A Novel Variant of *Inpp5f* Is Imprinted in Brain, and Its Expression Is Correlated with Differential Methylation of an Internal CpG Island

Jonathan D. Choi,^{1†} Lara A. Underkoffler,^{2†} Andrew J. Wood,¹ Joelle N. Collins,²
Patrick T. Williams,² Jeffrey A. Golden,⁴ Eugene F. Schuster, Jr.,¹
Kathleen M. Loomes,³ and Rebecca J. Oakey^{1*}

Department of Medical and Molecular Genetics, Guy's, King's and St. Thomas' School of Medicine, King's College London, 8th Floor, Guy's Tower, London SE1 9RT, United Kingdom,¹ and Division of Human Genetics,²
Division of Gastroenterology,³ and Department of Pathology,⁴ The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania 19104

Received 20 January 2005/Returned for modification 19 February 2005/Accepted 23 March 2005

Using a tissue-specific microarray screen in combination with chromosome anomalies in the mouse, we identified a novel imprinted gene, *Inpp5f_v2* on mouse chromosome 7. Characterization of this gene reveals a 3.2-kb transcript that is paternally expressed in the brain. *Inpp5f_v2* is a variant of the related 4.7-kb transcript, *Inpp5f*, an inositol phosphatase gene that is biallelically expressed in the mouse. *Inpp5f_v2* uses an alternative transcriptional start site within an intron of *Inpp5f* and thus has a unique alternative first exon. Whereas other imprinted transcripts have a unique first exon located within intron 1 of a longer transcript variant (such as at the *Gnas* and *WT1* loci), *Inpp5f_v2* is the first example of which we are aware in which the alternative first exon of an imprinted gene is embedded in a downstream intron (intron 15) of a transcript variant. The CpG island associated with the nonimprinted *Inpp5f* gene is hypomethylated on both alleles, a finding consistent with biallelic expression, whereas the CpG island present 5' of *Inpp5f_v2* is differentially methylated on the maternal versus paternal alleles consistent with its imprinting status.

Imprinting is the parent-of-origin-specific unequal expression of the alleles of a gene. Such imprinted genes frequently play important roles in mammalian growth and developmental processes (2, 12, 13). In the mouse, more than 70 imprinted genes have been identified (3, 22), although the extent of imprinting in the mouse and human genomes is not yet fully known. The identification of novel imprinted genes is of value in assessing the extent of imprinting, as well as for understanding aspects of epigenetic gene control and mammalian development.

Expression profiling has been used to identify novel imprinted transcripts by using chromosome anomalies in the mouse (7) and by using parthenogenotes versus normally fertilized embryos or androgenotes (16, 20, 23). We describe here an imprinted gene identified on mouse chromosome 7 (7), along with the genomic and epigenetic features characteristic of imprinting.

Through the identification and characterization of imprinted loci such as the *Igf2/H19* region (4, 30), the *Dlk-Gtl2* region (34), and the BWS region (24, 31), it has emerged that imprinted genes are frequently clustered in domains (28) and are associated with differentially methylated regions (DMRs) (9), often coinciding with imprinting control regions (35, 40). Other notable features include their frequent association with

alternatively spliced transcripts (26), CTCF binding sites (14), CpG islands (40), and direct repeats (27, 42). Alternative splicing and antisense transcripts have been particularly well studied at the *Gnas* locus on mouse chromosome 2, which elicits a complex pattern of parent of origin- and promoter-dependent expression (39). Alternative splice forms of the imprinted *Grb10* gene are expressed from either or both alleles in a tissue-specific manner (5), and whether alternative splicing has a mechanistic relationship to imprinting will be better understood by further study.

The novel imprinted gene identified from this microarray screen is a variant of *Inpp5f*, an inositol phosphatase gene that maps to distal mouse chromosome 7 at 116,088,744 to 116,174,591 bp, ca. 14 Mb proximal to the *H19/Igf2* cluster, in a region that shares conserved linkage with human chromosome 10q26. We have cloned and characterized the *Inpp5f_v2* transcript, which is expressed only in the brain, and allele-specific assays have shown that it is expressed only from the paternal allele. Differential methylation of a CpG island has been found between the maternal (methylated) and paternal (unmethylated) alleles consistent with allele-specific expression. The related longer transcript, *Inpp5f* is biallelically expressed in all tested tissues and is associated with a different hypomethylated CpG island.

MATERIALS AND METHODS

The tissue sources, RNA preparation, microarray protocols, and selection of this transcript were as described previously (7), and *Inpp5f_v2* is imprinted gene number 2 from that study. Microarrays were performed for Table 1 as follows. A total of 7 µg of total brain RNA as described in reference 7 was labeled by using standard Affymetrix protocols. Briefly, cDNA was synthesized with the Invitrogen double-stranded cDNA synthesis kit and biotin labeled by using the Enzo

* Corresponding author. Mailing address: Department of Medical and Molecular Genetics, Guy's, King's and St. Thomas' School of Medicine, King's College London, 8th Fl., Guy's Tower, London SE1 9RT, United Kingdom. Phone: 020-7188-3714. Fax: 020-7188-2585. E-mail: rebecca.oakey@genetics.kcl.ac.uk.

† J.D.C. and L.A.U. contributed equally to this study.

TABLE 1. Probe set values

Probe set	Region	Signal log ratio	P	Maternal signal	Paternal signal
1433542_at	3'	-3.6	0.99998	3008.9	38804.5
1447757_x_at	3'	-3.5	0.99997	380.7	5760.0
1442100_at	5'	0	0.111714	297.0	230.7

bioarray labeling kit. Labeled probe was fragmented and hybridized to 430 A&B 2.0 arrays for 16 h by using standard Affymetrix protocols. The arrays were then washed on an Affymetrix fluidics station 450 and scanned on an Affymetrix scanner 3000, and data were extracted by using GCOS software.

