- and EST-supported novel transcriptional units cis-antisense to human PCDH exons

Human PCDH gene or exon	Mouse ortholog	Human sense strand expression	Mouse sense strand expression	Number and list of human antisense ESTs	Number of mouse antisense ESTs	Number of exons in the human antisense transcript
PCDHα1 variable exon	Pcdhα1	+	+	1 flcDNA BC019862	0	1
PCDHα9 variable exon	Pcdhα7 variable exon	+	+	10 ESTs AI247431, AA437139, AI015406, AW236832, BF510424, AI632155, BE502295, AI968755, BG719192, AA927013	0	1
PCDHα11 variable exon	Pcdhα11 variable exon	+	+	1 EST BI562026	0	1
PCDHα12 variable exon	no one-to-one orthologue	+	not applicable	2 ESTs AW134813, BG150305	0	1
PCDHα13 variable exon	Pcdhα12 (Wu et al. 2001)	+	+	1 EST CN428072	0	2
PCDHαc1 variable exon	Pcdhαc1 variable exon	+	+	1 EST AA772180	0	1
PCDHβ1 single-exon gene	Pcdhβ1 single-exon gene	+	+	1 EST BX097466	0	2
PCDHβ3 single-exon gene	Pcdhβ3 single-exon gene	+	+	1 flcDNA BC016751 and 6 ESTs BG773164, AI825548, AW183407, AW136319, AI633930, AI990795	0	3
PCDHβ10 single-exon gene	Pcdhβ10 single-exon gene	+	+	1 EST BX113296	0	2
PCDHβ16 single-exon gene	Pcdhβ16 single-exon gene	+	+	4 ESTs AI684744, AI376824, AW079889, AA453283	0	1
PCDHψ5 single-exon unprocessed β-class pseudogene	no one-to-one ortholog	−	not applicable	17 ESTs BE463846, AI623196, AW294155, AI969803, AI652903, BE672046,AI954650, AW025845, AI219898, AW593514, AW241532, AI692192, AA437196, AA296664, AA757142, AA904724, AI911128	not applicable	3
PCDHψ3 unprocessed γ-class pseudogenic variable exon	Pcdhγb8 variable exon	−	+	2 ESTs AW070271, BX281564	0	1

Fig. 1. Antisense-strand canonical polyadenylation signal and distinct genomic footprint of a novel human transcriptional unit on the negative strand of PCDHβ3.

Second, the orientation of these polyadenylation signals and splice sites universally conformed to submitter-indicated transcription orientation of the ESTs and flcDNAs comprising the novel TUs; for example, AATAAA, or an associated consensus variation, was found within the 50 bp nearest the submitter-indicated 3′ end of the ESTs and flcDNAs, suggesting that there was no artifactual reversal of transcript sequences in GenBank/dbEST. Finally, BL2SEQ and Spidey pairwise alignments of ESTs and flcDNAs comprising the novel TUs against known PCDH exons were invariably in the antisense (plus/minus) orientation.

For initial assessment of interspecies conservation of PCDH antisense transcription, we identified the true orthologues or nearest homologues (Wu et al. 2001; Vanhalst et al. 2001) of the 12 human antisense-overlapped PCDH exons in the mouse and manually curated all EST-to-genome alignments corresponding to the mouse exons and adjacent genomic sequences. No evidence of antisense-strand transcription was seen in the mouse, suggesting that antisense transcripts at these specific locations are not conserved (Table 1).

Comparative Sequence Analysis of PCDH Cis-Antisense Transcripts

To further investigate the possibility that human PCDH cis-antisense transcripts are not evolutionarily conserved, we focused on the three transcripts with the greatest extent of EST support: anti-PCDHα12, anti-PCDHβ3, and anti-PCDHψ5—all of which are putatively noncoding. Anti-PCDHα12, anti-PCDHβ3, and anti-PCDHψ5 encode the longest ORFs, sized at 121, 49, and 149 amino acids, respectively. The ORFs contain no conserved domains and no similarities to any known proteins outside of low-complexity regions.

Since splice sites and polyadenylation signals are major contributors to transcript structure and boundary definition, the absence of these sequence elements in a nonhuman species would indicate either a major interspecies difference in antisense transcript structure or a lack of the antisense transcript. To determine the extent to which these elements are conserved in mammalian genomic sequences and to estimate the time at which they first arose in evolution, we searched for antisense-strand splice sites and polyadenylation signals at orthologous genomic locations in one nonhuman great ape (chimpanzee), one old world primate (rhesus macaque), and mouse. Results are summarized in Fig. 2.

All three cis-antisense transcripts were characterized by canonical polyadenylation signals in human. AATAAA is the most frequent polyadenylation signal in mammals, while AGTAAA and CATAAA are acceptable variants of the broader polyadenylation hexamer consensus, occurring at frequencies of 2.83% and 1.82%, respectively, in the FANTOM2 mouse cDNA collection (Carninci et al. 2003). All signals were fully conserved in chimpanzee and rhesus, with the exception of the rhesus genomic location equivalent to the human anti-PCDHα12 AGTAAA polyadenylation signal, which contained a strongly noncanonical AGTACA (the C is supported by a Q40 peak in TraceDB accession 331289929 and also remains in the January 2005 rhesus genome assembly at UCSC). The full conservation in chimpanzee, however, implies that the whole-genome shotgun assembly

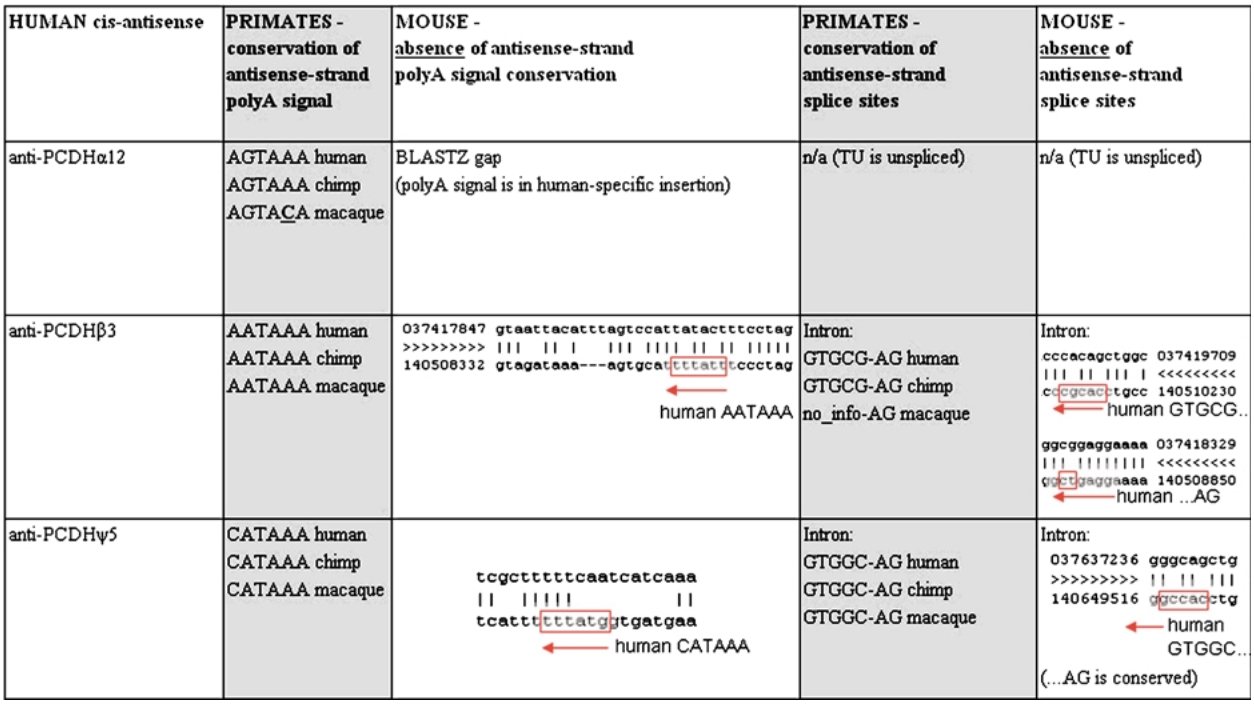

Fig. 2. Conservation of canonical polyadenylation signals and splice junctions on the PCDH antisense strand.
Columns 3 and 5: mouse is at top, human is at bottom, and short alignments are excised out of a substantially longer context of one-to-one sequence-level orthology represented by the pairwise sequence alignment underlying the UCSC Chained BLASTZ Alignments track. Arrows indicate direction of transcription of the human PCDH cis-antisense transcripts; boxes indicate consensus poly(A) signals and splice sites, all of which are on the reverse strand of alignments shown. Column 4: "no info" denotes splice site located in genomic sequence not currently covered in the *Macaca mulatta* division of Trace DB.

method used to derive the chimpanzee genomic sequence, while potentially less accurate than the BAC/PAC tiling path method used to assemble the human genome (for an in-depth discussion see Green 2002), was not problematic for this analysis.

