CHAPTER 2: RECENT ASSEMBLY OF AN IMPRINTED DOMAIN FROM NON-IMPRINTED COMPONENTS

This chapter consists of a research publication from 2006 that examines the evolution of the imprinted region responsible for the Prader-Willi and Angelman syndromes.

*Rapkins RW, ***Hore TA**, Smithwick M, Ager E, Pask AJ, et al. (2006) Recent assembly of an imprinted domain from non-imprinted components. *PloS Genet* **2**: e182.

The search for marsupial orthologues of imprinted genes at the Prader Willi-Angelman Syndrome domain had been ongoing in the Graves laboratory for five years prior to my involvement in the project, and was being pursued by fellow student Robert Rapkins. It had been monumentally unsuccessful, yielding only paralogous sequences. I took the search into monotremes, which proved to be the key to unlock the whole mystery; the genes were hard to isolate because the domain simply was not there in marsupials or monotremes, implying that it was a recent construction in eutherians.

As joint first authors* in this publication, Robert Rapkins and I gathered all materials and performed all experiments. Co-authors assisted with some tissue collection for detecting polymorphisms in tammar, or performed minor hybridisation experiments.

Specifically, I performed all experiments on platypus, including screening and characterisation of the critical *UBE3A-CNGA3* BAC, which was the first indication that the human and mouse arrangement of the PWS-AS region is a recent construct confined to the eutherian lineage. I also extracted and characterised the platypus *UBE3A* cDNA sequence and determined its imprinting status by quantitative primer extension analysis of single nucleotide polymorphisms, which I discovered by screening DNA from several animals.

Robert Rapkins and I jointly performed all bioinformatic analyses, wrote the first draft of the manuscript and generated all figures, which were edited and commented upon by our supervisors Jenny Graves and Janine Deakin and other co-authors. Robert Rapkins, Jenny Graves and I shared liaison with editorial staff.

This paper documents the first instance of an imprinted domain that was constructed by massive genomic rearrangement, from non-imprinted elements. It has had a major impact on the field, because prior to this demonstration the expectation was that all large imprinted domains evolved imprinting at the same time as *IGF2* and *IGF2R*, during the early radiation of therian mammals.

OPEN ACCESS Freely available online PLoS GENETICS

Recent Assembly of an Imprinted Domain from Non-Imprinted Components

Robert W. Rapkins[1☯], Tim Hore[1☯], Megan Smithwick[2], Eleanor Ager[3], Andrew J. Pask[3], Marilyn B. Renfree[3], Matthias Kohn[4], Horst Hameister[4], Robert D. Nicholls[5], Janine E. Deakin[1], Jennifer A. Marshall Graves[1*]

1 Australian Research Council Center for Kangaroo Genomics and Research School of Biological Sciences, Australian National University, Canberra, Australia, **2** Department of Genetics, La Trobe University, Melbourne, Australia, **3** Department of Zoology, University of Melbourne, Melbourne, Australia, **4** Department of Medical Genetics, University of Ulm, Ulm, Germany, **5** Department of Pediatrics, Children's Hospital of Pittsburgh, Pittsburgh, Pennsylvania, United States of America

Genomic imprinting, representing parent-specific expression of alleles at a locus, raises many questions about how—and especially why—epigenetic silencing of mammalian genes evolved. We present the first in-depth study of how a human imprinted domain evolved, analyzing a domain containing several imprinted genes that are involved in human disease. Using comparisons of orthologous genes in humans, marsupials, and the platypus, we discovered that the Prader-Willi/Angelman syndrome region on human Chromosome 15q was assembled only recently (105–180 million years ago). This imprinted domain arose after a region bearing *UBE3A* (Angelman syndrome) fused with an unlinked region bearing *SNRPN* (Prader-Willi syndrome), which had duplicated from the non-imprinted *SNRPB/B′*. This region independently acquired several retroposed gene copies and arrays of small nucleolar RNAs from different parts of the genome. In their original configurations, *SNRPN* and *UBE3A* are expressed from both alleles, implying that acquisition of imprinting occurred after their rearrangement and required the evolution of a control locus. Thus, the evolution of imprinting in viviparous mammals is ongoing.

Citation: Rapkins RW, Hore T, Smithwick M, Ager E, Pask AJ, et al. (2006) Recent assembly of an imprinted domain from non-imprinted components. PLoS Genet 2(10): e182. DOI: 10.1371/journal.pgen.0020182

Introduction

Genomic imprinting refers to the silencing of a gene or region according to its parent of origin. Among vertebrates, imprinting is specific to mammals, with about 83 mammalian genes shown to be imprinted. About half are paternally expressed (that is, the maternally derived allele is suppressed) and half maternally expressed (the paternally derived allele is suppressed) [1]. Many genes subject to imprinting are involved with either developmental disorders or cancer (sometimes both), so understanding how and why genomic imprinting evolved and how it functions is therefore of compelling interest to medicine as well as biology.

Imprinting is an important model system for studying epigenetic regulation—an accelerating field of biology that focuses on how identical DNA sequences are differentially expressed to produce different phenotypes. The molecular mechanism of imprinting resembles X chromosome inactivation in females, another mammal-specific epigenetic phenomenon, suggesting that X inactivation and autosomal imprinting may share a common origin [2,3].

We do not yet understand what selective forces eschewed the benefits of diploidy in favor of parental imprinting, and there are many hypotheses to account for the seemingly perverse evolution of hemizygosity at these loci. Perhaps the most interesting and widely debated is the parental conflict hypothesis [4], now developed into the kinship hypothesis (reviewed [5]), which proposes that imprinting evolved in response to the antagonistic interests of parental genomes.

The origin, as well as the mechanism, of imprinting can be investigated by comparing gene arrangement and expression between divergent species. The observation that genes imprinted in human and mouse *(IGF2, M6P/IGF2R)* are not imprinted in chicken [6,7] implies that imprinting is specific to mammals. Nonetheless, the gene content and arrangement of human imprinted domains is highly conserved in chicken [8] and other vertebrates. For instance, the content and arrangement of coding genes in the human Beckwith-Wiedemann imprinted cluster (including *IGF2, H19, ASCL2, KCNQ1*, and *CDKN1C*) is largely shared with birds and fish, but the non-coding regulatory *H19* RNA is missing, along with large stretches of repetitive sequences and retroelements and several sequences thought to exert local control of imprinting [9,10].

The transition of a region from a non-imprinted state in fish and chicken to an imprinted state in placental mammals could therefore be dissected by comparing orthologous regions with the most divergent mammal groups. Marsupials and monotremes diverged from placental mammals 180 and 210 million years ago (MYA), respectively [11], so fill the 310-MY evolutionary void that separates birds and reptiles from humans and mice. This permits the reconstruction of gene content and arrangement over a long evolutionary period and provides informative sequence comparisons with high

Editor: Wolf Reik, The Babraham Institute, United Kingdom

Received June 23, 2006; **Accepted** September 11, 2006; **Published** October 27, 2006

DOI: 10.1371/journal.pgen.0020182

Abbreviations: BAC, bacterial artificial chromosome; FISH, fluorescence in situ hybridization; ICR, imprint control region; kb, kilobase; MYA, million years ago; PWS-AS, Prader-Willi/Angelman syndrome; snoRNA, small nucleolar RNA

* To whom correspondence should be addressed. E-mail: Jenny.Graves@anu.edu.au

☯ These authors contributed equally to this work.

Synopsis

Humans and other mammals have two copies of the genome. For most genes, both copies are active. However, some genes are active only when they are inherited from the father, others only when inherited from the mother. These "imprinted" genes are clustered in domains that are controlled coordinately. Only mammals show genomic imprinting. It is not understood how or why genes became imprinted during mammalian evolution. The authors used comparisons between humans and the most distantly related mammals, marsupials and monotremes, to discover how one of these imprinted domains evolved. The authors studied an imprinted domain on human Chromosome 15, mutations which cause Prader-Willi and Angelman syndromes (PWS-AS). They discovered that the PWS and AS genes lie on different chromosomes in kangaroos and platypus and are not imprinted. Other imprinted genes in the domain, including the putative control region, are absent from the genome and derived from copies of genes from yet other chromosomes. The arrangement in kangaroos and platypus is present also in the chicken genome, so it must be ancestral. This study concludes that the PWS-AS imprinted region was assembled relatively recently from non-imprinted components that were moved together or copied from all over the genome.

signal to noise ratios [12]. Importantly, these "alternative mammals" represent the transition between egg-laying and viviparous animals. Monotremes lay eggs, like reptiles. Marsupial young are born at an early developmental stage and complete development attached to a teat (often protected in a pouch). These modes of reproductive strategy represent major differences in the level of maternal investment, as well as the ability of paternally derived genes to influence maternal resources.

Genomic imprinting has been demonstrated in marsupials for *IGF2* [6], *PEG1/MEST* [13], and *IGF2R* [14]. However, *IGF2* and *IGF2R* show biallelic expression in monotremes [14,15]. A comparative study, specifically of the non-imprinted *IGF2* in platypus with the imprinted opossum, mouse, and human locus, reveals that the absence of *cis*-acting elements, such as short interspersed transposable elements and an intergenic conserved inverted repeat containing putative CTCF-binding sites, may be important for *IGF2* imprinting [16].

The occurrence of imprinting in marsupials, but not monotremes, would date the emergence of genomic imprinting in vertebrates to after 210 MYA, when therian mammals diverged from the egg-laying monotremes and before the divergence of marsupials and placentals 180 MYA. These limited data are consistent with the hypothesis that imprinting evolved after viviparity, as would be expected if it is selected as a response to parental conflict. However, this important conclusion is rather tenuous, since it is based upon expression data from only three of the ~80 genes imprinted in placental mammals [17].

We have therefore made a detailed comparison of the arrangement and expression of orthologs of another cluster of imprinted genes in the three major mammal groups (placentals, marsupials, and monotremes) and other vertebrates.