5' RACE for full-length clone identification. IMAGE clone 575575 (accession number AA124959) was identified by using Incyte Genomics GEM microarrays. To further characterize this paternally expressed/maternally imprinted brain-specific transcript, 5' RACE (rapid amplification of cDNA ends) was performed to obtain a full-length transcript. Poly(A)⁺ RNA was derived from newborn mouse brain and RACE-ready cDNA was made by using a SMART RACE cDNA amplification kit (Clontech). The RACE primer was designed from the 5'-most end of the AA124959 sequence, 5'-TCCCATGAGTTAGCCAGCTT G-3'. RACE was performed by using the GeneAmp 9700 PCR system (Perkin-Elmer): 5 cycles at 94°C for 5 s and 72°C for 3 min; 5 cycles at 94°C for 5 s, 70°C for 10 s, and 72°C for 3 min; 20 cycles at 94°C for 5 s, 68°C for 10 s, and 72°C for 3 min. The cDNA sequence from *Inpp5f_v2* was compared to genomic sequence derived from a BAC clone identified by screening the Research Genetics genomic BAC library with IMAGE clone 575575 cDNA. The BAC was sequenced with ABI sequencing technology and analyzed by using Sequencher (Genecodes). *Inpp5f_v2* was assigned this name in accordance with the mouse nomenclature committee (Lois Maltis at the Jackson Laboratories).

Allele-specific assays in intersubspecies hybrids. Newborn brain and kidney RNA from C57BL/6J (B6), *Mus mus castaneus* (cast), cast × B6, and B6 × cast intersubspecies hybrids was reverse transcribed into cDNA. Polymorphisms were identified by sequencing *Inpp5f_v2* in the B6 and cast strains. The *Inpp5f_v2* exon 1 polymorphism is at 689 bp of exon 1 and was assayed by amplifying with primers 5'-ACCTAAGTCCGATGGCGTTCTC-3' and 5'-TTTCTATTCTTTC CAGGTCTTCTAGG-3' to distinguish a G in B6 and a T in cast sequence. *Inpp5f* was assayed for imprinting by a polymorphism unique to this longer transcript. This polymorphism was located in exon 5 at 252 bp, is an A in B6 and a G in cast, and is assayed by using the primers 5'-CATACTGCCTGCTGATG GAGTCAC-3' and 3'-AGAAAAGGAGAACTGGAGCGG-3'.

Northern blot analysis. A MessageMap Northern blot (Stratagene) of adult BALB/c mouse tissues was probed, stripped, and rehybridized with probes specific for different regions of *Inpp5f_v2*. The probe for exon 6 was generated with primers 5'-TTCAGAAGAGTCCAGCAGAACCC-3' and 5'-CCATATTCAG GATGACTGCCTG-3'. The probe for exons 2 to 4 was generated with primers 5'-AAGTGCTGCTGCTGCTGTCTAAC-3' and 5'-TTGCCATCTTCTTCAG GACTACG-3'. An actin probe was used as a loading control. The probe for exons 11, 12, and 13 of *Inpp5f* (long transcript) was generated from primers 5'-AACCTCACGGATGCCATTTC-3' and 5'-TCACCCTTTAGAGCAGCAG

TCC-3'. The probe for exons 13 and 14 was generated from primers 5'-TGAC TCCATCAGCAGGCAGTATG-3' and 5'-CTGTAGGCATCCTTGAACCGA C-3'. A Northern blot with mouse brain and kidney RNA was probed with an exon 1-specific probe generated from primers 5'-ATGCGACCATTGTCTCCG TG-3' and 5'-CATTCTGAAAAGTGTCTGCTTGAGC-3'.

Bioinformatics. Expressed sequence tags (ESTs) were identified from the NCBI database, the TIGR database, and the UCSC genome browser. The mouse genomic sequence for *Inpp5f_v2* analyzed was 116,165,000 to 116,175,000 nucleotides (UCSC May 2004 assembly, NCBI build 33). The human genomic sequence obtained for chromosome 10 was nucleotides 121,566,025 to 121,579,332 (NCBI build 35) determined by the Vistaplot program. The mouse sequence was assembled from RACE product sequencing and from the ESTs with accession numbers BB639524, BB646689, AW561896, and BE305393. The mouse genomic sequence was obtained by sequencing a BAC clone and from the UCSC genome browser sequences in the region of the ESTs listed above.

Programs used for sequence analysis. CpG plot (<http://bioweb.pasteur.fr/sequal/interfaces/cpgplot.html>) was used to report the incidence of CpG islands. The parameters used are a CG content of >50%, an observed/expected ratio of >0.6, and a window of 200 bp with a minimum length of 250 bp. A Vista plot (<http://www.gsd.lbl.gov/vista/index.shtml>) was used for visualizing global DNA sequence alignments (10). The genome assemblies and regions used were 116,165,000 to 116,175,000 bp for chromosome 7 from the May 2004 (mm 5) NCBI build 33 of the mouse genome and 121,566,025 to 121,579,332 bp for chromosome 10 from the human May 2004 (hg 17) NCBI build 35 of the human genome prealigned with SLAGAN (6). Exon 1 was shaded manually because Vistaplot autoscaling does not currently contain the novel murine *Inpp5f_v2* gene. The sequence used for the CTCF binding site search was CCGCNGG NGNC (25, 41), which was analyzed with the nucleic acid pattern search tool FUZZNUC in EMBOSS (<http://ngfnblast.gbf.de/EMBOSS>). Direct repeats were detected by using the Repeatmasker2 software from the University of Washington (<http://repeatmasker.genome.washington.edu/>).

Section in situ hybridization. Section in situ hybridization was performed on fixed CD1 mouse embryos harvested from 12.5 to 16.5 days postcoitum (dpc). The mouse embryos were fixed in 4% paraformaldehyde and embedded in paraffin wax, and sections were cut 7 µm thick. Radioactive riboprobes were transcribed in the presence of ³⁵SUTP. Sense (control) and antisense (test) probes were generated for the test gene. The EST IMAGE:575575 was used as the template for the probe cloned into the EcoRI and NotI sites of a modified pT7T3 vector. The region of probe was exon 6 of *Inpp5f_v2*, and thus the signal represents both long and short transcripts. The antisense template was linearized with EcoRI and transcribed with T3 polymerase. The sense template was linearized with NotI and transcribed with T7 polymerase. The protocol for radioactive section in situ hybridization that was followed was essentially as described in references 29 and 38. The slides were counterstained with Hoechst and the signal, visualized by using fluorescence microscopy, was photographed.