In contrast, no antisense-strand polyadenylation signals were found in mouse. The AGTAAA near the 3′ end of anti-PCDHα12 localized to a human-specific insertion in the global human/mouse BLASTZ alignment. Although the sequence containing the AATAAA of anti-PCDHβ3 was found in the mouse, two of the six bases differed between human and mouse due to single-nucleotide substitutions, thereby completely abolishing the polyadenylation signal consensus in mouse. Similarly, three of the six bases of the CATAAA polyadenylation signal utilized by human anti-PCDHψ5 diverged from the polyadenylation signal consensus in mouse.

Although all human anti-PCDHα12 EST evidence indicates a single-exon transcript, the other two human antisense transcripts were spliced, allowing an assessment of splice donor and acceptor conservation in homologous nonhuman sequences. The splice donor/splice acceptor sequence of the intron used by all anti-PCDHβ3 transcripts except AI633930 was GTGCG-AG and completely conserved in chimpanzee. The splice acceptor was conserved in rhesus as well, although the splice donor lacked genomic sequence coverage at the time of analysis due to incompleteness of the rhesus trace archive. The anti-PCDHψ5 transcript had two introns, with alternative splicing within the second. The long second intron variant, represented most frequently in anti-PCDHψ5 ESTs, had a splice donor (GTGGC) and splice acceptor (AG) completely conserved at orthologous locations in both chimpanzee and rhesus.

However, splice site conservation of PCDH antisense-strand transcriptional units did not extend to mouse. The GTGCG splice donor and AG splice acceptor utilized by human anti-PCDHβ3 did not exist in mouse due to two single-base substitutions in the donor and one in the acceptor sequence. Two of the five bases of the human anti-PCDHψ5 splice donor were substituted in mouse with nonconsensus bases as well. Thus, multispecies comparison of PCDH antisense transcript splice sites and polyadenylation signals within orthologous sequence context demonstrates conservation of these sequence elements in the primate genomes we considered but not in mouse.

Experimental Validation of Primate-Specific PCDH Cis-Antisense Transcripts Coexpressed with Corresponding Sense Exons in Brain

To test the hypothesis that PCDH cis-antisense transcripts are primate-specific, and to validate the

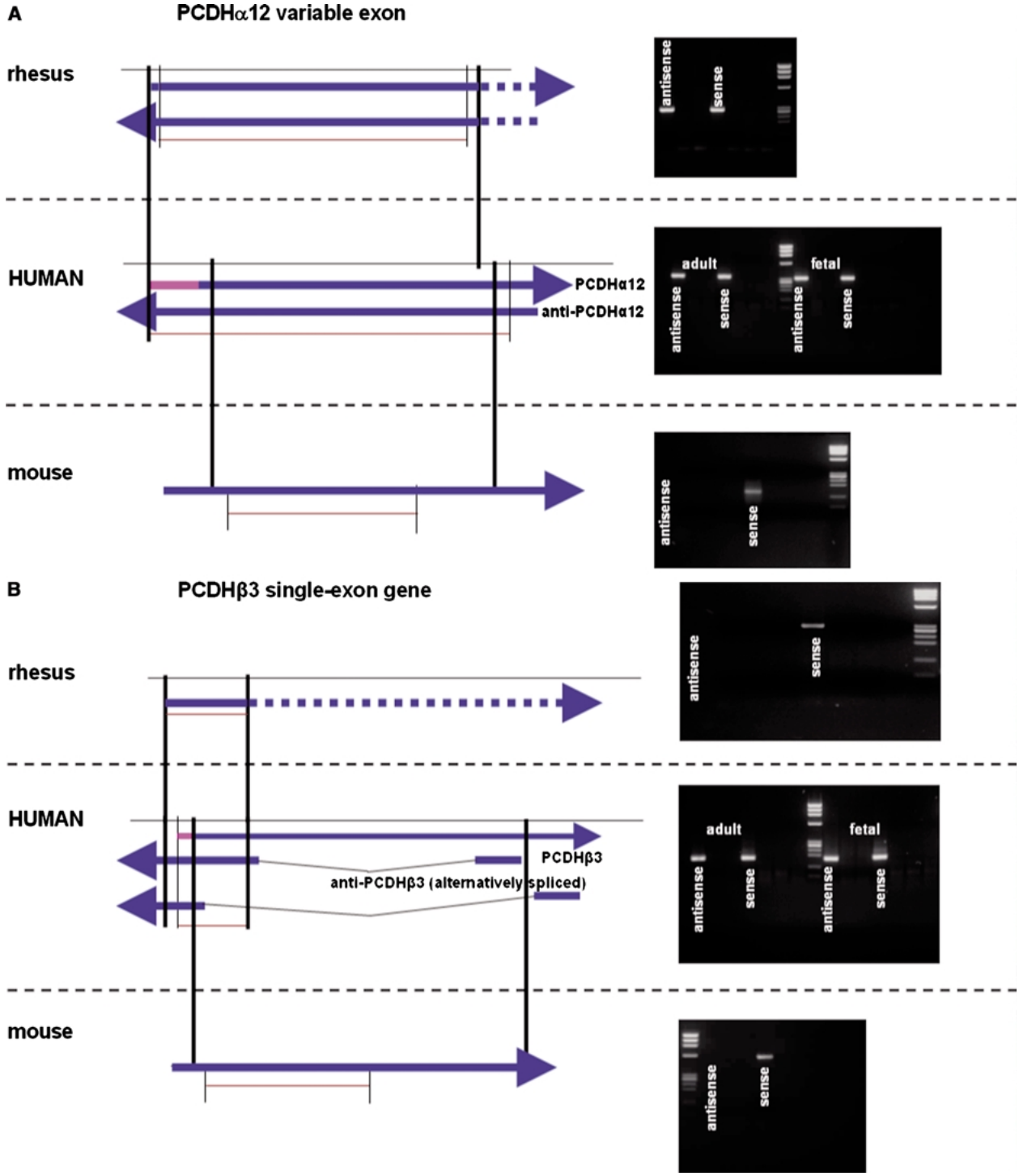

Fig. 3. Multispecies analysis of genomic structure and transcriptional activity of targeted portions of the protocadherin gene cluster.
A The human PCDHα12 variable exon and orthologous regions in rhesus and mouse. **B** The human PCDHβ3 single-exon gene and orthologous regions in rhesus and mouse. **C** The human PCDHψ5 single-exon β-class unprocessed pseudogene; the orthologous region in rhesus; and the mouse Pcdhβ15 gene, whose two closest primate homologues are PCDHψ5 and PCDHβ15. **D** The human PCDHβ15 single-exon gene; the orthologous region in rhesus; and the mouse Pcdhβ15 gene, whose two closest primate homologues are PCDHψ5 and PCDHβ15. SSRTPCR only. QPCR amplicons are *not* shown. **Left** side of each module (genomic structure): thick black dashed horizontal lines separate species within a gene grouping. For genes where both genomic DNA and transcripts are shown, the transcripts are indicated by horizontal arrows pointing in the direction of transcription below genomic DNA. Solid arrows indicate transcripts, or portions of transcripts, documented by our sequenced RTPCR products and/or by public flcDNA/EST evidence. Dotted arrows indicate portions of transcripts which are inferred to exist based on sequence homologies, but which are outside of our sequenced RTPCR products and lack public flcDNA/EST support. Thin black solid horizontal lines are genomic DNA sequences (human except for PCDHβ15, which is rhesus) and flcDNA sequences (mouse and human PCDHβ15). Interspecies thick vertical lines demarcate genomically equivalent sequence positions, based on one-to-one orthology for all genes except mouse Pcdhβ15 and its partners, in which case they are based on one-to-two homology to the primate genes.