Prader-Willi and Angelman syndromes (PWS and AS) are phenotypically distinct disorders associated with abnormalities (usually deletions) of a cluster of imprinted genes on human Chromosome 15q11-q13 [18] that amongst other things, influence feeding behavior. The regulation of imprinted genes in the PWS-AS domain has been studied in detail in humans and mice. The region comprises the AS and PWS domains (Figure 1A), within which deletions cause one or another disease. In the distal AS region lie two genes, *UBE3A* (thought to be solely responsible for AS) and *ATP10A*, both maternally expressed (paternally silenced) in brain [19–22]. The larger, more proximal PWS domain encompasses five paternally expressed (maternally silenced) genes responsible for Prader-Willi syndrome, including *SNRPN*, which encodes the SmN antigen. A large paternal transcript originating upstream of the *SNURF-SNRPN* genes liberates several classes of small nucleolar RNAs (snoRNAs) [23]. The imprinted expression of the AS and PWS domains is orchestrated by a bipartite imprinting control region (ICR) located within a 35-kilobase (kb) region which encompasses the *SNRPN* promoter [24,25]. Splice variants of the *SNURF-SNRPN* transcript, which are anti-sense to *UBE3A*, may provide the regulatory link between the ICR (in the PWS domain) and the imprinted genes of the AS domain [26].

The evolutionary history of the PWS-AS imprinted region is unknown, although recent retrotranspositions into the mouse domain have been noted [27]. We therefore cloned and characterized marsupial and monotreme orthologs of genes in the human 15q11-q13 region. To our astonishment, we discovered that marsupial and monotreme AS and PWS genes lie on different chromosomes, and we used bioinformatic analysis to show that this constitutes the ancestral arrangement. Both genes are biallelically expressed in marsupials and monotremes. Other genes from the PWS-AS region are absent from the marsupial and monotreme genomes. Thus, rearrangement of PWS-AS genes and acquisition of retrotransposed genes and key regulatory elements occurred much later in placental mammals, allowing their coordinate imprinted expression.

Results

We examined the evolutionary origin of genes in and adjacent to the human PWS-AS imprinted region by isolating, mapping, and assessing transcription of their orthologs in marsupials, platypus, chicken, and fish.

Isolation of Marsupial and Monotreme Homologs of PWS-AS Genes

We screened bacterial artificial chromosome (BAC) and cDNA libraries from the model kangaroo *Macropus eugenii* (tammar wallaby) for several human 15q11-q13 genes. cDNA clones containing *GABRB3* and *HERC2* and two BACs containing the AS gene *UBE3A* were confirmed by partial sequencing.

Numerous attempts over several years to isolate other imprinted genes, using PCR amplification, Southern blot analysis, and screening several genomic and cDNA libraries, were consistently unsuccessful. Screening for *MKRN3* identified a hitherto unknown intron-containing source gene makorin *(MKRN1)* [28]. Our conclusion that the marsupial genome lacks *MKRN3* was supported by its absence from the *Monodelphis domestica* (opossum) database. No clones containing tammar *MAGEL2* or *NDN* were ever recovered, and no *MAGEL2* or *NDN* sequences were found in the opossum database.

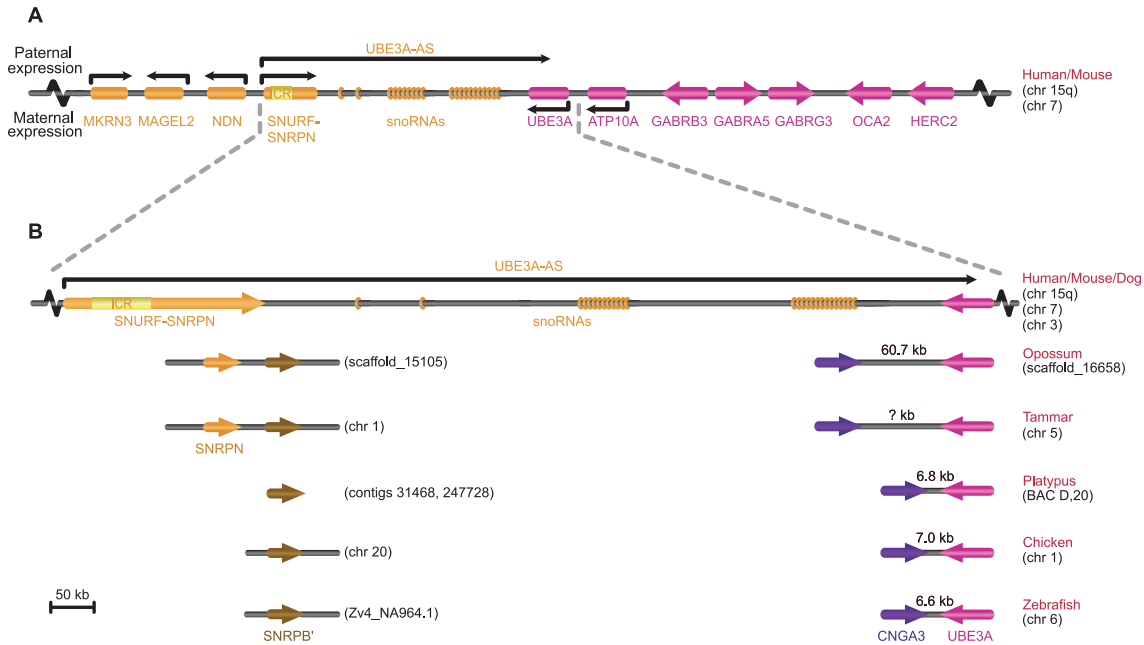

Figure 1. Genes of the PWS-AS Imprinted Domain

(A) The PWS-AS imprinted domain in human and mouse. Paternal or maternal expression is indicated by arrows. The yellow region of *SNURF/SNRPN* represents the ICR. *UBE3A-AS* is an antisense transcript that includes arrays of untranslated snoRNA genes.

(B) Comparison of the eutherian *SNRPN–UBE3A* region with its ancestral arrangement in non-eutherian vertebrates. Pink represents genes that are co-linear in humans through to fish, purple represents human Chromosome 2 genes that are adjacent to the PWS-AS homologous region in the ancestral arrangement, orange represents eutherian-specific genes or elements, and brown the unlinked gene *SNRPB′* that duplicated to form *SNRPN*. In marsupials, monotremes, chicken, and fish, *UBE3A* lies close to a human Chromosome 2 gene *CNGA3*, and there are no snoRNAs. In zebrafish, chicken, and platypus, only *SNRPB′* is present, but in marsupials tandem duplication gave rise to *SNRPN*, which was relocated next to *UBE3A* in eutherians.

DOI: 10.1371/journal.pgen.0020182.g001

Similarly, attempts to clone tammar homologs of the imprinted PWS gene *SNRPN* from a genomic DNA lambda library resulted in the repeated isolation of a non-imprinted paralog *SNRPB/B′* [29], which lies on human Chromosome 20. Subsequently, we discovered a second *SNRPN*-like transcript encoding a predicted gene product that clusters with the SmN proteins, demonstrating orthology to *SNRPN* (Figure S1). A single tammar BAC was then obtained containing both *SNRPN* and *SNRPB′* homologs. Searching the opossum draft assembly for this sequence located a *SNRPN* ortholog directly adjacent to *SNRPB′* on scaffold 15105. Neither this scaffold, nor any other opossum sequence contained sequences with homology to human *SNURF*. We therefore conclude that the marsupial genome contains tandemly arranged *SNRPN-SNRPB′* sequences, and that *SNURF* is absent.

To explore still more ancient arrangements of these genes, we screened a platypus BAC library for PW-AS orthologs. No *SNRPN* homolog was ever obtained. Sequence retrieved from the platypus trace archive confirmed the presence of *SNRPB′*, but not *SNRPN*. A platypus BAC containing *UBE3A* was obtained and its identity confirmed by sequencing. Surprisingly, full sequencing of this BAC also identified the ortholog of a human Chromosome 2 gene, *CNGA3*, 6.8 kb from *UBE3A* in a tail-to-tail arrangement.

To determine whether *UBE3A* is adjacent to *CNGA3* also in marsupials, we screened the two tammar *UBE3A* BACs for *CNGA3*. We identified *CNGA3* in the larger BAC and confirmed its presence by sequencing. We also searched the opossum draft assembly, finding that *UBE3A* and *CNGA3* lie ~60 kb apart in a tail-to-tail arrangement on scaffold 16658, which also contained flanking genes from human Chromosomes 15q and 2 (Table 1).

Mapping PWS-AS Genes in Marsupials

We mapped tammar PWS-AS genes by fluorescence in situ hybridization (FISH) to determine whether their co-location on human 15q is conserved in marsupials. *GABRB3* and *HERC2* (lambda clones), as well as the BAC containing *UBE3A* and *CNGA3* co-localized in the middle of tammar Chromosome 5p (Figure 2). This was unexpected because tammar 5p has homology to the short arm of the human X chromosome (reviewed [30]). This location was also different to the localization on tammar 1q of a lambda clone originally thought to contain *SNRPN* [31], but now identified as *SNRPB′*. We mapped the tammar BAC containing both *SNRPN* and *SNRPB′* unequivocally to the middle of tammar Chromosome 1q (Figure 2).

Thus, the arrangement of marsupial and monotreme orthologs of PWS-AS genes differs from that in humans and

Table 1. Positions of Orthologs of Human PWS-AS Genes and Their Flanking Markers in Chickens and Three Groups of Mammals

Gene	Chicken	Platypus	Opossum (Scaffold)	Tammar	Dog	Mouse	Human
UTX	1			5p	X	X	Xp
DMD	1			5p	X	X	Xp
NR0B1	1			5p	X	X	Xp
ZFX	1			5p	X	X	Xp
EIF2S3	1			5p	X	X	Xp
PDHA1	1			5p	X	X	Xp
STS	1			5p	X	X	Xp
RGN	1		12679		X	X	Xp
PHF16	1		12679		X	X	Xp
RP2	1		12679		X	X	Xp
SLC9A7	1		12679		X	X	Xp
TUBGCP5	1		12679		3	7	15q
CYFIP1	1		12679		3	7	15q
NIPA2	1		12679		3	7	15q
NIPA1	1		12679		3	7	15q
HERC2	1		12679	5p	3	7	15q
OCA2	1		12679		3	7	15q
GABRG3	1		12679		3	7	15q
GABRA5	1		16658		3	7	15q
GABRB3	1		16658	5p	3	7	15q
ATP10A	1		16658		3	7	15q
UBE3A	1	Oa_Bb358D20 (BAC)	16658	5p	3	7	15q
CNGA3	1	Oa_Bb358D20 (BAC)	16658	5p	10	1	2q
INPP4A	1		16658		10	1	2q
UNC50	1		16658		10	1	2q
MGAT4A	1		16658		10	1	2q
SNURF	—		—		3	7	15q
SNRPN	—		15105	1q	3	7	15q
MAGEL2	—		—		3	7	15q
NDN	—		—		3	7	15q
MKRN3	—		—		3	7	15q
KLF13	10		13378		3	7	15q
APBA2	10		13378		3	7	15q
TJP1	10		13378		3	7	15q
CHRNA7	10		13602		3	7	15q
SNRPB	20	31468, 247728 (contigs)	15105	1q	24	2	20p

Genes are color-coded as for Figures 1 and 4. Chicken Chromosome 1 genes are listed in order of their position.
DOI: 10.1371/journal.pgen.0020182.t001

mice. *UBE3A* and flanking genes do not co-localize with *SNRPN*, but instead share synteny with genes on human Chromosomes Xp and 2. To determine which arrangement is ancestral, we extended our comparison to other vertebrates, making use of information in public databases.