Bisulfite mutagenesis and sequencing. The methylation status of CpG dinucleotides within the two CpG islands associated with the promoters of *Inpp5f* (CpG1) and *Inpp5f_v2* (CpG2) (Fig. 1) were investigated in tissues from the F₁ progeny of C57BL/6J-Ei (B6) mothers and *Mus mus castaneus* (cast) fathers at 8 weeks of age. A total of 0.5 µg of B6 × cast genomic DNA was digested with

FIG. 1. Schematic representation (not drawn to scale) of mouse *Inpp5f_v2* and *Inpp5f*. (a) Mouse *Inpp5f_v2* with six exons, including a unique first exon located within an intron of *Inpp5f*. (b) Full-length *Inpp5f* with 20 exons. (c) The positions of two CpG islands and the SAC domain are indicated. Putative CTCF binding sites in *Inpp5f_v2* are indicated by horizontal bars.

FIG. 2. Allele-specific assays of *Inpp5f_v2* and *Inpp5f*. The allele-specific assay uses intersubspecies hybrids to confirm *Inpp5f_v2* imprinting in brain and biallelic expression of *Inpp5f* in brain and kidney. (a) A polymorphism was detected in the unique exon 1 of *Inpp5f_v2* at 665 bp (arrow in B6 and cast traces) and assayed by PCR and sequencing to distinguish a G in B6 and a T in the cast sequence in brain. *Inpp5f_v2* is maternally imprinted/paternally expressed in brain (arrow in BXC and CXB traces). (b) *Inpp5f* was assayed for imprinting by a polymorphism unique to full-length *Inpp5f* in exon 5 at 126 bp (arrow in B6 and cast traces). The SNP is an A in B6 and a G in cast. Brain and kidney were assayed by using PCR and sequencing and, in both cases, biallelic expression was detected (BXC and CXB traces). (c) RT-PCR assay of exon 1 of *Inpp5f_v2* indicates amplification from brain but not kidney. Lane 1, marker; lane 2, blank; lane 3, *Inpp5f_v2* cast $\times$ B6 brain; lane 4, *Inpp5f_v2* cast $\times$ B6 brain –RT; lane 5, cast $\times$ B6 kidney; lane 6, cast $\times$ B6 kidney –RT; lane 7, blank; lane 8, positive control cast $\times$ B6 brain; lane 9, positive control cast $\times$ B6 kidney.

EcoRI and converted with 3.25 M sodium meta-bisulfite in 0.93 M hydroxyquinone essentially as described in reference 8. Desulfonated converted DNA was amplified with primers specific for CpG1 and CpG2 (Fig. 1). A region from CpG1 associated with the nonimprinted *Inpp5f* transcript was amplified in a nested PCR amplification with the primers forward 5'-TATAGTTTTAYGTTG GGGAGG (where Y is a mix of C and T) and reverse 5'-AAAAAATACAC TAAAAAATAACC-3' for 35 cycles at 55°C and 1.5 mM MgCl₂, followed by a second round with forward primer 5'-GGTATGGAGTTTATAGTTAA GGAT-3' and reverse primer 5'-CCTAAAAACAAAAACTCCCC-3' for 35 cycles at 55°C and 1.5 mM MgCl₂. The product size is 269 bp with two C/A polymorphisms between B6 and cast at positions 162 and 165, respectively. The primers used for the CpG2 island upstream of the imprinted *Inpp5f_v2* transcript were also nested and were forward primer 5'-TTAGGATTAGAGTATTTGT AGAAA-3' and reverse primer TTTACAAAAAATAACAACCCCACTA-3' for 35 cycles at 55°C and 1.5 mM MgCl₂, followed by forward primer 5'-TTTG GTAGTTTTTGTATTATTAAGT-3' and reverse primer 5'-ACCCCACTAAC ACTTAACCAATAAT-3' for 35 cycles at 55°C and 1.5 mM MgCl₂. This gives a 367-bp product with a T/G (B6/cast) polymorphism at position 36. The PCR products were gel purified by using the QiaEXII (Qiagen) kit and cloned by using the TOPO TA cloning system (Invitrogen). Individual clones were isolated by using a Qiaprep minispin kit and sequenced from T3 and T7 primers by using standard ABI sequencing technology (Big Dye v3.1) to determine the methylation status of the CpG dinucleotides and the parental origin of each strand. Only DNA strands >95% converted were used for analysis. All strands were known to be derived from unique template strands on the basis of either epigenotype or an unconverted cytosine occurring outside a CpG dinucleotide (data not shown).

RESULTS

Tissue-specific imprinting. We have used chromosome anomalies in combination with microarrays to identify a novel imprinted gene on mouse chromosome 7. A cDNA represent-

ing *Inpp5f_v2* (Fig. 1) was differentially expressed in uniparental duplicated proximal chromosome 7 maternal versus paternal brain RNA by using microarray analysis but was not differentially expressed in newborn liver, heart, or kidney and thus was predicted to be tissue-specifically imprinted (7). An allele-specific assay was used to confirm from which allele(s) this gene was expressed. A polymorphism in the unique exon 1 of this gene was used to distinguish the maternal and paternal alleles in mouse intersubspecies hybrids (Fig. 2a) and confirmed that *Inpp5f_v2* is paternally expressed in brain. The related, longer transcript *Inpp5f* is expressed but not imprinted in both brain and kidney. A polymorphism was assayed in exon 5 of *Inpp5f*, which is unique to the larger transcript, this demonstrated biallelic expression in brain and kidney (Fig. 2b). The *Inpp5f_v2* assay in kidney RNA failed to amplify a PCR product (Fig. 2c, lanes 5 and 6) using exon 1-specific primers. Control reverse transcription-PCR (RT-PCR) of an unrelated sequence confirmed the presence of template in both brain and kidney samples (Fig. 2c, lanes 7 and 8). Thus, the *Inpp5f_v2* transcript is not expressed in kidney at levels detectable by RT-PCR.