(Continued on page 81)

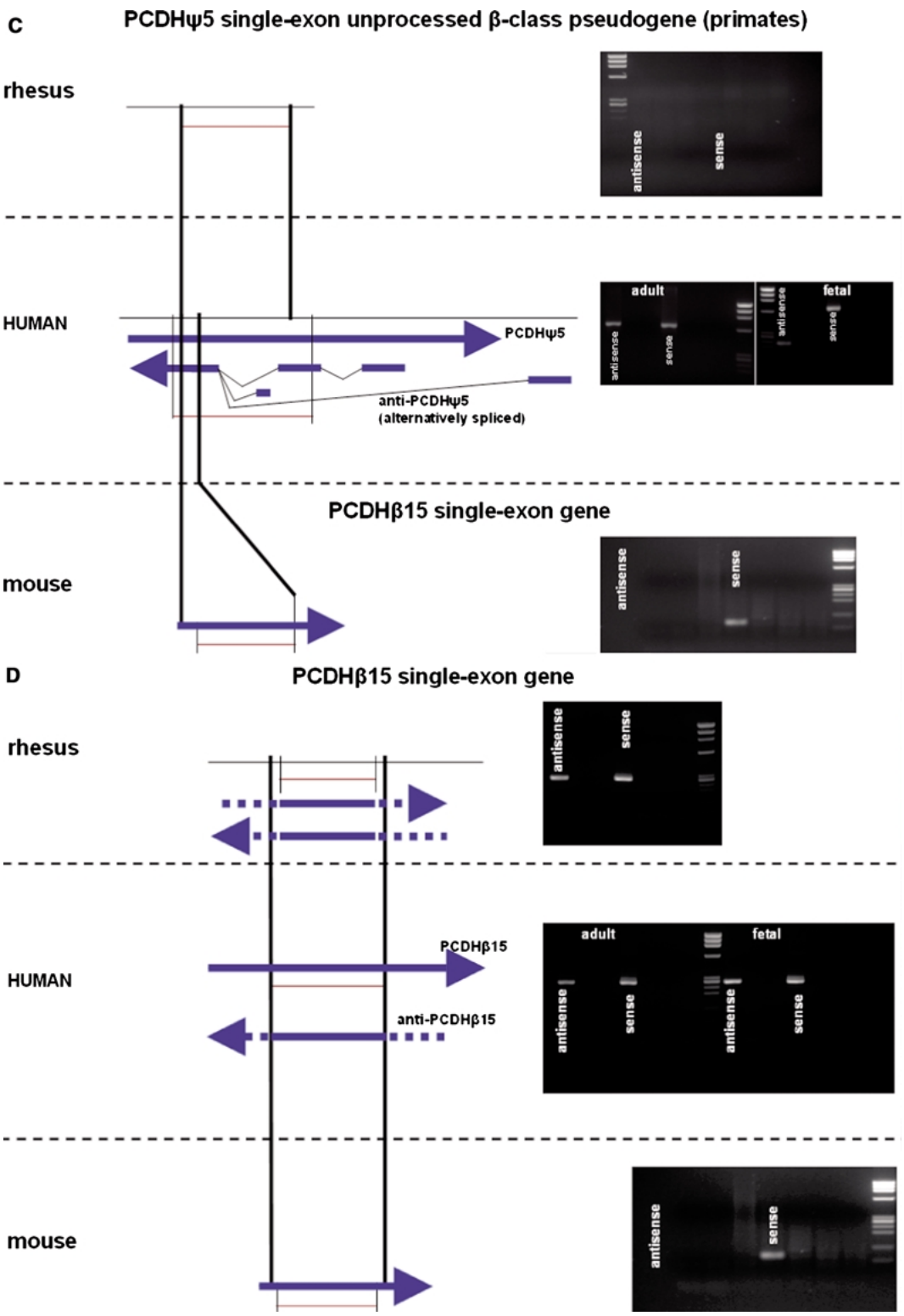

Fig. 3. (Continued from page 8). Primary RTPCR amplicons are shown as thin horizontal lines bounded by vertical lines. See supplementary information for accession numbers, sequence coordinates, and nested amplicon locations. Not to scale. **Right** side of each module (SSRTPCR results): all PCR products on gels are nested. Identities of all products were confirmed by sequencing. (All chromatograms are on file; data not shown.) ''Antisense'' is a nested PCR product obtained after gene-specific, strand-specific, single-primer RT with antisense-specific RT primer. ''Sense'' is a nested PCR product obtained after gene-specific, strand-specific, single-primer RT with sense-specific RT primer. Controls on gels, left to right, are as follows. (1) Mock RT. + RT –primer. Outer and nested PCR as usual. Shows no contamination of RT and buffer with genomic DNA. (2) Mock RT. –RT + primer. Outer and nested PCR as usual. Shows absence of genomic DNA in starting RNA sample. (3) Mock RT. –RT –primer. Outer and nested PCR as usual. Shows no contamination of primer aliquots with genomic DNA. Human and rhesus ''antisense'' lanes are followed by 1 and 2; ''sense'' lanes, by 1, 2, and 3. Mouse ''antisense'' and ''sense'' lanes are both followed by 1, 2, and 3. Templates: human—adult brain and fetal brain total RNA; rhesus—adult brain total RNA; mouse—pooled whole-body adult and fetal total RNA. Additional details pertaining to this figure are given in the supplementary information.

expression of the sense and antisense transcripts in brain, we performed gene-specific, strand-specific RT followed by nested PCR and sequencing in an attempt to detect the anti-PCDHα12, anti-PCDHβ3, PCDHα12, and PCDHβ3 transcripts in human, rhesus, and mouse and anti-PCDHψ5 and PCDHψ5 in human and rhesus. In addition, we performed the same orientation-specific transcription assay on the mouse Pcdhβ15 locus, which is in a one-to-two homologous relationship with the human PCDHβ15 and PCDHψ5 genes (Vanhalst et al. 2001), and on the human and rhesus PCDHβ15 genes, even though they did not have antisense-strand flcDNAs or ESTs.

Figure 3 summarizes the genomic structure of all PCDH loci in all species vis-à-vis the sense and antisense transcript exon-intron structures. Exon-intron structures were determined by curation of EST-to-genome alignments, prior to experimental validation. For human, antisense-specific RT primers were designed from sequences which, based on flcDNA and EST evidence, were exonic with respect to the antisense but not the sense transcripts. In addition, the orientation-specific nature of our single-primer RT reactions (see Materials and Methods) assures the strand specificity of results.

If antisense transcript sequence and structure are conserved in a nonhuman primate, then antisense transcript boundaries and splice sites can be predicted from the human sequence. We refer to these aligned nonhuman positions as positional equivalents of the human genomic structure elements. For rhesus, orthologous regions were localized, and primers were designed within these positionally equivalent spans. The position equivalencies are indicated by vertical lines in Fig. 3. For mouse, genomic sequence conservation appeared limited to sense-strand exon boundaries. Therefore, mouse SSRTPCR amplicons covered solely portions of the Pcdh exons which were positionally equivalent to antisense-covered portions of human PCDH exons.

SSRTPCR results are presented in Fig. 3 to the right of the genomic diagrams. Transcription in both directions was detected in human adult and fetal brain for the unspliced PCDHα12 amplicon. Though the primers were initially designed solely for antisense-specific SSRTPCR, the sense signal indicates that the sense-strand TSS is located at least 63 bp upstream of its previously reported location at bp 16625 of AC005609.1. Transcription in both directions was detected in adult rhesus brain as well. However, no antisense-specific signal was detected in the mouse total-body fetal and adult sample. Therefore, the cis-antisense transcript overlapping the PCDHα12 variable exon is primate-specific and is coexpressed with PCDHα12 in brain.

Similarly, transcription of PCDHβ3 was observed in both directions in human adult and fetal brain total RNA samples. The presence of the sense-strand signal suggesting that the PCDHβ3 TSS is at least 136 bp upstream of its previously known location at bp 78155 of AC005754.1. In contrast with PCDHα12, no antisense transcription could be detected in the rhesus region equivalent to the 3' terminal exon of the human anti-PCDHβ3 transcript. The detection of sense PCDHβ3 transcript in the rhesus suggests that, as in human, the rhesus PCDHβ3 TSS is upstream (at least 341 bp) of the location predicted based on the human TSS defined by the full-length cDNA AF217755. Consistent with our interpretation of the multispecies sequence alignment showing primate specificity of the cis-antisense transcripts, no antisense transcription was seen in the mouse equivalent of the antisense-overlapped portion of human PCDHβ3 (Fig. 3B).