Other Vertebrate Genomes Reveal the Ancestral Arrangement of PWS-AS Genes

We searched the UCSC database (http://www.genome.ucsc.edu) for dog, chicken, and fish homologs of genes in and near the human PWS-AS region, including *UBE3A*, *GABRB3*, and *HERC2* (tammar PWS-AS orthologs), as well as *CNGA3* and genes adjacent to this region in tammar (lying on Xp in human). Orthologs of these genes all lay on scaffolds mapping to chicken Chromosome 1 (Table 1). A block of nine human 15q11-q13 genes (from *UBE3A* to the centromeric *TUBGCP5*) was found sandwiched between large blocks of human Xp and Chromosome 2 genes on chicken Chromosome 1. *CNGA3*, flanking a block of human Chromosome 2 genes, lies only 2 kb from *UBE3A*. In zebrafish, although the synteny groups are duplicated and somewhat broken up, *UBE3A* and *CNGA3* lie on Chromosome 6 along with several other genes from the human X, 15, and 2 blocks and are separate from genes that flank this region in human, which lie on Chromosomes 7 and 20.

Attempts to retrieve *SNURF-SNRPN* and *MKRN3* sequence from the chicken and zebrafish databases yielded only the ancestral *SNRPB′* and *MKRN1*, suggesting that non-mammal vertebrates lack *SNURF-SNRPN* and *MKRN3*. No sequences orthologous to *MAGEL2* and *NDN* were found in chicken and fish genomes (Table S2).

The finding that human 15q11-q13 genes lie between blocks of human Xp and Chromosome 2 genes in chicken and fish, as well as marsupials and monotremes, and that *MKRN3*, *MAGEL2*, *NDN*, and *SNURF* are absent, implies that this is an ancestral vertebrate arrangement. Its occurrence also in marsupials and monotremes implies that the ancestral arrangement was retained in marsupials and monotremes, but the region was rearranged and augmented more recently in placental mammals.

Are *SNRPN* and *UBE3A* Imprinted in Marsupials and Monotremes?

Of the greatest interest was to determine whether marsupial and monotreme *SNRPN* and *UBE3A* are imprinted

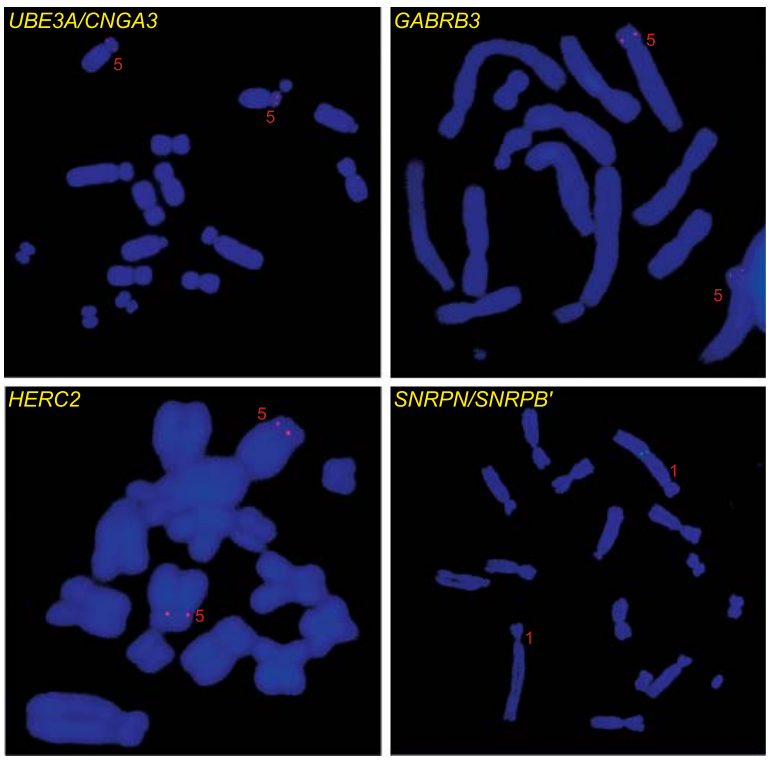

Figure 2. FISH Localization of BACs Containing Tammar Homologs of PWS-AS Genes
UBE3A/CNGA3, GABRB3, and *HERC2* localize to tammar wallaby Chromosome 5p and *SNRPN/SNRPB* to Chromosome 1q.
DOI: 10.1371/journal.pgen.0020182.g002

like their human orthologs, even though they do not share the gene arrangement of the placental PWS-AS region.

Expressed polymorphisms in these two genes in marsupials were therefore sought by screening DNA from 60 tammar wallabies. Sequence analysis revealed a *SNRPN* polymorphism in one heterozygote and a *UBE3A* polymorphism in two. Brain RNA (the tissue showing imprinting of both *UBE3A* and *SNRPN* in placental mammals) was extracted from these animals and a fragment amplified from the resulting cDNA template using gene-specific primers that spanned introns. Sequencing the PCR products revealed expression at similar levels of both parental alleles for both genes (Figure 3).

We also investigated the expression of monotreme *UBE3A* (but not *SNRPN,* since it appears to be absent from the monotreme genome). DNA from five platypuses was screened, and a *UBE3A* polymorphism detected in one heterozygous animal, from which brain tissue was isolated and RNA extracted. Amplification and sequencing from brain cDNA revealed expression of both alleles (Figure 3). Because the alleles differed by a base pair insertion, this result was confirmed by primer extension assays (Figure S2).

Expression studies therefore show that *UBE3A* (the Angelman syndrome gene) is biallelically expressed in marsupial and monotreme brain. The Prader-Willi gene *SNRPN,* represented by a tandem duplication of the non-imprinted gene *SNRPB′* in marsupials, also does not appear to be imprinted.

Discussion

Our results demonstrate that the PWS-AS imprinted region of placental mammals was assembled from a variety of disparate genomic elements and suggest that imprinting was acquired relatively recently in the placental lineage. These conclusions follow from differences in arrangement and expression of *SNRPN* and *UBE3A* in the three major mammal groups, placentals, marsupials, and monotremes.

SNRPN and *UBE3A* lie together in all placental mammals investigated, including the fully sequenced human, mouse, and dog. In the basal placental clades, Afrotheria and Xenarthra, sequence assembly is not sufficiently advanced to ascertain gene arrangement, but chromosome painting experiments reveal that human Chromosome 15q is represented as a single block in elephant and armadillo [32,33]. In contrast, we demonstrate that the two primary PWS-AS loci *SNRPN* and *UBE3A* lie on different chromosomes in marsupials and monotremes, and that other PWS-AS loci have no orthologs in marsupials or monotremes. Since this gene arrangement is shared by birds and fish, it must be ancestral.

This implies that a major rearrangement occurred in the placental lineage to unite *UBE3A* and *SNRPN.* The evolutionary breakpoints lie between *SLC9A7* and *TUBGCP5* and between *UBE3A* and *CNGA3* in the ancestral sequence. Fission between *UBE3A* and *CNGA3* was evidently preceded by

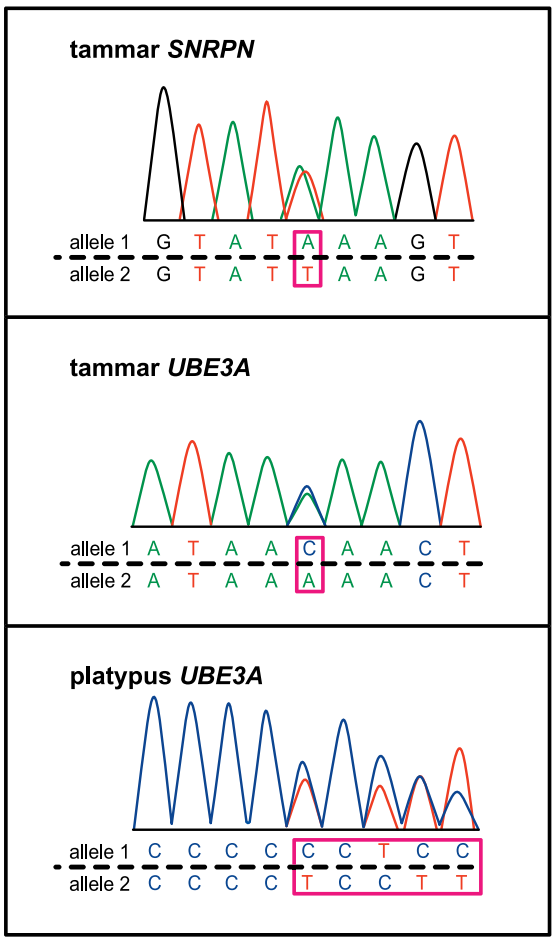

Figure 3. Biallelic Expression Demonstrated by Sequencing Brain cDNA from Heterozygous Animals for Alleles of *SNRPN* and *UBE3A* in Tammar Wallaby and *UBE3A* in Platypus

Alleles differ at an A/T polymorphism at base pair 67 of the 3′ UTR of tammar *SNRPN,* a C/T polymorphism at base pair 247 of exon 5 in tammar *UBE3A,* and an insertion polymorphism of a C at base pair 179 of the 3′ UTR in platypus *UBE3A* (marked by boxes).