The variant transcript was characterized by a number of approaches. First, the *Inpp5f_v2* sequence was obtained by using 5' RACE, followed by sequencing; the cDNA and genomic sequences were aligned, and the exon/intron boundaries were assigned and revealed six exons (Fig. 1a and 3a). Although the *Inpp5f_v2* gene may encode a protein, we have no

FIG. 3. Comparison of mouse *Inpp5f_v2* and human *INPP5F_v2* genomic sequences. (a) Schematic diagram of the *Inpp5f_v2* gene with exons 1 to 6 shown in green. (b) CpG island detection with thresholds: CpG > 0.6; % observed/expected ratio GC > 50; and window size, 200 bp. (c) Conservation analysis of human to mouse by using VistaPlot with the mouse genome as the base genome. Conserved regions are shaded as follows: exons in blue, 3' untranslated region in light blue, and conserved noncoding sequence in pink (70% threshold). The feature bar above the graph indicates repeat elements: LTR in pink, SINE in green, DNA in orange, and other repeats in light green. (d) Schematic representation of human transcript variant *INPP5F_v2*, RefSeq NM_198330. (e) A conserved CpG island is also detected in the human region.

direct evidence that a protein is made. However, several upstream ATGs are present in the imprinted transcript, and an in-frame ATG near the 3' end of exon 1 of *Inpp5f_v2* is predicted to code for the entire C-terminal half of *Inpp5f*. The transcripts were further characterized by Northern blot analyses and a brain-specific ~3.2-kb transcript was identified by using probes containing exons 2 to 4 and exon 6 of *Inpp5f_v2* (Fig. 4a and b). An additional, larger transcript (~4.7 kb) was detected on Northern blots in all tissues, including brain (Fig. 4a, b, d, and e). Probes specific to the larger transcript covering exons 11 to 14 detect the larger transcript in all tissues tested but do not detect the shorter transcript, as expected (Fig. 4d and e). Thus, the larger transcript represents full-length *Inpp5f*. Figure 4f shows that exon 1 is specific to brain and absent in kidney. The UCSC genome browser reveals two additional putative brain-specific transcripts; these were not detected by Northern blot analysis.

Bioinformatics analysis. The sensitivity of the microarray screen used to detect differential expression of *Inpp5f_v2* in maternal versus paternal disomies was further validated by analysis of *Inpp5f/Inpp5f_v2* features on a different platform

(Affymetrix Genechips) hybridized with maternal versus paternal duplicated chromosome 7 probes (Table 1). This illustrated the sensitivity of the microarray screen, since 3' features representing both *Inpp5f* and *Inpp5f_v2* detect differential expression between the maternal and paternal duplications, whereas 5' features representing only the nonimprinted *Inpp5f* transcript show a lack of differential expression (Table 1).

Comparative sequence analysis identified a gene in humans with sequence similarity to *Inpp5f* called *hSac2* or *INPP5F* (19). No mouse orthologues of *hSac2/INPP5F* have been previously described. A human-mouse conservation analysis of *Inpp5f_v2* was performed by using VistaPlot. Five exons are highly conserved (Fig. 3c) in a region of conserved linkage between mouse chromosome 7 and human chromosome 10. Exon 1 of *Inpp5f_v2* is unique and begins between exons 15 to 16 of the full-length *Inpp5f* gene (Fig. 1). Although exon 1 shows a slightly lower degree of conservation between mice and humans (Fig. 3c), a high degree of conservation is maintained in parts of the CpG island.

Imprinted genes are frequently associated with CTCF binding sites, and multiple putative CTCF binding site sequences

FIG. 4. Northern blot analyses of transcripts showing the brain specific 3.2-kb *Inpp5f_v2* band and a larger band present in all tissues. Probes used are for exon 6 (a) and exons 2 to 4 (b) of *Inpp5f_v2* and β -actin as a loading control (c), as well as probes from exons 11, 12, and 13 (d) and exons 13 and 14 (e) of the longer *Inpp5f* transcript. Exons 11 to 14 of the *Inpp5f* gene are outside of the *Inpp5f_v2* transcript and show no brain-specific ~3.2-kb band. (f) Exon 1 of *Inpp5f_v2* shows a 3.2-kb band specific to the brain and absent in kidney.

were detected within the CpG island associated with *Inpp5f_v2* by sequence searching (Fig. 1). In addition, a consensus CTCF direct repeat ($n = 2$) is located further 3' between exons 4 and 5 in an intronic sequence. The human *INPP5F_v2* CpG island showed no exact matches to the CTCF consensus but, allowing for two mismatches, 13 consensus sequences were detected in the forward direction and 14 in the reverse orientation. The structure of the human *INPP5F_v2* transcript predicted from Refseq clone NM_198330 (the orthologue of the mouse *Inpp5f_v2* gene) is shown in Fig. 3d and is similar to the mouse.

Spatial and temporal expression. A low-magnification 14.5-dpc embryo parasagittal section (lateral to the midline) illustrates the brain specific expression (Fig. 5a). A higher-power sagittal image (Fig. 5b) shows expression in the lateral cortex. There is strong labeling primarily outside the ventricular (proliferative) zone (VZ) dorsally and ventrally, although some VZ labeling is also seen dorsally and frontally. A high-power view of a 12.5-dpc embryo highlighting the reduced expression in the VZ and strong expression in the maturing regions of the midbrain and hindbrain is shown in Fig. 5c. There is an obvious ventral gap in the expression extending from the VZ that could represent the isthmus. Expression is seen in the hindbrain and in a small part of the tectum. The area that will become the brainstem is strongly labeled, whereas the rhombic lip (more dorsally located precursor of the cerebellum) shows less labeling as does the more rostral portion of the tectum (Fig. 5d). At

14.5dpc, expression is seen throughout the nervous system with slightly less expression in the ventricular zones. Figure 5e shows expression in the dorsal root ganglia lateral to the spinal cord and separated by vertebral bodies. Expression is seen in the tectum with a gradient of expression from the rostral to the caudal region (Fig. 5f). Expression is again primarily outside the ventricular zone, although the decreased expression caudally may be due to less developmental progress. Strong expression is also seen in the ventral midbrain and hindbrain, primarily outside the VZ. At this stage, relatively strong expression is seen in the rhombic lip, adjacent to the choroid plexus. At 16.5-dpc expression is seen in the cerebral cortex, the olfactory bulb, the ganglionic eminence, the thalamus, the hippocampus, and the retina. Slightly less expression is seen in the inferior colliculus compared to the superior colliculus (data not shown). The probe used was from exon 6 of *Inpp5f_v2* so it also reflects expression of *Inpp5f*.