Human PCDHψ5 is an unprocessed single-exon ß-class PCDH pseudogene, originating from an ancient tandem duplication within the PCDHB cluster, and is known to be transcribed in both sense (Vanhalst et al. 2001) and antisense (this study) orientations. We also show sense-strand transcription of PCDHψ5 in human fetal and adult brain. As expected for a PCDHB-class gene, the transcript is unspliced. Although numerous ESTs in this locus suggest the existence of a spliced cis-antisense transcript, antisense-specific SSRTPCR shows a spliced product of the appropriate size only in fetal brain, whereas an unspliced antisense transcript is found in adult brain. The corresponding region of the rhesus PCDHψ5 was not transcriptionally active in adult brain in either orientation (Fig. 3C).

Multiple attempts to detect PCDHα12, PCDHβ3, and PCDHψ5 antisense strand transcription in human by Northern blotting were unsuccessful with both Integrated DNA Technologies and standard T4 polynucleotide kinase labeling protocols, using both Clontech and U.S. Biologicals fetal and adult multitissue blots. Since an ACTB control Northern blot was successful, PCDH cis-antisense transcript levels are likely below the threshold of detection by Northern analysis but detectable by SSRTPCR.

To determine whether the paralogous PCDHβ15 and PCDHψ5 transcriptional units, originating from a duplication in the primate lineage after the primate-rodent divergence, both possess cis-antisense transcripts, we applied our SSRTPCR to human PCDHβ15 and to the orthologous rhesus sequence. Sense and antisense transcripts were detected in both fetal and adult human brain. However, they were also detected in adult rhesus brain, even though rhesus brain appeared to lack PCDHψ5 transcription. Therefore, expression profile differences in brain between two paralogous protocadherin genes, PCDHψ5 and PCDHβ15, may exist between human and rhesus.

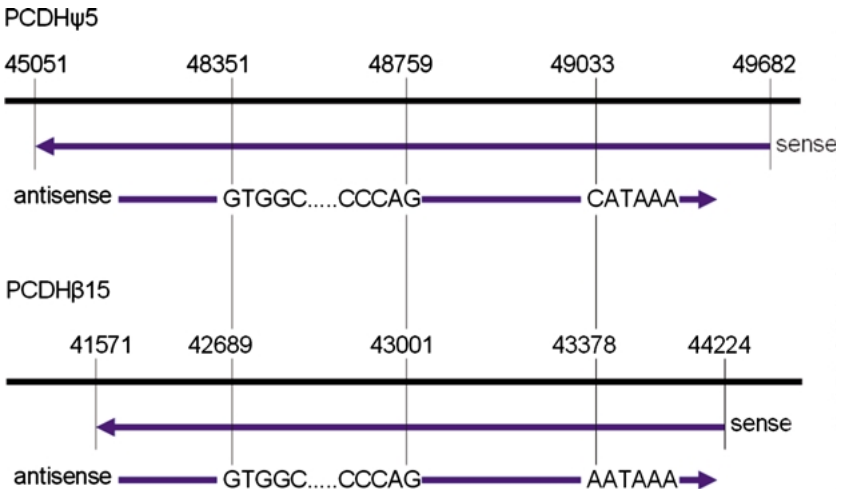

Fig. 4. Antisense-strand splice site (GT-AG) and polyadenylation signal (_ATAAA) conservation between human PCDHψ5 and human PCDHβ15. All genomic coordinates are on AC005752.1. Genomic DNA is at top. Transcribed sequence and transcription direction are below. GT...AG text indicates introns. Not to scale. The splice site and polyadenylation signal conservation is observed in the following one-to-one paralogous pairwise sequence alignment context: 47590–48988 vs 41921–43330, 85% identity, 1% in gaps; 49043–49434 vs 43386–43777, 73% identity; 49528–49760 vs 43881–44113, 71% identity.

		HUMAN BRAIN		RHESUS BRAIN	MOUSE TOTAL BODY
		fetal	adult	(adult)	(pooled fetal & adult)
PCDHα12	**sense** (known gene)	+	+	+	+
	antisense (human ESTs seen)	+	+	+	-
PCDHβ3	**sense** (known gene)	+	+	+	+
	antisense (human ESTs seen)	+	+	-	-
PCDHψ5	**sense** (known gene)	+	+	-	gene not in mouse
	antisense (human ESTs seen)	+	+	-	gene not in mouse
PCDHβ15	**sense** (known gene)	+	+	+	+
	antisense (no ESTs)	+	+	+	+

Fig. 5. Summary of protocadherin sense and antisense expression in human, rhesus, and mouse.

We detected PCDHβ15 antisense transcripts using both SuperScript II and the much more stringent tagged primer, exonuclease I-ThermoScript RT system. Sequence comparison of PCDHψ5 and PCDHβ15 demonstrates that antisense-strand canonical splice sites specifying anti-PCDHψ5 major isoform intron 2, as well as the antisense-strand canonical polyadenylation signal of anti-PCDHψ5, are conserved on the antisense strand of human PCDHβ15 (Fig. 4). This conservation of key genomic structure elements of the cis-antisense transcriptional units between paralogues is consistent with our detection of PCDHβ15 cis-antisense transcription. This suggests that cis-antisense arose at the ancestral PCDHβ15 locus prior to the PCDHβ15-PCDHψ5 gene duplication.

As expected from lack of sequence conservation, no murine antisense transcription could be detected in total-body pooled RNA in the homologous Pcdhβ15 region (Fig. 3C). This supports the emergence of cis-antisense transcription at this locus after the primate/rodent divergence. Our qualitative assessment of sense- and antisense-strand transcription of PCDHα12, PCDHβ3, and PCDHβ15 in the brain in all three species, as well as PCDHψ5 in human and rhesus, is summarized in Fig. 5.

Cis-Antisense Transcription Is Associated with Lower Levels of PCDH mRNA in Quantitative Orthologue Expression Comparisons

Our study is the first to report protocadherin sense expression quantitation in mammalian brains. To test for a correlation between the level of a sense transcript and the presence of its cis-antisense partner in the PCDH endogenous antisense system, we used a Roche LightCycler to quantitatively compare the expression levels of each sense transcript (PCDHα12, β3, and β15) across the three species (Fig. 6), taking into account the presence or absence of cis-antisense transcription. The presence of cis-antisense transcripts was visually associated with lower sense expression levels.

To quantify the relationship between cis-antisense incidence and sense expression levels, we considered separately each of the three sets of gene expression measurements from paralogous gene sets (Fig. 6). The maximum observed quantitated expression level within each set was denoted as 100%. No expression was denoted as 0%. For each paralogous gene, the species results were separated into two subsets: those samples from species with SSRTPCR-confirmed cis-antisense transcripts and those without evidence of any cis-antisense expression. Mean quantitated