DOI: 10.1371/journal.pgen.0020182.g003

expansion of the interval from ~7.0 kb (chicken, fish, and monotremes) to >60 kb (marsupials) (Figure 1B). The *UBE3A* region then fused with an ancestral region represented by chicken Chromosome 10 that bears homologs of genes in human 15q11-q13 and flanking genes located distal to *HERC2* (Figure 4).

The presence of *MKRN3, MAGEL2,* and *NDN* in the genomes of dogs, mice, and humans, but their absence from chicken, fish, and the recently sequenced opossum and platypus genomes (each of which are 6× coverage or greater), implies that these genes were all acquired by the placental PWS-AS domain 180–90 MYA. They arrived independently by retrotransposition from paralogs on different chromosomes. The intronless *MKRN3* arose by retrotransposition from the intron-containing source gene *MKRN1* on human Chromosome 20 [28], and the intronless *MAGEL2* and *NDN* arose by retrotransposition from genes on different sites on the X. Since no snoRNA arrays lie near the ancestral *UBE3A* or *SNRPN,* these must also have been seeded from other locations. Thus, the expanded and fused region was evidently a target for insertions and rearrangements. It seems to have remained unstable, since the mouse and human regions differ by several retrotranspositions [27], and the region contains breakpoint hotspots that are the source of many PWS and AS deletions [34].

We conclude from our expression studies that *UBE3A* and *SNRPN* are biallelically expressed in marsupial and monotreme brain. Because imprinted expression of *UBE3A* is found within neurons, but not glial cells of the brain [35], it is possible that we could not detect some level of allelic attenuation for *UBE3A* in marsupials and monotremes. However, considering that the brain of AS patients produces only about 10% *UBE3A* expression [22], we consider that even semi-quantitative methods such as direct sequencing should have detected this level of imprinting in marsupial and monotreme brain. Our finding that *SNRPN* and *UBE3A* are not imprinted in their original locations is consistent with the hypothesis that rearrangement was required for the establishment of imprinting.

Since *UBE3A* and *SNRPN* appear not to be imprinted in marsupials and monotremes, we would not expect to find sequences that control imprinting of these genes, so the absence of *SNURF* and the ICR from the marsupial genome is particularly telling. *SNURF* lies in the ICR (as defined by the region of shortest deletion in human PWS patients) and the *UBE3A* anti-sense transcript originates either at its 5′ end or at alternative upstream exons [36]. We propose that coordinate regulation of the PWS-AS domain required introduction of these sequences into the unstable fused region.

The time at which chromosome rearrangement occurred and imprinting was acquired can be deduced from the mammalian phylogeny (Figure 4). Gene arrangement in the PWS-AS domain is conserved at least between dog, mouse, and human, and the region was syntenic even in the most distantly related placentals, implying that it predated the placental radiation ~105 MYA. The ancestral arrangement in chicken and fish is shared by marsupials and monotremes, which diverged from placentals 180 and 210 MYA, respectively. Major rearrangements therefore occurred after placentals and marsupials diverged 180 MYA, but before the placental radiation 105 MYA. Our demonstration that *SNRPN* and *UBE3A* are also biallelically expressed in marsupials shows that imprinting evolved considerably later (180–80 MYA) in this region than, for instance, in the *IGF2* region. This is consistent with the view that imprinting evolved after the evolution of viviparity in therian mammals, but shows, also, that at least some loci evolved imprinting considerably after this time. The recent findings that the imprinted gene *Nnat* is eutherian-specific [37], and *DLK1* is not imprinted in the opossum [38], also suggest that novel genes could be imprinted more recently. Thus, imprinting has been acquired at different times in different domains, after the evolution of viviparity, and suggests that viviparity is a necessary but not a sufficient condition for the evolution of imprinting.

Is there any significance in our finding that the PWS-AS

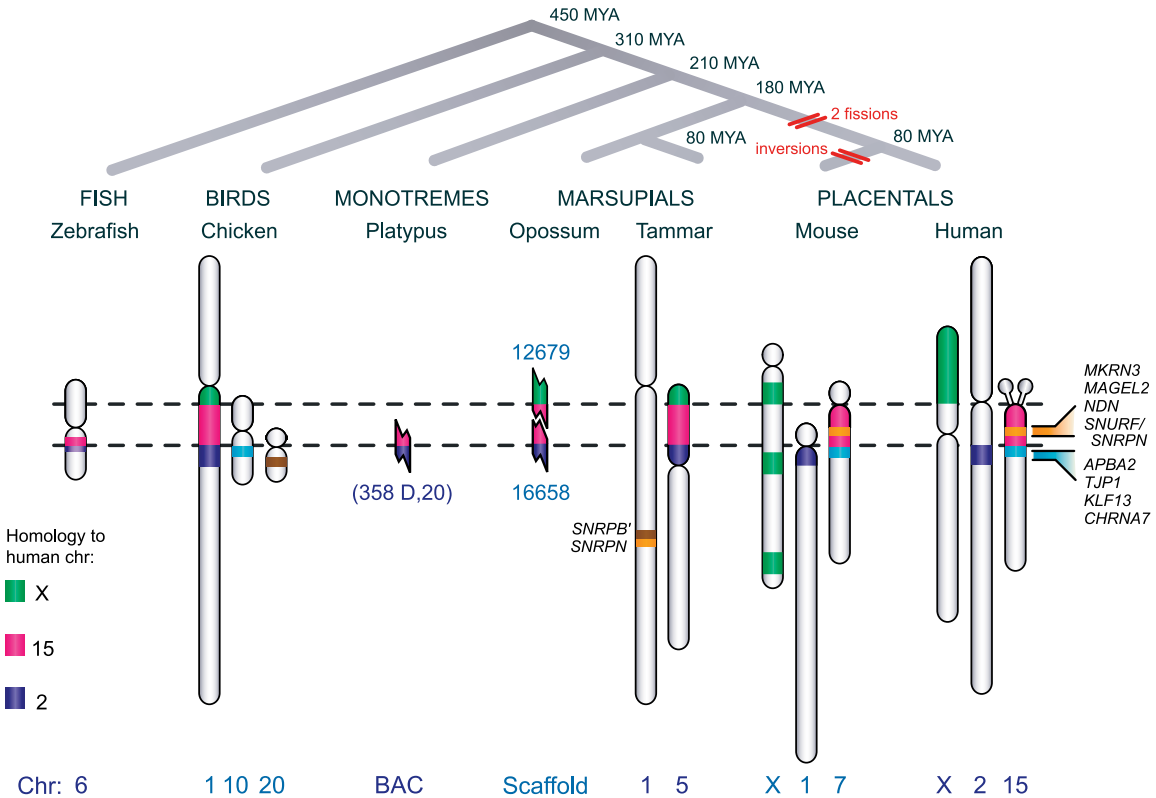

Figure 4. Assembly of the PWS-AS Imprinted Region in Placental Mammals during Vertebrate Evolution

Relationships between fish, birds/reptiles, and the three mammal groups are presented as a phylogeny (top). In the ancestral arrangement, shared by marsupials, monotremes, birds, and fish, the block of imprinted human 15q genes (pink) is flanked by human X (green) and human 2 (purple) blocks. These three blocks were separated by at least two fissions and were rearranged next to an unlinked block of genes (on chick Chromosome 10, pale blue) to make up the present regions of human Chromosome 15q (and with two more inversions, of mouse Chromosome 7). *SNRPB′* (brown) is present on a different chromosome in fish, birds, and marsupials, but its duplicate *SNRPN* (orange) is transposed next to *UBE3A* in placentals. Other human 15q genes absent from non-placental vertebrates (orange) were independently added to the imprinted region in the placental lineage.

DOI: 10.1371/journal.pgen.0020182.g004

imprinted domain once shared synteny with the genome region that became part of the placental X and was recruited into the X inactivation system in placental mammals? There has been continuing speculation that genomic imprinting and X chromosome inactivation are related by descent [2,3]. One possibility is that the entire region was added to the ancestral therian X and recruited to the X inactivation system, and then the PWS-AS region was subsequently relocated to an autosome, carrying elements that control *cis*-silencing. This could also explain the acquisition by this human Chromosome 15 region of *MAGE* (cancer-testis antigen) genes, nearly all of which accumulated on the X chromosome in placentals [39].

In conclusion, we show that the Prader-Willi/Angelman imprinted domain on human Chromosome 15q11–13 was assembled relatively recently from unlinked and non-imprinted components in a mammalian ancestor. Two non-imprinted regions fused 105–180 MYA, and several retroposed genes and snoRNAs from different regions were independently inserted. We propose that genomic rearrangement early in the eutherian lineage was required for the acquisition of imprinting at this locus.

Materials and Methods

Tissue samples. Tammar wallaby tissue samples were sourced at the Research School of Biological Sciences, Australian National University, Canberra, and the Department of Zoology, University of Melbourne, according to the Animal Experimentation Ethics Committee. Platypus samples were supplied courtesy of Dr. F. Grützner, Australian National University, Canberra, Australia.

Nucleic acid extraction, amplification, and sequencing. Total genomic DNA was extracted from tail clippings of 20 pouch young (py) tammar wallabies (collected June 2005), from the brain of three py tammars (collected November 2003), and from ear punches of 30 adult tammars (collected March 2005) using the DNeasy Tissue Kit (Qiagen, Valencia, California, United States) or the protocol outlined in [40]. Total genomic DNA was extracted from two male and one female platypus brain samples (collected September 2004). High-copy number plasmids were extracted using either the Wizard Plus SV Minipreps DNA Purification System (Promega, Madison, Wisconsin, United States) according to the manufacturer's instructions or by the protocol outlined by Sambrook [40]. BAC DNA was extracted using the Wizard Plus SV Minipreps DNA Purification System (Promega).