Differential methylation in the brain-specific imprinted transcript. CpG analysis of the mouse genomic sequence identified two CpG islands associated with the *Inpp5f* gene: one near the 5' end of *Inpp5f* (CpG1) and the other located intronically to *Inpp5f*, but immediately 5' of *Inpp5f_v2* (CpG2) (Fig. 1c and 3b). These two regions were investigated as potential DMRs by using bisulfite conversion. Bisulfite sequencing of the CpG1 island shows that both alleles are hypomethylated, as would be expected for a nonimprinted transcriptionally active gene (Fig. 6a). In the brain, CpG2, 5' of *Inpp5f_v2* was differentially methylated between the maternal and paternal alleles. Most of the strands were methylated on the maternal allele, whereas most strands were unmethylated on the paternal allele (Fig. 6b). In kidney, all strands appeared almost completely methylated on the maternal allele. In contrast, on the paternal allele nine strands were unmethylated and six were methylated (Fig. 6c). Statistical analysis showed that the numbers of maternally methylated and unmethylated strands in brain compared to kidney is not statistically significant. A standard goodness-of-fit with Yates' continuity correction gives a χ^2 of 1.341 and a P value of 0.2469. Similarly, the Fisher's exact test for count data, for the small sample size, returns a P value of 0.22, a finding consistent with the χ^2 result.

DISCUSSION

Using chromosome anomalies and microarrays to detect imprinted genes. Several approaches have been devised to discover imprinted genes, and several have been detected by using microarrays to assay either parthenogenetic versus normally fertilized embryos (16) or versus androgenetic embryos (20). A large-scale screen for imprinted transcripts has been performed by using FANTOM cDNAs on RIKEN microarrays (23). Differentially expressed genes were identified, of which a proportion will be nonimprinted genes due to differences in development between androgenetic and parthenogenetic embryos at 9.5 dpc (32, 37). Our experimental design has been mindful of cell type differences and matches developmental stage both for embryo and isolated tissue comparisons. Further, the sensitivity of this screen is underscored by additional microarray experiments that were able to discriminate between the transcriptional variants of the *Inpp5f* gene. Differences in expression between maternal and paternal alleles for 3' fea-

FIG. 5. In situ hybridization of mouse embryo sections from 12.5 and 14.5 dpc. Signal is shown as yellow staining, and nonexpressing cells have a blue appearance. (a) Parasagittal section (fairly far lateral) from 14.5-dpc embryo showing expression primarily in the brain. At the top (which is highlighted in panel b) is the cerebral cortex. (b) Higher-power image of a 12.5-dpc embryo showing strong expression in the lateral cortex but sparing the ventricular zone. (c) High-power view of a 12.5-dpc embryo highlighting the reduced expression in the ventricular zone and strong expression in the maturing regions of the midbrain and hindbrain. (d) Expression is seen in the hindbrain and tectum. (e) Expression is seen in the 12.5-dpc dorsal root ganglia lateral to the spinal cord separated by the vertebral disks (blue, nonstaining cells). (f) Expression is seen in the tectum and largely outside the ventricular zone (14.5 dpc). Strong expression is also shown in the ventral midbrain and hindbrain and rhombic lip. Scale bars: a, 200 μ m; b to f, 100 μ m. VZ, ventricular zone.

tures representing *Inpp5f* and *Inpp5_v2* were detected, whereas similar expression was detected for 5' features representing *Inpp5f* alone (Table 1), further validating this approach.

***Inpp5f_v2*, *Inpp5f*, and *hSac2/INPP5F*.** *Inpp5f_v2* is located on distal mouse chromosome 7, band position 7F3 at 116.1 Mb (UCSC Genome Browser using the Mouse May 2004 assembly), ca. 14 Mb proximal of the *H19/Igf2* cluster. This region of the genome shares conserved linkage with human chromosome 10q26.13 in the region of the human gene *hSac2/INPP5F*. Comparative mouse and human sequence analysis has identified sequence similarity between *Inpp5f_v2* and the *hSac2/INPP5F* gene. No mouse orthologue of *hSac2/INPP5F* has previously been described. *Inpp5f_v2* is a transcriptional variant of *Inpp5f*. Inositol phosphatases are involved in endocytosis and exocytosis at nerve terminals, a process essential for clathrin coat shedding and synaptic vesicle recycling (1), which is important in the developing brain and in postnatal survival. It has been shown that full-length human INPP5F has inositol 4,5-bisphosphatase activity in vitro (19). However, unlike *hSac2/INPP5F*, *Inpp5f_v2* does not contain a Sac domain but rather is similar to the variable C-terminal domain of *hSac2/INPP5F* (Fig. 1). It may be that *Inpp5f_v2* originated independently of the Sac domain of *Inpp5f* and that the larger transcript coopted

the exons of *Inpp5f_v2* as its C-terminal domain, the function of which is unknown.

Transcript size, sequence features, and tissue distribution of *Inpp5f_v2*. Northern analyses reveal a brain-specific 3.2-kb transcript and a larger 4.7-kb transcript present in all tested tissues (Fig. 4). The 3.2-kb *Inpp5f_v2* transcript is imprinted in brain (Fig. 2a) and the longer *Inpp5f* transcript is biallelically expressed in both brain and kidney (Fig. 2b and c). In situ hybridization studies illustrate that *Inpp5f_v2* and *Inpp5f* are expressed in brain structures and the developing nervous system over a range of developmental time points in the mouse embryo (Fig. 5).