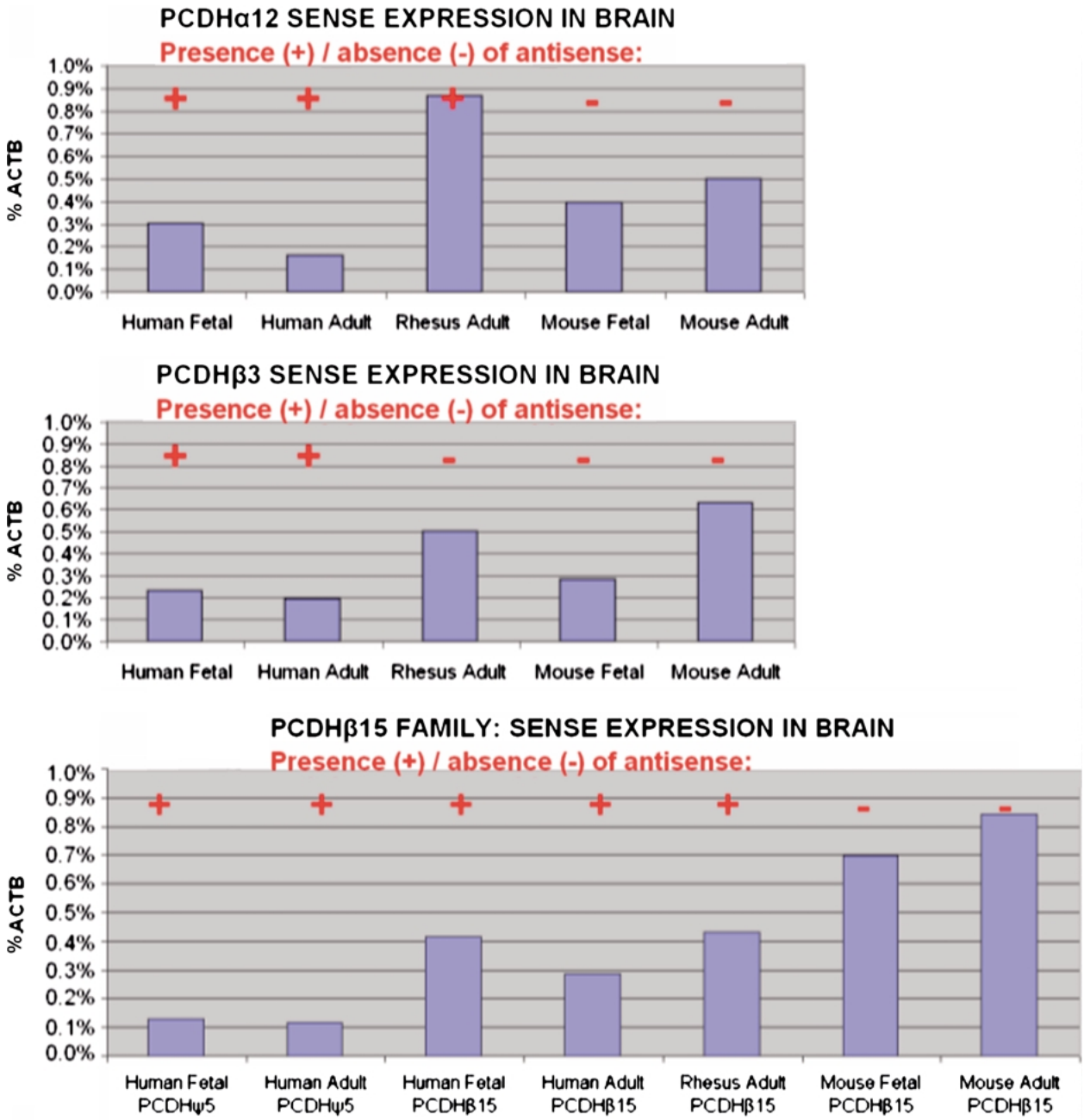

Fig. 6. Quantitated expression levels of sense-strand protocadherin transcripts, as a percentage of the ß-actin transcript levels within the same samples.

expression levels, as percentages of the maximum, were computed for each subset (adult and fetal expression levels for the same gene in the same species were considered different data points). For all three gene sets, the mean expression levels in the presence of cis-antisense were two to three times lower than the levels in the absence of cis-antisense (Table 2). This shows a strong trend toward lower sense transcript levels in the presence of a cis-antisense moiety.

The 10 expression level measurements in samples showing expression of cis-antisense transcripts were further compared to the 7 measurements taken from loci and species where cis-antisense coexpression was not detected. The difference between the lower PCDH sense expression levels in the presence of cis-antisense and higher sense expression levels in the absence of cis-antisense was statistically significant (*t*-test for two independent samples: $p = 0.038$). Taken together, these results suggest an inverse relationship between levels of cognate sense and antisense transcripts through evolution: i.e., in the presence of a cis-antisense transcript, the level of the sense transcript is reduced.

Table 2. Percentage means of PCDH sense expression levels with and without endogenous cis-antisense

Orthologous/homologous gene group	Mean expression level without cis-antisense (% of max.)	Mean expression level with cis-antisense (% of max.)
PCDHα12[a]	52%	27%
PCDHβ3	75%	34%
PCDHβ15/PCDHψ5	91%	33%

[a]RhesusPCDHα12 was considered to be an aberrant outlier and was not used in means calculation.

Putative Human-Specific PCDHψ5 Expression in Brain

PCDHψ5 has the highest number of antisense-strand ESTs (17) of any PCDH transcript. This robust cis-antisense transcription may account for its low level relative to that of the three PCDH genes. In addition, our repeated attempts to detect a PCDHψ5 sense transcript in rhesus adult brain both qualitatively and quantitatively were unsuccessful (data not shown). Therefore, PCDHψ5 transcription in adult brain may take place in human but not rhesus.

Discussion

Identification and Experimental Validation of PCDH Cis-Antisense Transcripts in Human and Rhesus

Despite EST evidence for PCDH cis-antisense transcription, antisense transcripts at this locus had not been validated by orientation-specific RTPCR prior to our study. We demonstrate that cis-antisense transcription in the brain, associated in all cases with simultaneous sense-strand transcription, occurs at four human PCDH exons and at two of the four rhesus orthologues of those exons but does not take place at the mouse counterparts of any of these exons.

Partial Conservation Between Human and Rhesus and Lack of Conservation Between Primate and Mouse Cis-Antisense Transcripts at Orthologous PCDH Exons

We show that the broad framework of protocadherin sequence diversity and divergence between mammalian species is characterized by the birth of novel cis-antisense transcripts after the primate/rodent divergence (Fig. 7). Furthermore, we demonstrate that PCDHβ3 cis-antisense transcription, as well as transcription from both strands of the PCDHψ5 pseudogene, occurs in human, but not rhesus, adult brain. In the case of PCDHβ3, the conservation of the cis-antisense splice sites and polyadenylation signals between human and rhesus suggests that the gene birth predated the human-rhesus divergence, while expression pattern differences between human and rhesus appeared after the divergence, although the possibility that the gene birth was due to a human-specific sequence change cannot be formally excluded.

Although pseudogenes and pseudogenized exons account for just 5 of 58 (9%) of the PCDHβ genes and PCDHα/γ variable exons in the human PCDH gene cluster, 2 of the 8 cis-antisense transcripts (25%) overlap the PCDH pseudogenic elements. We propose an antisense-mediated exon turnover model that can explain the association of cis-antisense transcripts and pseudogenes. This model starts with de novo birth of a robustly transcribed cis-antisense TU at the genomic location of an existing sense-strand PCDH exon as a chance event. By competitive transcriptional interference with the sense strand, and/or by posttranscriptional facilitation of sense mRNA decay, the appearance of anti-PCDH transcripts during evolution could have decreased or eliminated the translation of the corresponding PCDH proteins. This would be an alternative to promoter mutations as an evolutionary stratagem to attenuate the expression of paralogous genes. The initial function of such cis-antisense transcripts, in effect, is to convert a gene into a pseudogene. The utility of the cis-antisense thus evolved is then to suppress the useless transcription of that pseudogene. In the resulting absence of purifying selection on the PCDH sense transcripts, mutations that made their exons pseudogenic would accumulate. Accordingly, any synaptic connection formation patterns specified by those proteins prior to the birth of the antisense transcripts would disappear from the set of possible patterns.

Examination of the mouse Pcdh cluster using the UCSC Genome Browser revealed antisense to the Pcdhβ12 gene and to five Pcdhγ variable exons: a10, a11, a12, b7, and c3. The human orthologues/closest homologues of these mouse Pcdh exons do not match any ESTs in the antisense orientation. EST-supported human anti-PCDH transcripts are limited to the α and β PCDH clusters, whereas mouse demonstrates cis-antisense transcription in the Pcdhγ cluster (entirely unaffected by antisense in human) as well as at Pcdhβ exons whose human equivalents lack antisense. While protocadherin cis-antisense transcripts are thus not a primate-specific phenomenon, cis-antisense transcription in the α and γ portions of the region may have first appeared in primates and rodents respectively after the two lineages diverged.