Total RNA was extracted from the brains of polymorphic platypus and tammar wallabies using the RNeasy Mini Kit (Qiagen) according

Table 2. Description of Probes Used and DNA Libraries Screened

Probe Description	Probe Source	DNA Library	Library Source
Full-length cDNA of human *SNRPN*	R. D. Nicholls	λ ZAP II phage *M. eugenii* cDNA library	Clontech Laboratories Incorporated, Palo Alto, California, United States
3.0-kb fragment from human *GABRB3* cDNA	R. D. Nicholls	λ EMBL3 phage library (average insert size:15–20 kb) containing *M. eugenii* genomic sequence	Steven Wilcox [45]
1.1-kb product from human *HERC2* cDNA	R. D. Nicholls		
213-bp product from exon 7 of *CNGA3* in platypus	Amplified using primers: CNGA3F1 (GAGACCAGGACCACTACCC) CNGA3R1 (GGTCAGGGTGGACCAGTAGA)	Me_Kba BAC library containing *M. eugenii* genomic sequence in 11-fold coverage	Arizona Genomics Institute, Tuscon, Arizona, United States
1,712-bp product encompassing exon 10 and surrounding intronic sequence of *UBE3A* in tammar	Amplified using primers: UBEint9F (GCTGATGGAGTATCCCTCTGTG)UBEint10R (AGGTTTTGCTGAGCCAGAAAG)		
112-bp product encompassing last exon and 3′ UTR of *SNRPN* in tammar	Amplified using primers: SnrpnEx10UTRF (GCGCCCACCAAGACCTTA)SnrpnEx10UTRR (AAATCAATTTAAAAGGAGCACACT)		
280-bp cDNA probe from exons 10–13 of *UBE3A* in tammar	Amplified using primers: UBE3A1F (TACTTATTCAGACCAGAAGA)UBE3A1R (GCAAGTATGAGATGTAGGTAAC)	BAC library containing *M. eugenii* genomic sequence in 2.2-fold coverage	Victorian Institute of Animal Science, Victoria, Australia
268-bp product from exon 14 of *UBE3A* in platypus	Amplified using primers: PlatyUbe3aF (TTTCGGAGAGGTTTTCATATGGTG)PlatyUbe3aR (TTTCGGAGAGGTTTTCATATGGTG)	Oa_Bb BAC library containing *Ornithorhynchs anatinus* genomic sequence in 11-fold coverage	Clemson University Genomics Institute, Clemson, South Carolina, United States
Overgo probe from exon 2 of *SNRPB* in platypus	Oligonucleotides used: SNRPBI-Ova (GCATCTTCATCGGGACCTTCAAGG)SNRPBI-Ovb (ATGTGCTTGTCGAAGGCCTTGAAG)		

DOI: 10.1371/journal.pgen.0020182.t002

to the manufacturer's instructions or by RNaWIZ RNA Isolation reagent. Total RNA was reverse-transcribed using the SuperScript III First-Strand Synthesis System for RT-PCR (Invitrogen, Carlsbad, California, United States) according to manufacturer's instructions.

PCR amplification was conducted on a MJ Research PTC-200 thermal cycler using a 20-μl reaction, including up to 200 ng of template DNA, 1 × PCR reaction buffer (Roche Applied Sciences, Basel, Switzerland), 200 μM dNTPs (Roche Applied Sciences), 0.8 μM of each forward and reverse primer (Table 3), and 1 unit of *Taq* (Roche Applied Sciences). The cycling conditions were 94 °C, 2 min; 35 × (94 °C, 30 s; 54–60 °C, 1 min; 72 °C, 2 min); 72 °C, 10 min.

DNA sequencing was carried out by the Washington University Genome Sequencing Center, St. Louis Missouri, United States or the Australian Genome Research Facility, Brisbane, Australia.

DNA library screening. 50 ng of purified DNA probe was denatured in boiling water for 3 min and radioactively labeled with 50 μCi of [α-^{32}P] dCTP using the Megaprime labeling kit (Amersham, Little Chalfont, United Kingdom) according to the manufacturer's instructions. To remove unincorporated nucleotides the labeled probe was run through a ProbeQuant G-50 Micro column (Amersham Pharmacia Biotech) according to the manufacturer's instructions or through a Sephadex G-50 (Pharmacia) column by centrifugation at 1,000 *g* for 1 min. Library screening using radioactive overgos was also undertaken on the platypus library as described previously [41].

Table 3. PCR Products Used in Expression Studies

Product Description	Primers Used
247-bp product from exons 9–10 and 3′ UTR of *SNRPN* in tammar wallaby	SNRPN ex9 F (CCTCCTCCTGGAATGAGACC)SnrpnEx10UTRR (AAATCAATTTAAAAGGAGCACACT)
517-bp product from exon 5 and exon 6 of *UBE3A* in tammar wallaby	UBE3A ex6 F (GCAAAGCATCTAATAGAACGC)UBE3A ex6 R (GCAGCTTTTTCCTTTTCATCTT)
508-bp product from exon 14 and 3′ UTR of *UBE3A* in platypus	UBEUTR_F4 (TTACCCACATCTCACACTTGCTTT)UBEUTR_R8 (ACACGGTCTACAAACGATGG)

DOI: 10.1371/journal.pgen.0020182.t003

Probes were hybridized to various DNA libraries (Table 2) and exposed to autoradiography film at −70 °C until sufficiently exposed.

FISH. Mitotic metaphase chromosomes were prepared from tammar wallaby fibroblast cell lines grown from ear punctures of adult male wallabies. FISH was carried out as previously described [31]. Probes were fluorescence-labeled by the BioNick Labeling System (Invitrogen) in accordance with the manufacturer's instructions. Probes were detected using fluorescein isothiocyanate (FITC)-conjugated avidin and biotin-conjugated anti-avidin antibody (Vector Laboratories, Burlingame, California, United States). Chromosomes and cell nuclei were counterstained with 1 μg/ml 4,6-diamidino-2-phenylindole (DAPI) in 2 X SSC for 1 min and mounted.

Fluorescence signals were visualized and captured using a Zeiss (Oberkochen, Germany) Axioplan epifluorescence microscope equipped with a CCD (charge-coupled device) camera (RT-Spot, Diagnostic Instruments, Sterling Heights, Michigan, United States). Images were manipulated using IPlab imaging software on an Apple Macintosh computer (Cupertino, California, United States). Gray scale images were captured with source images superimposed into color images.

Expression analysis. Tammar py heterozygous for an expressed polymorphism ranged in age from 104 d to 116 d. PCR products used in expression studies were amplified using primers shown in Table 3. PCR products were purified prior to sequencing using the QIAquick gel extraction kit (Qiagen). Single nucleotide primer extension (SNUPE) assays were undertaken on adult platypus samples using the MassARRAY analyzer system (Sequenom, San Diego, California, United States) by the staff at the Australian Genome Research Facility.

Bioinformatic analysis. Human protein sequence from genes of interest (Table S1) were extracted and interrogated against the mouse (assembly: March 2005), dog (assembly: May 2005), opossum (assembly: October 2004), chicken (assembly: February 2004), and zebrafish (June 2004) databases deposited on the UCSC genome browser (http://www.genome.ucsc.edu) using the BLAT algorithm with default settings [42] (Table S2). In an attempt to identify highly diverged homologs of proteins *SNURF*, *SNRPN*, *MAGEL2*, and *NDN* the more sensitive tBLASTn algorithm [43] was used on databases accessible through the Ensembl website (http://www.ensembl.org). No significant alignments in addition to those found using BLAT were produced. Sequence from the draft platypus genome assembly (version 5.0), related to *SNRPB*, was extracted from the Washington University Genome Sequencing Center (http://genome.wustl.edu/tools/blast) using tBLASTn.

Members of the Sm family of proteins (Table S3) were aligned using the ClustalW program (http://www.ebi.ac.uk/clustalw) with

default parameters. Phylogenetic analysis was performed with the maximum parsimony method using PAUP* version 4.0 b 10 [44]. Gaps were treated as missing data. Most-parsimonious trees were searched using a heuristic strategy, starting trees were obtained via stepwise addition for tree-bisection-reconnection (TBR) branch swapping, one tree was held at each step during stepwise addition, and a maximum of 1,000 best trees were saved in each replicate. 1,000 replications were performed for the bootstrap analysis. A pictorial consensus tree was created using the Phylodendron tree printer (http://iubio.bio.indiana.edu/treeapp/treeprint-form.html).

Supporting Information

Figure S1. Relationships between Vertebrate Sm Proteins Encoded by the Genes *SNRPN* (SmN) and *SNRPB/B′* (SmB) Generated Using Maximum Parsimony Analysis

Marsupial (opossum and tammar) SmN sequences are sister to placental SmN and distant from vertebrate SmB. Support for tree topology is indicated by bootstrap values (1,000 replicates).

Found at DOI: 10.1371/journal.pgen.0020182.sg001 (200 KB PDF).

Figure S2. Primer Extension Assay

Relative allelic concentrations of PCR products amplified from the *UBE3A* gene of platypus brain genomic DNA and cDNA for the three individuals sampled. Standard deviation indicated by error bars.

Found at DOI: 10.1371/journal.pgen.0020182.sg002 (200 KB PDF).

Table S1. Human Protein Sequences Used in Study

Found at DOI: 10.1371/journal.pgen.0020182.st001 (37 KB DOC).

Table S2. Location of Putative Orthologs Aligned to Human Proteins of Interest in Mouse, Dog, Opossum, Tammar Wallaby, Platypus, Chicken, and Zebrafish

Percent identity between query and target sequence is given, along with the coverage of this alignment relative to query sequence length. An asterisk (*) denotes that no significant alignment was found. % Id, percent identity; % cov, percent coverage.

Found at DOI: 10.1371/journal.pgen.0020182.st002 (207 KB DOC).

Table S3. Sm Protein Family Members Used in Phylogenetic Studies and Their Accession Numbers

Found at DOI: 10.1371/journal.pgen.0020182.st003 (21 KB DOC).

Text S1. Expression Studies

A primer extension assay was used to confirm the biallelic expression of *UBE3A* in platypus brain (Figure S2), because the polymorphism identified was a deletion and therefore difficult to test using direct sequencing. DNA was amplified in triplicate from genomic and cDNA samples extracted from the brain of three platypus individuals. One individual was homozygous for the C allele (I), another homozygous for the deleted allele (II), and a third heterozygous for both alleles (III). Each sample was subjected to a primer extension assay capable of quantifying the relative amount of each allele. In accordance with sequencing data, the heterozygous individual showed biallelic expression of the polymorphic site (Figure 1). Interestingly, there appeared to be an unexpected increase in the concentration of the deleted allele for all samples. This is most likely to be an experimental artifact, due to the polymorphism being located within a poly-C tract and resulting in primer slippage.

Found at DOI: 10.1371/journal.pgen.0020182.sd001 (20 KB DOC).

Acknowledgments

We thank H. M. Gehring for assistance with wallaby RNA extractions.