Approximately 40 to 60% of the multiexon genes in the mouse and human transcriptomes exhibit alternative splicing (21, 43). This facilitates increased complexity of the mammalian proteome and altered protein function. In many cases, the alternative use of a first exon is regulated by alternative promoters, conferring tissue specificity to the different isoforms of a gene. A recent genome survey reports that >2,000 genes in the mouse and >3,000 genes in humans have alternative first exons spliced to a common 3' transcript (44) and, in many cases, conserved CpG islands are associated with the variable exon. *Inpp5f_v2* uses an alternative transcriptional start site

FIG. 6. Bisulfite mutagenesis of the methylation status of CpG dinucleotides in genomic DNA. Unfilled circles represent unmethylated CpGs, and filled circles represent methylated CpGs; shaded circles indicate CpGs that could not be determined unequivocally. (a) BXC (B6 $\times$ cast) brain DNA with primers specific for CpG1, maternal, and paternal alleles are hypomethylated. (b) BXC (B6 $\times$ cast) brain DNA with primers specific for CpG2, the maternal allele appears hypermethylated and the paternal allele appears hypomethylated. (c) BXC (B6 $\times$ cast) kidney DNA with primers specific for CpG2. The maternal allele is hypermethylated; the paternal allele has some methylated and some unmethylated CpGs.

between exons 15 and 16 of *Inpp5f* and thus is the first example we are aware of in which the unique first exon is embedded within an intron near the 3' end of the gene rather than the selection among first exons seen at the *Gnas* locus (26). At the *Gnas* locus, three CpG islands are associated with three alternative start sites sharing common 3' exons. The transcription of *Gnas* is dependent on the methylation state of the CpG islands, as can be seen in the alternative use of the first exon of *Nesp*, *Gnasxl* (26), or *Gnas1A* (18). In humans, the alternative *WT1* transcript *AWT1* is similarly organized, is paternally expressed, and consists of exons 2 to 10 of the *WT1* gene. However, like at the *Gnas* locus, this transcript uses a novel 5' exon located in the first intron of the nonimprinted *WT1* gene (11).

Differential methylation. Differential methylation has been associated with most of the characterized imprinted genes, such as those in the *H19/Igf2* region (36) and at the *Dlk1/Gtl2* locus (34). In both of these domains, CpGs are hypermethylated on the silent paternal allele and hypomethylated on the expressing maternal allele. The methylation status of two CpG islands associated with *Inpp5f* and *Inpp5f_v2* was determined. The CpG2 is differentially methylated in brain, where the expressing paternal allele is unmethylated and the silent maternal allele is methylated (Fig. 6b). This differentially methylated region is likely to be involved in the imprinted expression of *Inpp5f_v2* in brain. However, three maternal alleles out of thirteen were hypomethylated (Fig. 6b), but no corresponding maternally derived transcription was detected in the allele-

specific RT-PCR assay. Although the in situ hybridization data suggest that expression of the *Inpp5f_v2* transcript varies in different regions of the brain (Fig. 5), this is unlikely to account for these findings since differential methylation is also present in nonexpressing kidney tissue (Fig. 6c), albeit to a lesser extent. One possible explanation is that this DMR exerts its role in the imprinting of *Inpp5f_v2* at an early stage of embryonic development and is not absolutely required for maintenance of the imprint in neonatal and adult tissues. A similar situation has been found in the Angelman/Prader-Willi Syndrome region, where differential methylation established in the germ line and maintained at 7.5 dpc is lost by adulthood (15). Bisulfite sequencing analysis of CpG2 in gametes and early-stage embryos would determine whether this DMR is germ line derived and progressively lost during pre- and postnatal development.

CpG1 is hypomethylated on both parental alleles (Fig. 6a), and *Inpp5f* is biallelically expressed in both kidney and brain. The differential methylation pattern at CpG2 is maintained in tissues where *Inpp5f_v2* is not expressed, suggesting that a brain-specific transcriptional regulation mechanism other than methylation may be involved and possibly imposed in addition to the epigenetic silencing of the maternal allele. For *Inpp5f_v2*, methylation is related to imprinted gene expression but not to tissue specificity; this is observed in the human *NDN* gene, where a promoter-associated CpG island exhibits differ-

ential methylation independently of tissue-specific transcriptional status (17).

One of the elements involved in the control of the reciprocal imprinting of the *H19* and *Igf2* genes is the vertebrate enhancer blocking protein, CTCF, which binds to sites in DNA to block promoter-enhancer interactions (4, 14, 33). Multiple CTCF binding site sequences were detected within the *Inpp5f_v2* CpG2 in mice (Fig. 1). However, functional CTCF binding sites cannot be predicted from sequence alone and provide only an indication of potential binding *in vivo*.

Many imprinted genes are clustered, suggesting that they reside in domains that are epigenetically modified, resulting in the coordinate regulation of more than one imprinted gene. Although the identification of a DMR near *Inpp5f_v2* is consistent with an imprinted domain, one might expect one or more additional imprinted genes to be present. Inspection of flanking sequence in combination with microarray differential expression data in this region revealed a cluster of differentially expressed transcripts 3' to this gene. Further study of this flanking cluster of imprinted genes will shed light on whether *Inpp5f_v2* resides alone or is coordinately regulated as part of a larger region.

ACKNOWLEDGMENTS

This study was supported by PHS grant GM58759 from the National Institutes of Health (R.J.O.) and by The Wellcome Trust (R.J.O.). J.D.C. is supported by an IRG fellowship from the GKT School of Medicine, King's College London, and A.J.W. is supported by a studentship from the Generation Trust.

We thank Colin V. Beechey for the T65H mouse tissues and careful reading of the manuscript. We thank Kathryn Woodfine for Affymetrix GeneChip hybridizations and Trevelyan Menheniott and Reiner Schulz for analysis of Affymetrix microarray data and statistical analysis.

