Comparative phylogenetic analysis reveals multiple non-imprinted isoforms of opossum *Dlk1*

Jennifer R. Weidman,^{1,2} Kristin A. Maloney,^{1–3} Randy L. Jirtle^{1–3}

¹Department of Radiation Oncology, Duke University Medical Center, Durham, North Carolina 27710, USA

²University Program in Genetics and Genomics, Duke University Medical Center, Durham, North Carolina 27710, USA

³Program in Cell and Molecular Biology, Duke University Medical Center, Durham, North Carolina 27710, USA

Received: 30 August 2005 / Accepted: 18 October 2005

Abstract

Imprinted genes are monoallelically expressed in a parent-of-origin manner and were previously identified in both marsupials and eutherians, but not in monotremes. Phylogenetic comparison of imprinted domains is a powerful tool for investigating the molecular and adaptive evolution of this unique gene regulatory mechanism. Herein, we report that multiple transcripts of *Dlk1* (*Delta*, *Drosophila*, *Homolog-like 1*) are expressed in the opossum, but none are imprinted. Thus, we provide the first example of a reciprocally imprinted gene domain in which imprinting evolved in a common ancestor to eutherian rather than therian mammals. Moreover, the reciprocally imprinted *Meg3* (*Maternally Expressed Gene 3*), found downstream of *Dlk1* in eutherian mammals, is absent in the opossum. We propose that the *Meg3* sequence integrated into the eutherian *Dlk1* domain via a LINE-1 element and that *Dlk1* became imprinted in eutherian mammals only after this downstream integration. These findings clearly demonstrate that imprinted genes did not all evolve before the divergence of marsupials and eutherians.

Introduction

Dlk1 (*Delta*, *Drosophila*, *Homolog-like 1*) encodes for a cell-surface transmembrane protein demonstrated to be involved in the development and dif-

ferentiation of adipose, mesenchymal, neuroendocrine, and hepatopoietic tissues (Laborda 2000). *DLK1* in both mouse and human synthesizes a transmembrane protein containing six EGF-like (epidermal growth factor-like) repeats that are closely related to the EGF-like repeats of the invertebrate proteins delta, serrate, notch, and lin-12 (Smas and Sul 1993; Fleming 1998; Kobayashi et al. 2000). The functional significance of the EGF domains is not fully understood; however, EGF-like repeats are commonly found in the extracellular domain of membrane-bound proteins or in proteins known to be secreted (Garces et al. 1999). The EGF domains of *DLK1* include six cysteine residues that are linked by a total of 18 disulfide bonds. It has been suggested that the six EGF-like repeats conserved between mouse and human *DLK1* mediate its function as a negative regulator of adipocyte differentiation in a manner similar to that by which other EGF-like proteins govern cell fate decisions in invertebrates (Smas and Sul 1993).

Dlk1 is expressed in multiple embryonic tissues, yet its expression is downregulated in most tissues after birth in eutherians (Murphy et al. 2005), with the exception of preadipocytes, pancreatic β cells, thymocytes, and adrenal cells (Carlsson et al. 1997; Hansen et al. 1998; Kaneta et al. 2000). Interestingly, the callipyge mutation in sheep interferes with the normal postnatal downregulation of *Dlk1* expression in fast twitch muscles (Freking et al. 2002; Murphy et al. 2005). *Dlk1* is not only temporally regulated in eutherian mammals, it is also imprinted and preferentially expressed from the paternal allele (Schmidt et al. 2000; Wylie et al. 2000). Imprinting is regulated by epigenetic chromosomal modifications that lead to monoallelic, parent-of-origin-dependent expression (Reik and Walter 2001; Murphy and Jirtle 2003). In eutherian mammals, *Dlk1* is juxtaposed to

URL: <http://www.genem印rint.com>

Nucleotide sequence data reported are available in the GenBank database under the accession number AY972866.