Cis-Antisense Is Associated with Lower Sense Expression in Orthologue Comparisons

Cis-antisense transcripts have been hypothesized to bind their sense counterparts upon coexpression in

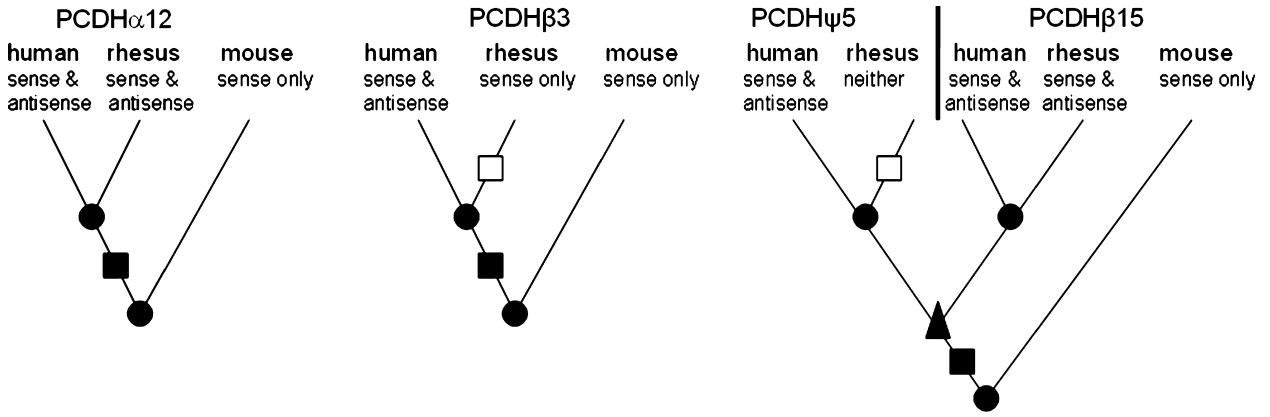

Fig. 7. Evolutionary map of mammalian PCDHα12, PCDHβ3, and PCDHβ15/PCDHψ5 sense and cis-antisense transcription, presented as simplified gene trees.
Filled circles indicate species divergences. Filled squares indicate de novo birth of cis-antisense transcripts. Open squares indicate the loss of cis-antisense expression in brain along an evolutionary lineage. Filled triangle indicates a tandem gene duplication. The origin of PCDHβ15/PCDHψ5 cis-antisense from a single ancestral copy is inferred from the sequence conservation shown in Fig. 4.

the same cell, preventing translation of the protein-coding sense mRNA. Therefore, molar excess of an antisense RNA might be predicted to decrease the effective copy number of the sense RNA in the cell, since only antisense transcripts will remain after most of the sense has been sequestered into RNA duplexes targeted for degradation. Competitive transcriptional inhibition of one member of a sense-antisense pair by another might also be expected to cause the levels of the two to be inversely related. Together, these arguments comprise the basis of the "sense-high, antisense-low" hypothesis, stipulating that, if antisense functions by downregulating sense, then the levels of the two should be inversely related.

Our results were mostly consistent with this hypothesis. In all three orthologue comparisons (PCDHα12, PCDHβ3, and PCDHβ15), expression level of the same orthologue relative to the intraspecies ACTB standard was higher in mouse than in human (i.e., the Pcdhα12/Actb ratio in mouse was higher than the PCDHα12/ACTB ratio in human), allowing for the possibility of sense transcript depletion in human by the primate-specific cis-antisense. In adult brain, the lowest PCDHβ3 sense expression level was seen in human, consistent with the hypothesis that human-specific PCDHβ3 cis-antisense downregulates the sense. For PCDHβ15, the mouse is the only species lacking cis-antisense at that locus. Consistent with our sense-high, antisense-low hypothesis, the mouse exhibited the highest sense expression level. In aggregate, percentage mean expression levels in tissues without cis-antisense were approximately two- to threefold higher than in tissues expressing cis-antisense ($p = 0.038$). Altogether, our combined qualitative and quantitative PCR evidence supports an inverse correlation between sense expression levels and the presence of antisense transcripts in the mammalian PCDH system.

Biological Significance of PCDH Cis-Antisense Transcription

Cis-antisense transcription covering variable or alternative exons within extended combinatorial gene clusters may be a widespread phenomenon not limited to protocadherin genes. Extensive cis-antisense transcription over the V segments of the mouse immunoglobulin heavy chain region has been recently demonstrated (Bolland et al. 2004). In this report, we document novel cis-antisense TUs within the conserved PCDH gene cluster in diverged mammalian lineages and show evidence for quantitative regulation of gene expression by cis-antisense transcripts. We also demonstrate that an antisense-mediated regulatory mechanism arose at specific exons after the primate-rodent divergence. Such a mechanism would be consistent with species-specific evolutionary pressures on PCDH genes (Vanhalst et al. 2001), might provide a recent parallel to the vertebrate-specific origin of the protocadherins (Frank and Kemler 2002), and suggests a role for cis-antisense in the regulation of synaptic plasticity in human brain (Cheng et al. 2002). In view of the potential coexpression of sense and antisense transcripts from this locus in primate brains, as well as the apparently recent evolutionary origin of PCDH cis-antisense transcription, the antisense transcripts merit consideration as factors contributing to the complexity of primate brains and behaviors. Our work identifies those cis-antisense components that can be experimentally manipulated in functional studies using transgenic models.

Acknowledgments. This work was funded in its entirety by the Agency for Science, Technology, and Research, Republic of Singapore, through Genome Institute of Singapore budget no. GIS/03-114102. We thank Zhang Tao for fruitful discussions of SSRTPCR artifacts and the ThermoScript protocol and Sanjay Gupta and Jane Thomsen for expert assistance with Northern blotting.

References

Blin-Wakkach C, Lezot F, Ghoul-Mazgar S, Hotton D, Monteiro S, Teillaud C, Pibouin L, Orestes-Cardoso S, Papagerakis P, Macdougall M, Robert B, Berdal A (2001) Endogenous Msx1 antisense transcript: in vivo and in vitro evidences, structure, and potential involvement in skeleton development in mammals. Proc Natl Acad Sci USA 98:7336–7341

Bolland DJ, Wood AL, Johnston CM, Bunting SF, Morgan G, Chakalova L, Fraser PJ, Corcoran AE (2004) Antisense intergenic transcription in V(D)J recombination. Nat Immunol 5:630–637

Carninci P, Waki K, Shiraki T, Konno H, Shibata K, Itoh M, Aizawa K, Arakawa T, Ishii Y, Sasaki D, Bono H, Kondo S, Sugahara Y, Saito R, Osato N, Fukuda S, Sato K, Watahiki A, Hirozane-Kishikawa T, Nakamura M, Shibata Y, Yasunishi A, Kikuchi N, Yoshiki A, Kusakabe M, Gustincich S, Beisel K, Pavan W, Aidinis V, Nakagawara A, Held WA, Iwata H, Kono T, Nakauchi H, Lyons P, Wells C, Hume DA, Fagiolini M, Hensch TK, Brinkmeier M, Camper S, Hirota J, Mombaerts P, Muramatsu M, Okazaki Y, Kawai J, Hayashizaki Y (2003) Targeting a complex transcriptome: the construction of the mouse full-length cDNA encyclopedia. Genome Res 13:1273–1289

Chen J, Sun M, Kent WJ, Huang X, Xie H, Wang W, Zhou G, Shi RZ, Rowley JD (2004) Over 20% of human transcripts might form sense–antisense pairs. Nucleic Acids Res 32:4812–4820

Courseaux A, Nahon JL (2001) Birth of two chimeric genes in the Hominidae lineage. Science 291:1293–1297

Frank M, Kemler R (2002) Protocadherins. Curr Opin Cell Biol 14:557–562

Green P (2002) Whole–genome disassembly. Proc Natl Acad Sci USA 99:4143–4144

Hastings ML, Milcarek C, Martincic K, Peterson ML, Munroe SH (1997) Expression of the thyroid hormone receptor gene, erbAalpha, in B lymphocytes: alternative mRNA processing is independent of differentiation but correlates with antisense RNA levels. Nucleic Acids Res 25:4296–4300

Imamura T, Yamamoto S, Ohgane J, Hattori N, Tanaka S, Shiota K (2004) Non-coding RNA directed DNA demethylation of Sphk1 CpG island. Biochem Biophys Res Commun 322:593–600

Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12:656–664

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. Genome Res. 12:996–1006

King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. Science 188:107–116

Kiyosawa H, Yamanaka I, Osato N, Kondo S, Hayashizaki Y (2003) Antisense transcripts with FANTOM2 clone set and their implications for gene regulation. Genome Res 13:1324–1334

Lavorgna G, Dahary D, Lehner B, Sorek R, Sanderson CM, Casari G (2004) In search of antisense. Trends Biochem Sci 29:88–94