REFERENCES

- Arai, Y., T. Ijuin, M. Itoh, T. Takenawa, S. Takashima, and L. E. Becker. 2001. Developmental changes of synaptotagmin expression in the human cerebrum and cerebellum. *Dev. Brain Res.* **129**:1–9.
- Barlow, D. P. 1995. Gametic imprinting in mammals. *Science* **270**:1610–1613.
- Beechey, C. V., B. M. Cattanch, and R. L. Selley. 2000. Standard ideogram/anomaly breakpoints of the Mouse MRC Mammalian Genetics Unit, Harwell, Oxfordshire, United Kingdom. [Online.] <http://www.mgu.har.mrc.ac.uk/imprinting/implink.html>.
- Bell, A. C., and G. Felsenfeld. 2000. Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene. *Nature* **405**:482–485.
- Blagitko, N., S. Mergenthaler, U. Schulz, H. A. Wollmann, W. Craigen, T. Eggermann, H. H. Ropers, and V. M. Kalscheuer. 2000. Human *GRB10* is imprinted and expressed from the paternal and maternal allele in a highly tissue- and isoform-specific fashion. *Hum. Mol. Genet.* **9**:1587–1595.
- Brudno, M., C. B. Do, G. M. Cooper, M. F. Kim, E. Davydov, E. D. Green, A. Sidow, and S. Batzoglou. 2003. NISC Comparative Sequencing Program LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res.* **13**:721–731.
- Choi, J. D., L. A. Underkoffler, J. C. Collins, S. M. Marchegiani, N. A. Terry, C. V. Beechey, and R. J. Oakey. 2001. Microarray expression profiling of tissues from mice with uniparental duplications of chromosomes 7 and 11 to identify imprinted genes. *Mammalian Genome* **12**:758–764.
- Clark, S. J., J. Harrison, C. L. Paul, and M. Frommer. 1994. High sensitivity mapping of methylated cytosines. *Nucleic Acids Res.* **22**:2990–2997.
- Constancia, M., B. Pickard, G. Kelsey, and W. Reik. 1999. Imprinting mechanisms. *Genome Res.* **8**:881–900.
- Couronne, O., A. Poliakov, N. Bray, T. Ishkhanov, D. Ryaboy, E. Rubin, L. Pachter, and I. Dubchak. 2003. Strategies and tools for whole-genome alignments. *Genome Res.* **13**:73–80.
- Dallosso, A. R., A. L. Hancock, K. W. Brown, A. C. Williams, S. Jackson, and K. Malik. 2003. Genomic imprinting at the WT1 gene involves a novel coding transcript (AWT1) that shows deregulation in Wilms' tumours. *Hum. Mol. Genet.* **13**:405–415.
- DeChiara, T. M., E. J. Robertson, and A. Efstratiadis. 1991. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**:849–859.
- Guillemot, F., T. Caspary, S. M. Tilghman, N. G. Copeland, D. J. Gilbert, N. A. Jenkins, D. J. Anderson, A. L. Joyner, J. Rossant, and A. Nagy. 1995. Genomic imprinting of *Mash2*, a mouse gene required for trophoblast development. *Nat. Genet.* **9**:235–241.
- Hark, A. T., C. J. Schoenherr, D. J. Katz, R. S. Ingram, J. M. LeVorse, and S. M. Tilghman. 2000. CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus. *Nature* **405**:486–489.
- Kantor, B., Y. Kaufman, K. Makedonski, A. Razin, and R. Shemer. 2004. Establishing the epigenetic status of the Prader-Willi/Angelman imprinting center in the gametes and embryo. *Hum. Mol. Genet.* **13**:2767–2779.
- Kobayashi, S., H. Wagatsuma, R. Ono, H. Ichikawa, M. Yamazaki, H. Tashiro, K. Aisaka, N. Miyoshi, T. Kohda, A. Ogura, M. Ohki, T. Kaneko-Ishino, and F. Ishino. 2000. Mouse *Peg9/Dlk1* and human *PEG9/DLK1* are paternally expressed imprinted genes closely located to the maternally expressed imprinted genes: mouse *Meg3/Gtl2* and human *MEG3*. *Genes Cells* **5**:1029–1037.
- Lau, J. C. Y., M. L. Hanel, and R. Wevrick. 2004. Tissue-specific and imprinted epigenetic modifications of the human *NDN* gene. *Nucleic Acids Res.* **32**:3376–3382.
- Liu, J., S. Yu, D. Litman, W. Chen, and L. S. Weinstein. 2000. Identification of a methylation imprint mark within the mouse *Gnas* locus. *Mol. Cell. Biol.* **20**:5808–5817.
- Minagawa, T., T. Ijuin, Y. Mochizuki, and T. Takenawa. 2001. Identification and characterization of a Sac domain-containing phosphoinositide 5-phosphatase. *J. Biol. Chem.* **276**:22011–22015.
- Mizuno, Y., Y. Sotomaru, Y. Katsuzawa, T. Kono, M. Meguro, M. Oshimura, J. Kawai, Y. Tomaru, H. Kiyosawa, I. Nikaido, H. Bono, M. Kadamura, T. Kono, G. A. Morris, P. A. Lyons, M. Oshimura, Y. Hayashizaki, Y. Okazaki, et al. 2002. *Asb4*, *Ata3*, and *Dcn* are novel imprinted genes identified by high-throughput screening using RIKEN cDNA microarray. *Biochem. Biophys. Res. Commun.* **290**:1499–1505.
- Modrek, B., and C. J. Lee. 2003. Alternative splicing in the human, mouse and rat genomes is associated with an increased frequency of exon creation and/or loss. *Nat. Genet.* **34**:177–180.
- Morison, I. M., C. J. Paton, and S. D. Cleverley. 2001. The imprinted gene and parent of origin effect database. *Nucleic Acids Res.* **29**:275–276.
- Nikaido, I., C. Saito, Y. Mizuno, M. Meguro, H. Bono, M. Kadamura, T. Kono, G. A. Morris, P. A. Lyons, M. Oshimura, Y. Hayashizaki, Y. Okazaki, et al. 2003. Discovery of imprinted transcripts in the mouse transcriptome using large-scale expression profiling. *Genome Res.* **13**(Suppl. 6B):1402–1409.
- Paulsen, M., O. El-Maarri, S. Engemann, M. Strodicke, O. Franck, K. Davies, R. Reinhardt, W. Reik, and J. Walter. 2000. Sequence conservation and variability of imprinting in Beckwith-Wiedemann syndrome gene cluster in human and mouse. *Hum. Mol. Genet.* **9**:1829–1841.
- Paulsen, M., S. Takada, N. A. Youngson, M. Behnchaib, C. Charlier, K. Segers, M. Georges, and A. C. Ferguson-Smith. 2001. Comparative sequence analysis of the imprinted *Dlk1-Gtl2* locus in three mammalian species reveals highly conserved genomic elements and refines comparison with the *Igf2-H19* region. *Genome Res.* **11**:2085–2094.
- Peters, J., S. F. Wroe, C. A. Wells, H. J. Miller, D. Bodle, C. V. Beechey, C. M. Williamson, and G. Kelsey. 1999. A cluster of oppositely imprinted transcripts at the *Gnas* locus in the distal imprinting region of mouse chromosome 2. *Proc. Natl. Acad. Sci. USA* **96**:3830–3835.
- Reik, W., and J. Walter. 2001. Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* **2**:21–32.
- Reik, W., and J. Walter. 1998. Imprinting mechanisms in mammals. *Curr. Opin. Genet. Dev.* **8**:154–164.
- Sassoon, D., and N. Rosenthal. 1993. Detection of messenger RNA by *in situ* hybridization, p. 384–406. *In* P. M. Wassarman and M. L. DePamphilis (ed.), *Methods in enzymology*. Academic Press, London, England.
- Schoenherr, C. J., J. M. LeVorse, and S. M. Tilghman. 2003. CTCF maintains differential methylation at the *Igf2/H19* locus. *Nature Genetics* **33**:66–69.
- Smilnich, N. J., C. D. Day, G. V. Fitzpatrick, G. M. Caldwell, A. C. Lossie, P. R. Cooper, A. C. Smallwood, J. A. Joyce, P. N. Schofield, W. Reik, R. D. Nicholls, R. Weksberg, D. J. Driscoll, E. R. Maher, T. D. Shows, and M. J. Higgins. 1999. A maternally methylated CpG island in *KvLQT1* is associated with an antisense paternal transcript and loss of imprinting in Beckwith-Wiedemann syndrome. *Proc. Natl. Acad. Sci. USA* **96**:8064–8069.
- Surani, M. A., S. C. Barton, and M. L. Norris. 1986. Nuclear transplantation in the mouse: heritable differences between parental genomes after activation of the embryonic genome. *Cell* **45**:127–136.
- Szabo, P., S. H. Tang, Rentsendorj, A., G. P. Pfeifer, and J. R. Mann. 2000. Maternal-specific footprints at putative CTCF sites in the *H19* imprinting control region give evidence for insulator function. *Curr. Biol.* **18**:607–610.
- Takada, S., M. Paulson, M. Trevalde, C.-E. Tsai, G. Kelsey, B. M. Cattanch, and A. C. Ferguson-Smith. 2002. Epigenetic analysis of the *Dlk1-Gtl2* imprinted domain on mouse chromosome 12: implications for imprinting control from comparisons with *Igf2-H19*. *Hum. Mol. Genet.* **11**:77–86.
- Thorvaldson, J., J. L. Duran, and M. S. Bartolomei. 1998. Deletion of the *H19* differentially methylated domain results in loss of imprinted expression of *H19* and *Igf2*. *Genes Dev.* **12**:3693–3702.
- Tremblay, K. D., J. R. Saam, R. S. Ingram, S. M. Tilghman, and M. S.