Author contributions. RWR, TH, MS, RDN, and JAMG conceived and designed the experiments. RWR, TH, MS, and RDN performed the experiments. RWR, TH, MK, HH, and JED analyzed the data. EA, AJP, and MBR contributed reagents/materials/analysis tools. RWR, TH, and JAMG wrote the paper.

Funding. This investigation was supported by the Australian Research Council grant to the ARC Center for Kangaroo Genomics.

Competing interests. The authors have declared that no competing interests exist.

References

1. Morison IM, Ramsay JP, Spencer HG (2005) A census of mammalian imprinting. Trends Genet 21: 457–465.
2. Ohlsson R, Paldi A, Graves JAM (2001) Did genomic imprinting and X chromosome inactivation arise from stochastic expression? Trends Genet 17: 136–141.
3. Reik W, Lewis A (2005) Co-evolution of X-chromosome inactivation and imprinting in mammals. Nat Rev Genet 6: 403–410.
4. Haig D, Westoby M (1989) Parent-specific gene-expression and the triploid endosperm. Am Nat 134: 147–155.
5. Haig D (2004) Genomic imprinting and kinship: How good is the evidence? Annu Rev Genet 38: 553–585.
6. O'Neill MJ, Ingram RS, Vrana PB, Tilghman SM (2000) Allelic expression of IGF2 in marsupials and birds. Dev Genes Evol 210: 18–20.
7. Yokomine T, Kuroiwa A, Tanaka K, Tsudzuki M, Matsuda Y, et al. (2001) Sequence polymorphisms, allelic expression status, and chromosome locations of the chicken IGF2 and MPR1 genes. Cytogenet Cell Genet 93: 109–113.
8. Dunzinger U, Nanda I, Schmid M, Haaf T, Zechner U (2005) Chicken orthologs of mammalian imprinted genes are clustered on macrochromosomes and replicate asynchronously. Trends Genet 21: 488–492.
9. Paulsen M, Khare T, Burgard C, Tierling S, Walter J (2005) Evolution of the Beckwith-Wiedemann syndrome region in vertebrates. Genome Res 15: 146–153.
10. Yokomine T, Shirohzu H, Purbowasito W, Toyoda A, Iwama H, et al. (2005) Structural and functional analysis of a 0.5-Mb chicken region orthologous to the imprinted mammalian Ascl2/Mash2-Igf2-H19 region. Genome Res 15: 154–165.
11. Woodburne MO, Rich TH, Springer MS (2003) The evolution of tribospheny and the antiquity of mammalian clades. Mol Phylogenet Evol 28: 360–385.
12. Wakefield MJ, Graves JAM (2003) The kangaroo genome. Leaps and bounds in comparative genomics. EMBO Rep 4: 143–147.
13. Suzuki S, Renfree MB, Pask AJ, Shaw G, Kobayashi S, et al. (2005) Genomic imprinting of IGF2, p57(KIP2), and PEG1/MEST in a marsupial, the tammar wallaby. Mech Dev 122: 213–222.
14. Killian JK, Byrd JC, Jirtle JV, Munday BL, Stoskopf MK, et al. (2000) M6P/IGF2R imprinting evolution in mammals. Mol Cell 5: 707–716.
15. Killian JK, Nolan CM, Stewart N, Munday BL, Andersen NA, et al. (2001) Monotreme IGF2 expression and ancestral origin of genomic imprinting. J Exp Zool 291: 205–212.
16. Weidman JR, Murphy SK, Nolan CM, Dietrich FS, Jirtle RL (2004) Phylogenetic footprint analysis of IGF2 in extant mammals. Genome Res 14: 1726–1732.
17. Wilkins JF, Haig D (2003) What good is genomic imprinting: The function of parent-specific gene expression. Nat Rev Genet 4: 359–368.
18. Knoll JH, Nicholls RD, Magenis RE, Graham JM Jr, Lalande M, et al. (1989) Angelman and Prader-Willi syndromes share a common Chromosome 15 deletion but differ in parental origin of the deletion. Am J Med Genet 32: 285–290.
19. Vu TH, Hoffman AR (1997) Imprinting of the Angelman syndrome gene, *UBE3A,* is restricted to brain. Nat Genet 17: 12–13.
20. Meguro M, Kashiwagi A, Mitsuya K, Nakao M, Kondo I, et al. (2001) A novel maternally expressed gene, *ATP10C,* encodes a putative aminophospholipid translocase associated with Angelman syndrome. Nat Genet 28: 19–20.
21. Herzing LB, Kim SJ, Cook EH Jr, Ledbetter DH (2001) The human aminophospholipid-transporting ATPase gene *ATP10C* maps adjacent to *UBE3A* and exhibits similar imprinted expression. Am J Hum Genet 68: 1501–1505.
22. Rougeulle C, Glatt H, Lalande M (1997) The Angelman syndrome candidate gene, *UBE3A/E6-AP,* is imprinted in brain. Nat Genet 17: 14–15.
23. Runte M, Huttenhofer A, Gross S, Kiefmann M, Horsthemke B, et al. (2001) The IC-SNURF-SNRPN transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for *UBE3A.* Hum Mol Genet 10: 2687–2700.
24. Ohta T, Buiting K, Kokkonen H, McCandless S, Heeger S, et al. (1999) Molecular mechanism of Angelman syndrome in two large families involves an imprinting mutation. Am J Hum Genet 64: 385–396.
25. Buiting K, Lich C, Cottrell S, Barnicoat A, Horsthemke B (1999) A 5-kb imprinting center deletion in a family with Angelman syndrome reduces the shortest region of deletion overlap to 880 bp. Hum Genet 105: 665–666.
26. Runte M, Kroisel PM, Gillessen-Kaesbach G, Varon R, Horn D, et al. (2004) SNURF-SNRPN and UBE3A transcript levels in patients with Angelman syndrome. Hum Genet 114: 553–561.
27. Chai JH, Locke DP, Ohta T, Greally JM, Nicholls RD (2001) Retrotransposed genes such as *Frat3* in the mouse Chromosome 7C Prader-Willi syndrome region acquire the imprinted status of their insertion site. Mamm Genome 12: 813–821.
28. Gray TA, Hernandez L, Carey AH, Schaldach MA, Smithwick MJ, et al. (2000) The ancient source of a distinct gene family encoding proteins featuring RING and C(3)H zinc-finger motifs with abundant expression in developing brain and nervous system. Genomics 66: 76–86.
29. Gray TA, Smithwick MJ, Schaldach MA, Martone DL, Graves JAM, et al.

(1999) Concerted regulation and molecular evolution of the duplicated SNRPB′/B and SNRPN loci. Nucleic Acids Res 27: 4577–4584.
30. Alsop AE, Miethke P, Rofe R, Koina E, Sankovic N, et al. (2005) Characterizing the chromosomes of the Australian model marsupial *Macropus eugenii* (tammar wallaby). Chromosome Res 13: 627–636.
31. Toder R, Wienberg J, Voullaire L, O'Brien PC, Maccarone P, et al. (1997) Shared DNA sequences between the X and Y chromosomes in the tammar wallaby—evidence for independent additions to eutherian and marsupial sex chromosomes. Chromosoma 106: 94–98.
32. Yang F, Alkalaeva EZ, Perelman PL, Pardini AT, Harrison WR, et al. (2003) Reciprocal chromosome painting among human, aardvark, and elephant (superorder Afrotheria) reveals the likely eutherian ancestral karyotype. Proc Natl Acad Sci U S A 100: 1062–1066.
33. Dobigny G, Yang F, O'Brien PC, Volobouev V, Kovacs A, et al. (2005) Low rate of genomic repatterning in Xenarthra inferred from chromosome painting data. Chromosome Res 13: 651–663.
34. Nicholls RD, Knepper JL (2001) Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. Annu Rev Genomics Hum Genet 2: 153–175.
35. Yamasaki K, Joh K, Ohta T, Masuzaki H, Ishimaru T, et al. (2003) Neurons but not glial cells show reciprocal imprinting of sense and antisense transcripts of Ube3a. Hum Mol Genet 12: 837–847.
36. Landers M, Bancescu DL, Le Meur E, Rougeulle C, Glatt-Deeley H, et al. (2004) Regulation of the large (approximately 1,000 kb) imprinted murine Ube3a antisense transcript by alternative exons upstream of Snurf/Snrpn. Nucleic Acids Res 32: 3480–3492.
37. Evans HK, Weidman JR, Cowley DO, Jirtle RL (2005) Comparative phylogenetic analysis of blcap/nnat reveals eutherian-specific imprinted gene. Mol Biol Evol 22: 1740–1748.
38. Weidman JR, Maloney KA, Jirtle RL (2006) Comparative phylogenetic analysis reveals multiple non-imprinted isoforms of opossum Dlk1. Mamm Genome 17: 157–167.
39. Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, et al. (2005) The DNA sequence of the human X chromosome. Nature 434: 325–337.
40. Sambrook J, Maniatis T, Fritsch EF (1989) Molecular cloning: A laboratory manual. Cold Spring Harbor (New York): Cold Spring Harbor Laboratory Press. 999 p.
41. McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, et al. (2001) A physical map of the human genome. Nature 409: 934–941.
42. Kent WJ (2002) BLAT—the BLAST-like alignment tool. Genome Res 12: 656–664.
43. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403–410.
44. Swofford DL (2000) PAUP: Phylogenetic analysis using parsimony and other methods, version 4 [computer program]. Sunderland (Massachusetts): Sinauer.
45. Wilcox SA (1995) The evolution of the mammalian sex chromosomes [thesis]. Melbourne (Australia): La Trobe University. pp. 40–43. Available from La Trobe University, Melbourne, Australia; 599.087322 W667e C2.