Correspondence to: Randy L. Jirtle, Box 3433, Duke University Medical Center, Durham, NC 27710, USA; E-mail: jirtle@radonc.duke.edu

the reciprocally imprinted gene, *Meg3* (*Maternally Expressed Gene 3*, also named *Gtl2*, but hereafter referred to as *Meg3*) (Schmidt et al. 2000; Wylie et al. 2000). The silencing of *Dlk1* on the maternal allele is proposed to be controlled by an intergenic differentially methylated region (DMR) located within the approximately 80 kb separating *Dlk1* and *Meg3* (Lin et al. 2003).

Imprinted genes have been linked to a number of human behavioral and developmental disorders and to a variety of pediatric and adult malignancies (Falls et al. 1999; Murphy and Jirtle 2003). *DLK1/MEG3* loss of imprinting is linked to tumorigenesis in a subset of human malignancies, including neuroblastomas, pheochromocytomas and Wilms' tumors (Astuti et al. 2005; Li et al. 2005). Loss of paternal *Dlk1* expression in the mouse leads to growth retardation, obesity, blepharophimosis, skeletal malformations, and increased serum lipid metabolites (Moon et al. 2002). In sheep, an intergenic mutation between *Dlk1* and *Meg3* results in the callipyge phenotype, characterized by fast twitch muscle hypertrophy (Freking et al. 2002; Smit et al. 2003). This hypertrophy is recapitulated by murine expression of *Dlk1* in skeletal muscle, demonstrating that the callipyge phenotype is correlated with ectopic expression of *Dlk1* protein (Davis et al. 2004; Murphy et al. 2005). Thus, a greater knowledge of the *cis*-acting elements necessary to establish and maintain imprinting at the *Dlk1/Meg3* locus is needed to understand the pathobiology that results from its dysregulation.

Three general classes of imprint regulation have been proposed (Ferguson-Smith 2000; Reik and Murrell 2000). Postulated to be the primordial imprint mark since it is the simplest in form (Reik and Walter 2001), allele-specific promoter hypermethylation for genes such as *Nnat* (Evans et al. 2001; John et al. 2001) and *U2af1-rs* (Nabetani et al. 1997) leads to repression of expression. Second, some genes such as *Igf2r* are regulated in *cis* by the production of an imprinted antisense RNA (Lyle et al. 2000; Sleutels et al. 2002). The third and most complex system involves the coordinated regulation of clusters of imprinted genes, including juxtapositioned reciprocally imprinted gene (JRIG) domains. For example, the two presently defined examples of JRIG domains, *IGF2/H19* and *DLK1/MEG3*, are organized and imprinted in a similar manner (Schmidt et al. 2000; Takada et al. 2000; Wylie et al. 2000). Previously, our phylogenetic analysis of *IGF2* in extant mammals with divergent imprint status identified strong candidates for mechanistic involvement in the evolution of imprinting at the *IGF2/H19* imprinted domain (Weidman et al. 2004).

One method for identifying *cis*-acting elements involved in imprint establishment and maintenance and for providing a greater understanding of the adaptive and molecular evolution of genomic imprinting is comparative phylogenetic analysis (Onyango et al. 2000; Paulsen et al. 2001; Amarger et al. 2002; Takada et al. 2002; Weidman et al. 2004; Suzuki et al. 2005). Comparative analyses have been previously conducted among eutherian mammals at the *Dlk1/Meg3* locus (Charlier et al. 2001; Paulsen et al. 2001), yet these studies are restricted by evolutionary relatedness (Murphy et al. 2001). A key factor in determining ancestral imprint marks and in effectively identifying putative control elements in humans is through analysis of imprinted regions in more distantly related species (Weidman et al. 2004).

The purpose of this study was to conduct a phylogenetic analysis of the *DLK1/MEG3* imprinted domain in the broader Theria mammalian group to further investigate the adaptive and molecular evolution of imprinting at JRIG domains by including the marsupial *Monodelphis domestica* (gray short-tailed opossum). Genomic imprinting is thought to have evolved with the advent of live birth in therian mammals, possibly because of an interparental genetic conflict to control maternal-dependent growth of the offspring (Haig and Graham 1991). This theory is supported by the finding that *IGF2* (*Insulin-like Growth Factor 2*) and *M6P/IGF2R* (*Mannose 6-phosphate/Insulin-like Growth Factor 2 Receptor*) are imprinted in both marsupials and eutherians (O'Neill et al. 2000; Killian et al. 2001a, 2001b; Suzuki et al. 2005). The phylogenomic comparison reported in this study, however, shows that imprinting at the *DLK1* locus is unique with respect to *IGF2* and *M6P/IGF2R* in that it evolved upon the establishment of a JRIG motif only in placental mammals subsequent to the divergence of marsupials and eutherians. Thus, we demonstrate for the first time that the JRIG domains identified in eutherian mammals did not arise in a single common ancestor.