- Bartolomei.** 1995. A paternal specific methylation imprint marks the alleles of the mouse H19 gene. *Nat. Genet.* **9**:407–413.
37. **Wilkins, A. S.** 1993. Genetic analysis of animal development, 2nd ed. Wiley-Liss, New York, N.Y.
38. **Wilkinson, D. G., and M. A. Nieto.** 1993. Detection of messenger RNA by in situ hybridization to tissue sections and whole mounts, p. 361–372. *In* P. M. Wassarman and M. L. DePamphilis (ed.), *Methods in enzymology*. Academic Press, London, England.
39. **Wroe, S. F., G. Kelsey, J. A. Skinner, D. Bodle, S. T. Ball, C. V. Beechey, J. Peters, and C. M. Williamson.** 2000. An imprinted transcript, antisense to *Nesp*, adds complexity to the cluster of imprinted genes at the mouse *Gnas* locus. *Proc. Natl. Acad. Sci. USA* **97**:3342–3346.
40. **Wutz, A., O. W. Smrzka, N. Schweifer, K. Schellander, E. F. Wagner, and D. P. Barlow.** 1997. Imprinted expression of the *Igf2r* gene depends on an intronic CpG island. *Nature* **389**:745–749.
41. **Wylie, A. A., S. K. Murphy, T. C. Orton, and R. L. Jirtle.** 2000. Novel imprinted DLJ1/GTL2 domain on human chromosome 14 contains motifs that mimic those implicated in IGF2/H19 regulation. *Genome Res.* **10**:1711–1718.
42. **Yoon, B. J., H. Herman, A. Sikora, L. T. Smith, C. Plass, and P. D. Soloway.** 2002. Regulation of DNA methylation of *Rasgrf1*. *Nat. Genet.* **30**:92–96.
43. **Zavolan, M., S. Kondo, C. Schonbach, J. Adachi, D. A. Hume, Y. Hayashizaki, T. Gaasterland, et al.** 2003. Impact of alternative initiation, splicing, and termination on the diversity of the mRNA transcripts encoded by the mouse transcriptome. *Genome Res.* **13**(Suppl. 6B):1290–1300.
44. **Zhang, T., P. Haws, and Q. Wu.** 2004. Multiple variable first exons: a mechanism for cell- and tissue-specific gene regulation. *Genome Res.* **14**:79–89.