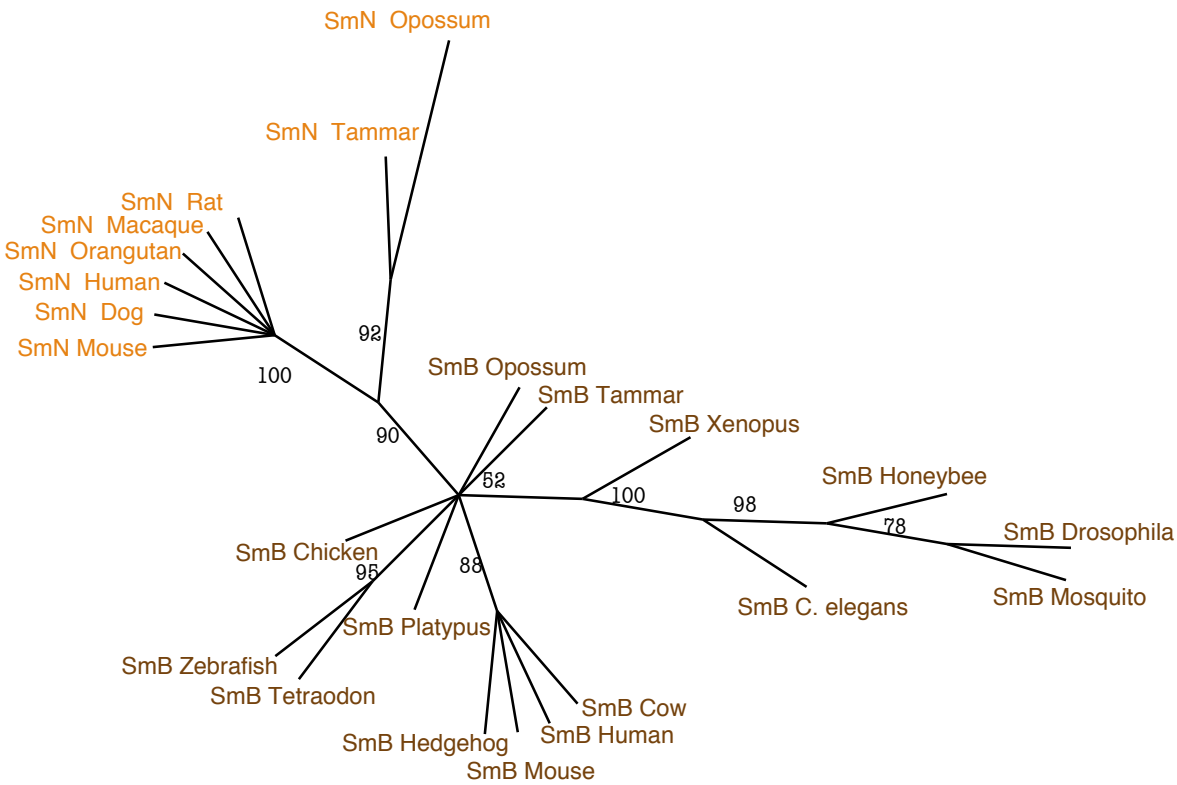

Figure S1. Relationships between Vertebrate Sm Proteins Encoded by the Genes *SNRPN* (SmN) and *SNRPB/B'* (SmB) Generated Using Maximum Parsimony Analysis. Marsupial (opossum and tammar) SmN sequences are sister to placental SmN and distant from vertebrate SmB. Support for tree topology is indicated by bootstrap values (1,000 replicates).

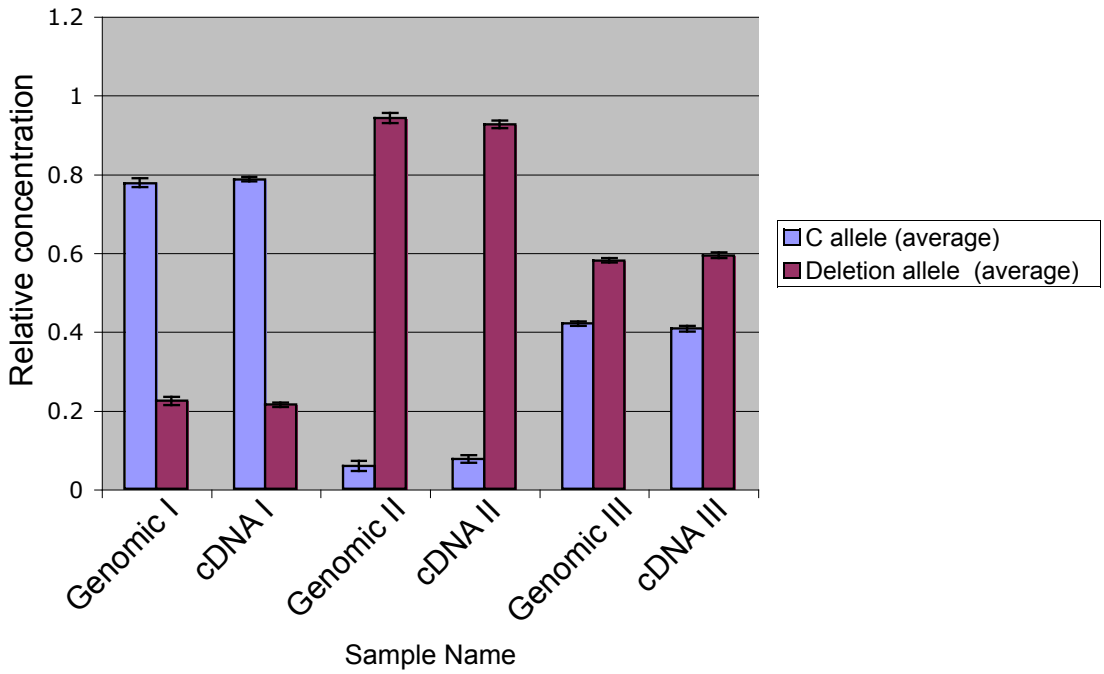

Figure S2. Primer Extension Assay. Relative allelic concentrations of PCR products amplified from the *UBE3A* gene of platypus brain genomic DNA and cDNA for the three individuals sampled. Standard deviation indicated by error bars. DNA was amplified in triplicate from genomic and cDNA samples extracted from the brain of three platypus individuals. One individual was homozygous for the C allele (I), another homozygous for the deleted allele (II), and a third heterozygous for both alleles (III). Each sample was subjected to a primer extension assay capable of quantifying the relative amount of each allele. In accordance with sequencing data, the heterozygous individual showed biallelic expression of the polymorphic site (Figure 1). Interestingly, there appeared to be an unexpected increase in the concentration of the deleted allele for all samples. This is most likely to be an experimental artifact, due to the polymorphism being located within a poly-C tract and resulting in primer slippage.

Table S1. Human protein sequences used in study.

Query	Accession number	Chromosome	Start	End
UTX	O15550	chrX	44489051	44725748
DMD	P11532	chrX	30911873	32797972
NR0B1	P51843	chrX	30082355	30087137
ZFX	P17010	chrX	23950516	23989147
EIF2S3	P41091	chrX	23832742	23854556
PDHA1	P08559	chrX	19121806	19137425
STS	P08842	chrX	7030970	7128035
RGN	Q15493	chrX	46696797	46708597
PHF16	Q92613	chrX	46600549	46674730
RP2	O75695	chrX	46452789	46495455
SLC9A7	Q96T83	chrX	46222643	46374721
TUBGCP5	Q96RT8	chr15	20384959	20424686
CYFIP1	NP_055423	chr15	20477223	20554478
NIPA2	NP_001008892	chr15	20557664	20572777
NIPA1	NP_653200	chr15	20600272	20637852
UBE3A	Q05086	chr15	23135379	23172002
ATP10A	NP_077816.1	chr15	23475583	23659336
GABRB3	P28472	chr15	24344035	24569202
GABRA5	P31644	chr15	24665488	24776123
GABRG3	Q99928	chr15	24799428	25451619
OCA2	Q04671	chr15	25674131	26000615
HERC2	NP_004658	chr15	26030506	26240174
APBA2	Q99767	chr15	27133379	27196625
TJP1	Q07157	chr15	27781093	27880188
KLF13	Q9Y2Y9	chr15	29406707	29451791
CHRNA7	P36544	chr15	30110089	30247948
CNGA3	Q16281	chr2	98444956	98472233
INPP4A	NP_001557	chr2	98595029	98662586
UNC50	NP_054763	chr2	98684740	98693282
MGAT4A	NP_036346	chr2	98700703	98801319
SNRPB	P14678	chr20	2391279	2396404
MKRN3	Q13064	chr15	21362022	21363543
MAGEL2	NP_061939.1	chr15	21440235	21441795
NDN	Q99608	chr15	21482494	21483457
SNURF	NP_073715	chr15	22758351	22764274
SNRPN	NP_003088.1	chr15	22771597	22774681

Table S2. Location of putative orthologs aligned to Human proteins of interest in Mouse, Dog, Opossum, Tammar Wallaby, Platypus, Chicken, and Zebrafish. Percent identity between query and target sequence is given, along with the coverage of this alignment relative to query sequence length. An asterisk (*) denotes that no significant alignment was found. % ID, percent identity; % cov, percent coverage.

Mouse					
Query	**Target Name**	**Target Start**	**Target End**	**% ID**	**% Coverage**
UTX	chrX	16401976	16516487	94	99
DMD	chrX	77797286	79851361	88	99
NR0B1	chrX	80852459	80856248	61	100
ZFX	chrX	88739457	88759512	95	95
EIF2S3	chrX	88851520	88873292	99	100
PDHA1	chrX	153723191	153732638	99	82
STS	*	*	*	*	*
RGN	chrX	18789319	18800423	88	92
PHF16	chrX	18718454	18756834	82	92
RP2	chrX	18615821	18636673	83	84
SLC9A7	chrX	18344906	18530625	92	96
TUBGCP5	chr7	50065373	50100363	88	93
CYFIP1	chr7	50141337	50199637	99	98
NIPA2	chr7	50202444	50214097	91	96
NIPA1	chr7	50248872	50289037	99	91
UBE3A	chr7	53550079	53583181	95	88
ATP10A	chr7	52936756	53106874	80	96
GABRB3	chr7	51863875	52097347	97	92
GABRA5	chr7	51680998	51763460	96	85
GABRG3	chr7	50993893	51659162	91	91
OCA2	chr7	50554068	50698884	76	69
HERC2	chr7	50320354	50500607	95	100
APBA2	chr7	58579102	58637178	91	88
TJP1	chr7	59181614	59238918	88	96
KLF13	chr7	57665288	57712324	86	61
CHRNA7	chr7	56872964	56985986	94	90
CNGA3	chr1	37548682	37556677	90	63
INPP4A	chr1	37652393	37702731	94	98
UNC50	chr1	37725771	37733290	95	92
MGAT4A	chr1	37739281	37830613	91	88
SNRPB	chr2	129686454	129691020	99	95
MKRN3	chr7	56192363	56193527	63	66
MAGEL2	chr7	56153454	56154855	65	91
NDN	chr7	56122145	56123120	77	93
SNURF	chr7	53771930	53776009	97	96
SNRPN	chr7	53759845	53764042	100	95