Long M, Deutsch M, Wang W, Betran E, Brunet FG, Zhang J (2003) Origin of new genes: evidence from experimental and computational analyses. Genetica 118:171–182

Misra S, Crosby MA, Mungall CJ, Matthews BB, Campbell KS, Hradecky P, Huang Y, Kaminker JS, Millburn GH, Prochnik SE, Smith CD, Tupy JL, Whitfied EJ, Bayraktaroglu L, Berman BP, Bettencourt BR, Celniker SE, de Grey AD, Drysdale RA, Harris NL, Richter J, Russo S, Schroeder AJ, Shu SQ, Stapleton M, Yamada C, Ashburner M, Gelbart WM, Rubin GM, Lewis SE (2002) Annotation of the *Drosophila melanogaster* euchromatic genome: a systematic review. Genome Biol 3:RESEARCH0083 Epub

Morishita H, Kawaguchi M, Murata Y, Seiwa C, Hamada S, Asou H, Yagi T (2004) Myelination triggers local loss of axonal CNR / protocadherin alpha family protein expression. Eur J Neurosci 20:2843–2847

Murata Y, Hamada S, Morishita H, Mutoh T, Yagi T (2004) Interaction with protocadherin-gamma regulates the cell surface expression of protocadherin-alpha. J Biol Chem 279:49508–49516

Nemes JP, Benzow KA, Moseley ML, Ranum LP, Koob MD (2000) The SCA8 transcript is an antisense RNA to a brain-specific transcript encoding a novel actin-binding protein (KLHL1). Hum Mol Genet 9:1543–1551

Noonan JP, Li J, Nguyen L, Caoile C, Dickson M, Grimwood J, Schmutz J, Feldman MW, Myers RM (2003) Extensive linkage disequilibrium, a common 167-kilobase deletion, and evidence of balancing selection in the human protocadherin alpha cluster. Am J Hum Genet 72:621–635

Olson MV, Varki A (2003) Sequencing the chimpanzee genome: insights into human evolution and disease. Nat Rev Genet 4:20–28

Podlowski S, Bramlage P, Baumann G, Morano I, Luther HP (2002) Cardiac troponin I sense-antisense RNA duplexes in the myocardium. J Cell Biochem 85:198–207

Prescott EM, Proudfoot NJ (2002) Transcriptional collision between convergent genes in budding yeast. Proc Natl Acad Sci USA 99:8796–8801

Schwartz S, Kent WJ, Smit A, Zhang Z, Baertsch R, Hardison RC, Haussler D, Miller W (2003) Human-mouse alignments with BLASTZ. Genome Res 13:103–107

Shendure J, Church GM (2002) Computational discovery of sense-antisense transcription in the human and mouse genomes. Genome Biol. 3:RESEARCH0044 Epub

Shibata S, Lee JT (2004) Tsix transcription- versus RNA-based mechanisms in Xist repression and epigenetic choice. Curr Biol 14:1747–1754

Stuart JJ, Egry LA, Wong GH, Kaspar RL (2000) The 3' UTR of human MnSOD mRNA hybridizes to a small cytoplasmic RNA and inhibits gene expression. Biochem Biophys Res Commun 274:641–648

Tatusova TA, Madden TL (1999) Blast 2 sequences – a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Lett 174:247–250

Thenie AC, Gicquel IM, Hardy S, Ferran H, Fergelot P, Le Gall JY, Mosser J (2001) Identification of an endogenous RNA transcribed from the antisense strand of the HFE gene. Hum Mol Genet 10:1859–1866

Tran N, Raponi M, Dawes IW, Arndt GM (2004) Control of specific gene expression in mammalian cells by co-expression of long complementary RNAs. FEBS Lett 573:127–134

Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet 34:157–165

Vanhalst K, Kools P, Vanden Eynde E, van Roy F (2001) The human and murine protocadherin-beta one-exon gene families show high evolutionary conservation, despite the difference in gene number. FEBS Lett 495:120–125

Vanhee-Brossollet C, Vaquero C (1998) Do natural antisense transcripts make sense in eukaryotes? Gene 211:1–9

Veeramachaneni V, Makalowski W, Galdzicki M, Sood R, Makalowska I (2004) Mammalian overlapping genes: the comparative perspective. Genome Res 14:280–286

Verona RI, Mann MR, Bartolomei MS (2003) Genomic imprinting: intricacies of epigenetic regulation in clusters. Annu Rev Cell Dev Biol 19:237–259

Wang X, Su H, Bradley A (2002) Molecular mechanisms governing Pcdh-gamma gene expression: evidence for a multiple promoter and cis–alternative splicing model. Genes Dev 16:1890–1905

Wheelan SJ, Church DM, Ostell JM (2001) Spidey: a tool for mRNA-to-genomic alignments. Genome Res 11:1952–1957

Wu Q, Zhang T, Cheng JF, Kim Y, Grimwood J, Schmutz J, Dickson M, Noonan JP, Zhang MQ, Myers RM, Maniatis T (2001) Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. Genome Res 11:389–404

Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MM, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR (2003) Empirical analysis of transcriptional activity in the Arabidopsis genome. Science 302:842–846

Yelin R, Dahary D, Sorek R, Levanon EY, Goldstein O, Shoshan A, Diber A, Biton S, Tamir Y, Khosravi R, Nemzer S, Pinner E, Walach S, Bernstein J, Savitsky K, Rotman G (2003) Widespread occurrence of antisense transcription in the human genome. Nat Biotechnol 21:379–386

Supplementary information for Materials and Methods

I. Qualitative PCR

a. conditions

Optimization using genomic DNA:

Human, Macaca mulatta and mouse genomic DNAs were used to optimize the PCR conditions for each amplicon prior to attempting RTPCR. All amplifications were nested. 500 ng of genomic DNA and 5 μM outer primers were used in the first-round PCR reaction for a total volume of 25 μl, and then 2 μl of firs- round PCR product and 10 μM nested primers were used in second-round PCR. Taq DNA polymerase (Promega) was used for all PCR reactions. Cycle conditions were: denaturation at 96°C for 4 min; 35 or 40 cycles of (96°C for 1 min, annealing temperature based on the primers, 72°C for 1 min); and final extension at 72°C for 10 min.

The optimized annealing temperatures were as follows:

	PCDHβ3		PCDHα12		PCDHβ15		PCDHψ5	
	Step 1	Step 2	Step 1	Step 2	Step 1	Step 2	Step 1	Step2
Mouse	59	52	57	59	57	56	-	-
Rhesus	59	58	53	53	53	53	58	53
Human	57	62	60	64	48	52	57	62

b. primer sequences

Human PCDHβ3:
outer forward 5′-GCCAGCAAGTCTGAGCCTCTG-3′
outer reverse 5′-TCTTGACTCGGACCCAGCCAG-3′
nested forward 5′-CAAGTCTGAGCCTCTGGA AAG-3′
nested reverse 5′-GCCTTTGTCTAAGAAATCG CTCTC-3′

Rhesus PCDHβ3:
outer forward 5′-CTCCTCCAGTAGCGTCAAC CAG -3′
outer reverse 5′-TCTTGACTCGGACCCAGCCAG-3′
nested forward 5′-CAACCAGGGTTGATAAGAAT AAACG-3′
nested reverse 5′-TTCTAGGGAATGTCTCCAAGC-3′

Mouse Pcdhβ3:
outer forward 5′-CTGGGTGGGTCTCTTGCT GGG-3′
outer reverse 5′-TCCCGTATCCAAGTCAGAG GC-3′
nested forward 5′-GAAGAACTGGCTGAAAG GAGAG-3′
nested reverse 5′-TCTGGGGCATTATCATTT GTG-3′

Human PCDHα12:
outer forward 5′-ATACCTCAGGCAAGCGATC CC-3′
outer reverse 5′-GTTCCCTTTCTCTGAACACC GGC-3′
nested forward 5′-TGTCCCAACTCAGAGGCC CTC-3′
nested reverse 5′-GCTCCGCCCGCACAGCTTC TC-3′

Rhesus PCDHα12:
outer forward 5′-AGCGTCCTGTCCACTATCA CC-3′
outer reverse 5′-GCAAGCGAACCCTTAAAAC TG-3′
nested forward 5′-GATTTACCTCCAGAAGGT CCC-3′
nested reverse 5′-CCTTAAAACTGACTGTCCC AACTC-3′