Dog					
Query	**Target Name**	**Target Start**	**Target End**	**% ID**	**% Coverage**
UTX	chrX	38693898	38905089	98	99
DMD	chrX	26820708	28110606	93	100
NR0B1	chrX	25349389	25354078	75	100
ZFX	chrX	19716698	19743870	97	96
EIF2S3	chrX	19615905	19633540	100	100
PDHA1	chrX	15291876	15305557	99	88
STS	chrX	4349468	4419017	79	90
RGN	chrX	40646435	40660867	91	91
PHF16	chrX	40568419	40627708	89	96
RP2	chrX	40420035	40468966	95	94
SLC9A7	chrX	40193973	40336251	97	92
TUBGCP5	chr3	35126572	35185446	94	97
CYFIP1	chr3	34940370	35013332	96	98
NIPA2	chr3	34914424	34937719	98	100
NIPA1	chr3	34846702	34894912	99	92
UBE3A	chr3	38254978	38290086	98	89
ATP10A	chr3	37688624	37831089	93	95
GABRB3	chr3	36723892	36945197	91	93
GABRA5	chr3	36509909	36589323	96	92
GABRG3	chr3	35809726	36481382	98	89
OCA2	chr3	35208770	35558737	75	86
HERC2	chr3	34567357	34808560	97	100
APBA2	chr3	41224598	41285096	90	89
TJP1	chr3	41853070	41935256	93	97
KLF13	chr3	40494286	40494574	98	21
CHRNA7	chr3	39708369	39833183	97	91
CNGA3	chr10	47357444	47373494	80	78
INPP4A	chr10	47212502	47267853	97	98
UNC50	chr10	47181239	47195374	98	92
MGAT4A	chr10	47058833	47174295	95	89
SNRPB	chr24	21673557	21679358	100	81
MKRN3	chr3	39273496	39274750	71	79
MAGEL2	chr3	39221582	39223139	82	98
NDN	chr3	39172076	39172934	92	84
SNURF	chr3	38706634	38711308	100	94

Opossum					
Query	**Target Name**	**Target Start**	**Target End**	**% ID**	**% Coverage**
UTX	scaffold_13259	897353	1147144	91	99
DMD	scaffold_16798	2039723	4082158	80	99
NR0B1	scaffold_13277	3162209	3165883	68	49
ZFX	scaffold_15001	6829094	6867590	94	95
EIF2S3	scaffold_15001	6661616	6676881	98	100
PDHA1	scaffold_14793	140165	152635	92	78
STS	scaffold_13648	1529424	1631162	71	88

RGN	scaffold_12679	1850467	1868249	79	87
PHF16	scaffold_12679	1893766	1943200	72	92
RP2	scaffold_12679	2134635	2152115	77	77
SLC9A7	scaffold_12679	2284724	2520910	89	90
TUBGCP5	scaffold_12679	2700968	2743015	83	90
CYFIP1	scaffold_12679	2817550	2912628	97	98
NIPA2	scaffold_12679	2917365	2923074	94	99
NIPA1	scaffold_12679	2960227	3019697	92	90
UBE3A	scaffold_16658	2000681	2067334	94	88
ATP10A	scaffold_16658	2589124	2884751	72	85
GABRB3	scaffold_16658	3943166	4204840	96	92
GABRA5	scaffold_16658	4432738	4551693	93	91
GABRG3	scaffold_12679	4693309	5071351	88	73
OCA2	scaffold_12679	3590688	4102983	88	61
HERC2	scaffold_12679	3169194	3373486	91	100
APBA2	scaffold_13378	2770235	2906606	77	85
TJP1	scaffold_13378	1854101	1931757	77	96
KLF13	scaffold_13378	3840428	3939057	57	61
CHRNA7	scaffold_13602	1306903	1511405	92	90
CNGA3	scaffold_16658	1923277	1939958	80	75
INPP4A	scaffold_16658	953281	1042914	88	98
UNC50	scaffold_16658	747305	761216	95	92
MGAT4A	scaffold_16658	619930	736494	80	88
SNRPB	scaffold_15105	181332	190945	98	100
MKRN3	*	*	*	*	*
MAGEL2	*	*	*	*	*
NDN	*	*	*	*	*
SNURF	*	*	*	*	*
SNRPN	scaffold_15105	160379	172878	94	100

Tammar					
Query	**Target Name**	**Target Start**	**Target End**	**% ID**	**% Coverage**
UBE3A	DQ985736	52	516	78	16 (Exons 9, 11)
GABRB3	DQ985733	1	50	92	10 (Exon 4)
HERC2	DQ985734	41	196	80	1 (Exon 20)
CNGA3	DQ985737	1	213	95	7 (Exon 7)
SNRPB	AF176323	1	720	98	100
SNRPN	DQ985732	1	720	95	100

Platypus					
Query	**Target Name**	**Target Start**	**Target End**	**% ID**	**% Coverage**
UBE3A	Oa_Bb_358D20 (AC153157)	24936	62953	89	100
CNGA3	Oa_Bb_358D20 (AC153157)	60943	55769	74	65 (Exons 5,6 and 7)
SNRPB	Contig_31468, (AAPN01400224) Contig_247728 (AAPN01346103)	1558 (31468)	413 (247728)	95	56 (Exons 2, 5 and 6)

Chicken					
Query	**Target Name**	**Target Start**	**Target End**	**% ID**	**% Coverage**
UTX	chr1	104551380	104664179	92	94
DMD	chr1	108251626	109037065	70	98
NR0B1	chr1	109365081	109366342	70	25
ZFX	chr1	111548006	111561852	88	95
EIF2S3	chr1	111583582	111596226	99	100
PDHA1	chr1	113631696	113639085	83	77
STS	chr1	119318135	119402371	54	80
RGN	chr1	123149368	123158318	82	71
PHF16	chr1	123169869	123187104	64	90
RP2	chr1	123236647	123241329	83	77
SLC9A7	chr1	123264670	123331702	83	84
TUBGCP5	chr1	123380151	123399712	81	93
CYFIP1	chr1	123419018	123469801	97	98
NIPA2	chr1	123472602	123476437	90	90
NIPA1	chr1	123491113	123494377	95	72
UBE3A	chr1	124935062	124955276	93	88
ATP10A	chr1	124712054	124820785	64	82
GABRB3	chr1	124459446	124487052	87	77
GABRA5	chr1	124292718	124339118	86	85
GABRG3	chr1	123988120	124283670	89	87
OCA2	chr1	123731609	123832426	75	55
HERC2	chr1	123531854	123635894	91	100
APBA2	chr10	6632517	6671352	72	85
TJP1	chr10	6801626	6842786	64	96
KLF13	chr10	6347546	6372312	51	61
CHRNA7	chr10	7435669	7473396	92	90
CNGA3	chr1	124962289	124963660	87	57
INPP4A	chr1	125219580	125262427	88	98
UNC50	chr1	125338067	125341228	85	92
MGAT4A	chr1	125347063	125416738	86	88
SNRPB	chr20	10267880	10270086	96	95
MKRN3	*	*	*	*	*
MAGEL2	*	*	*	*	*
NDN	*	*	*	*	*
SNURF	*	*	*	*	*
SNRPN	*	*	*	*	*

Zebrafish					
Query	**Target Name**	**Target Start**	**Target End**	**% ID**	**% Coverage**
UTX	chr9	32971650	33010431	56	84
DMD	chr1	4429041	4668091	13	95
NR0B1	*	*	*	*	*
ZFX	chr14	12381148	12382315	75	47
EIF2S3	chr24	11824002	11841685	98	95
PDHA1	chr5	12157945	12165204	79	78
STS	*	*	*	*	*

RGN	chr6	26177135	26180169	70	47
PHF16	chrNA	64322405	64339622	66	60
RP2	chr6	18690916	18691585	76	59
SLC9A7	chr6	18699088	18756493	78	78
TUBGCP5	chr6	18798400	18818005	64	80
CYFIP1	chr14	46365976	46403251	84	98
NIPA2	chr7	25441399	25450625	86	88
NIPA1	chr7	25453218	25456695	46	77
UBE3A	chr6	21085748	21096430	83	88
ATP10A	chr6	21017207	21028259	59	35
GABRB3	chr6	20568300	20659657	73	87
GABRA5	chr12	20349581	20376394	73	81
GABRG3	chr6	20780259	20937025	77	78
OCA2	chr6	21238248	21335124	70	52
HERC2	chr7	25475816	25552077	74	96
APBA2	chr7	23533804	23581194	67	80
TJP1	*	*	*	*	*
KLF13	chr9	26874589	26896812	83	27
CHRNA7	chr7	58379186	58397973	67	81
CNGA3	chr6	21077771	21078989	83	50
INPP4A	chr6	9301605	9358688	73	75
UNC50	*	*	*	*	*
MGAT4A	chr9	3708932	3759271	70	67
SNRPB	chrNA	8113050	8121800	93	57
MKRN3	*	*	*	*	*
MAGEL2	*	*	*	*	*
NDN	*	*	*	*	*
SNURF	*	*	*	*	*
SNRPN	*	*	*	*	*

Table S3. Sm Protein Family Members Used in Phylogenetic Studies and Their Accession Numbers

Species:Protein	**Accession number**
Mosquito:SmB	BX052864
C.elegans:SmB	NM_060947
Chicken:SmB	NM_204599.1
Cow:SmB	NP_001029903
Dog:SmN	XP_536165.1
Drosophila:SmB	NM_057573.3
Hedgehog:SmB	AF134826.1
Honeybee:SmB	XM_392871
Human:SmB	P14678
Human:SmN	NP_003088
Macaque:SmN	AB125194.1
Mouse:SmB	NP_033251.1
Mouse:SmN	NP_038698
Opossum:SmB'	AF134827
Opossum:SmN	DQ899953
Orangutan:SmN	CAH93359
Platypus:SmB	AAPN01400224, AAPN01346103
Rat:SmN	AAH87671
Tammar:SmB	AF176323
Tammar:SmN	DQ985732
Tetraodon:SmB	CR721452.2
Xenopus:SmB	BC041275.1
Zebrafish:SmB	NM_205667.1