Mouse Pcdhα12:
outer forward 5′-CTCCTGCTCTCGTTCCTGCT CCTCA-3′
outer reverse 5′-TTGGACGCCACCCTGAACAAG-3′
nested forward 5′-TCCTGCTCTCGTTCCTGCT CCT-3′
nested reverse 5′-CACCAGCTCCGTCAGCTCC AG-3′

Human PCDHβ15:
outer forward 5′-TCATACTCCCTTTATTACA GCTCTC-3′
outer reverse 5′-TCGGTCTCTAATCCTAAACA G-3′
nested forward 5′-TACTCCCTTTATTACAGCT CTCAGG-3′
nested reverse 5′-CTAAACAGGGCCACTTCTG TC-3′

Rhesus PCDHβ15:
outer forward 5′-TACTCCATTTCTTACAGCTC TCAGG-3′
outer reverse 5′-CGGTCTCGAATCCTAAACA GG-3′
nested forward 5′-TCCATTTCTTACAGCTCTC AGGAG-3′
nested reverse 5′-CTAAACAGGGCCACTTCT GTC-3′

Mouse Pcdhβ15:
outer forward 5′-TCTATTCAGGTGGTGGATG TCAACG-3′
outer reverse 5′-CGAATCCGAAACAGAGCC ACC-3′
nested forward 5′-TCAGGTGGTGGATGTCAA CGA-3′
nested reverse 5′-TCTCTGGTGAATTTTCTGG GA-3′

Human PCDHψ5:
outer forward 5′-CTCTGATGGGGCAGGTCT TTC-3′
outer reverse 5′-CAGCGGGAACAGCACGAAGAG-3′
nested forward 5′-TTCATTCACAAGCCCCATC CG-3′
nested reverse 5′-AGCACGAAGAGTGAGCTG TCG-3′

Rhesus PCDHψ5:
outer forward 5′-CCCTGCAGAGACTCGTAGT CCAG-3′
outer reverse 5′-GTGATGGGGCAGGTCTTTA TG-3′
nested forward 5′-TGCAGAGACTCGTAGTCC AGC-3′
nested reverse 5′-ATGGGGCAGGTCTTTATGG AA-3′

II. Quantitative PCR

a. LightCycler primer annealing temperatures

	PCDHβ3	PCDHα12	PCDHβ15	PCDHψ5	β-actin (ACTB)
Mouse	58	62	59	n/a	62
Rhesus	60	60	56	58	59
Human	59	58	63	58	62

b. primer sequences

Human PCDHβ3:
forward 5′-CAGAAGCCGAGACAGTGACAAT GAC-3′
reverse 5′-GACAGCCCAGATAAACATCCAG GTC-3′

Rhesus PCDHβ3:
forward 5′-GGCGAATTAAGACTGCGTTGGC -3′
reverse 5′-CAAATCCAGGGATGGCTTAGGT TTC-3′

Mouse Pcdhβ3:
forward 5′-AGGCTGCATTGTTTTGAGTTCA TTG-3′
reverse 5′-CACTGATGGTTTCAGGAAGCAA GC-3′

Human PCDHα12:
forward 5′-TGGATCACTCTTGTGTCGTTTTG GA-3′
reverse 5′-CAAATAACCGTCCTGGATGTGAA TG-3′

Rhesus PCDHα12:
forward 5′-CGACACGCTAATCAGGGCAAT G-3′
reverse 5′-GGTGGAAGTTCTGGACGTGAATG AC-3′

Mouse Pcdhα12:
forward 5′-CAGGTGACTTGCTCTTTGACGC C-3′
reverse 5′-TGTCTCGGGCTGTCACCACCAC-3′

Human PCDHβ15:
forward 5′-TCCATGTTTCCACTCGCACCCG-3′
reverse 5′-CACCGTCCACCGCTGTCAAGG-3′

Rhesus PCDHβ15:
forward 5′-TACCCAATATTCTGAGCCAGGA CTC-3′
reverse 5′-GGTGAAAAGTGAGGAAGACAAA ATT-3′

Mouse Pcdhβ15:
forward 5′-AACTGATGGCGGGGGTCTTTC-3′
reverse 5′-GGGATTGGATTGGTGAGCGATG-3′

Human PCDHψ5:
forward 5′-TTGTGTTTGAAATCACAGAGGG AGG-3′
reverse 5′-GCGTCTATCTAGCCCACCTGATC TT-3′

Rhesus PCDHψ5:
forward 5′-CAGGCTGTTGATTTGAAGATTC CAG-3′
reverse 5′-ACCAATTTTTGCAGACACTTCTC CA-3′

Human ACTB:
forward 5′-AGACCTTCAACACCCCAGCC-3′
reverse 5′-GTCACGCACGATTTCCCGCT-3′

Rhesus ACTB:
forward 5′-GGCTCTCTTCCAACCTTCCTTCC-3′
reverse 5′-CCAGACAGCACTGTGTTGGCGTAC-3′

Mouse Actb:
forward 5′-CGAGCACAGCTTCTTTGCAGCT C-3′
reverse 5′-GGAGCCGTTGTCGACGACCAGC-3′

III. Northern blotting – probe sequences
Human PCDHα12 cis-antisense transcript detection:
5′-CAGGCAAGCGATCCCTTAAAACTGATT GTCCCAACTCAGAGGCCCTCATT-3′

Human PCDHβ3 cis-antisense transcript detection:
5′-ACATGAGAAACTCCTCCAGTAGCGTC AACTAGG GTTGATAAGAATAATCG-3′

Human PCDHψ5 cis-antisense transcript detection:
5′-GATGGGGCAGGTCTTTCTGGAAAATGC ACTGTCATAATACAGGTGGTAGA-3′

Human ACTB transcript detection:
5′- TACAGGGATAGCACAGCCTGGATAGC AACGTACATGGCTGGGGTGTTGAA-3′

Supplementary information for figure 2

I. PCDHA12 variable exon
RHESUS
genomic DNA: TraceDB 331289929
amplicon: nt 455 – 61
region alignable to human PCDHA12: 464 – 34
identity to human: 95%

HUMAN
genomic DNA: AC005609.1
amplicon: 16688 – 16214
region alignable to rhesus PCDHA12: 16688 – 16258
5′ end extension of sense gene: 16688 – 16625
region alignable to mouse Pcdha12: 16608 – 16227

MOUSE
cDNA: AY013759
identity to human: 89%
region alignable to human PCDHA12: 18 – 399
amplicon: 23 – 189

II. PCDHB3 single-exon gene
RHESUS
genomic DNA: TraceDB 350010531
identity to human: 95%
region alignable to human PCDHB3: 310 – 743 (same as amplicon)

HUMAN
genomic DNA: AC005754.1
amplicon: 78290 – 78063
region alignable to rhesus PCDHB3: 78496 – 78063
region alignable to mouse Pcdhb3: 78134 – 76124
EST AI633930 (bottom of two isoforms) is the minor isoform of the cis-antisense.

MOUSE
cDNA: AY013785
identity to human: 81%
region alignable to human PCDHB3: 22 – 2032
amplicon: 61 – 816

III. PCDHps5 unprocessed single-exon pseudogene (primates)
RHESUS
genomic DNA: TraceDB 350110836
identity to human: 91%
region alignable to human PCDHps5: 720 – 122 (same as amplicon)

HUMAN
genomic DNA: AC005752.1
amplicon: 48920 – 48192
region alignable to rhesus: 48918 – 48310
region alignable to mouse Pcdhb15: 48918 – 48771
2-exon isoforms of the cis-antisense are minor: ESTs AI692192 and AI911128

MOUSE
cDNA: AF326315
identity to human PCDHps5: 77%
region alignable to human PCDHps5: 959 – 1106
amplicon: 994 – 1106

IV. PCDHB15 single-exon gene
RHESUS
genomic DNA: TraceDB 331319193
identity to human: 95%
region alignable to human PCDHB15: 294 – 581
amplicon: 297 – 580

HUMAN
cDNA: AF217742
amplicon: 832 – 1119 (same as region alignable to both rhesus PCDHB15 and mouse Pcdhb15)

MOUSE
cDNA: AF326315
identity to human PCDHB15: 81%
region alignable to human PCDHB15: 826 – 1113
amplicon: 994 – 1106