Materials and methods

Opossum genomic sequence. The *Monodelphis domestica* genome assembly was searched for *Dlk1* sequence based on homology to known *Dlk1* sequences (http://pre.ensembl.org/Monodelphis_domestica/). The *M. domestica* assembly has a base coverage of approximately 7.2 ×. A portion of *oDlk1* exon 5 sequence was annotated and used to generate primer sequences (see Opossum *Dlk1* cDNA sequence identification below).

Opossum tissue samples. Kidney, liver, and brain tissues were taken from young, gray short-tailed opossum offspring. Samples were kindly provided by Dr. Kathleen K. Smith, Duke University, Durham, NC, and were flash frozen and stored at -80°C before nucleic acid extraction.

Opossum *Dlk1* cDNA sequence identification. *oDlk1* cDNA sequence was determined with gene-specific primers using RLM-RACE as described by the manufacturer (Ambion, Inc., Austin, TX). RACE products were separated by electrophoresis on a 1.0% agarose gel, excised, and gel-extracted (GenElute, Sigma Chemical Co., St. Louis, MO). Upon nucleotide sequencing to confirm product identity (ABI 377 sequencer, Applied Biosystems, Foster City, CA), sequences were analyzed and assembled using EditSeq (DNASTAR, Inc., Madison, WI). To confirm *oDlk1* cDNA sequences, first-strand cDNA was oligo-dT primed and synthesized from DNaseI-treated RNA from opossum kidney, liver, and brain using Superscript II as recommended by the manufacturer (Invitrogen, Carlsbad, CA). RT-PCR in opossum young kidney, liver, and brain was performed using exon 1 primer *DLKF1* 5'-ATG TTC GCC TTG GGA TCC TT and exon 5 primer *CS-DLKe5.R2* 5'-AAT GCC AAG TAC AGT CGT CT with 1.5 U Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, CA), 10 pmol primers, 1.5 mM MgCl_2 , and 10 mM dNTPs in a 25- μl PCR reaction volume ($94^{\circ}\text{C} \times 30$ sec, $65^{\circ}\text{C} \times 30$ sec, and $72^{\circ}\text{C} \times 1$ min 10 sec for 35 cycles). Amplicons were subcloned into the pGEMT-easy vector, transformed, and plated according to the manufacturer's instructions (Promega, Madison, WI). DNA from single-colony-forming units was amplified by whole-cell PCR using standard T7 and SP6 primers and sequenced with an ABI 377 sequencer (Applied Biosystems, Foster City, CA). Each *oDlk1* transcript was aligned to the *M. domestica Dlk1* region (Ensembl scaffold_5934) using GeneJockey (Biosoft, Cambridge, UK).

Opossum *Dlk1* imprint analysis. DNA was isolated from opossum brain, liver, and kidney in Qiagen buffer ATL and proteinase K (Qiagen Sciences, Inc., Valencia, CA) followed by phenol:chloroform:isoamyl alcohol extraction and ethanol precipitation. Each animal was screened for polymorphisms in exons 4 and 5 of *oDlk1* by genomic DNA PCR. For exon 4, we used forward primer *CS-DLKe4.F1* 5'-ATC TCC AGC TGT GTG CT and reverse primer *CS-DLKe4.R1* 5'-GAC CTT GTG TTA TTA ATG G with 1.5 U Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, CA), 10 pmol

primers, 1.5 mM MgCl_2 , and 10 mM dNTPs in a 25- μl PCR reaction volume ($94^{\circ}\text{C} \times 30$ sec, $55^{\circ}\text{C} \times 30$ sec, and $72^{\circ}\text{C} \times 30$ sec for 35 cycles). For exon 5, PCR was conducted with forward primer *CS-DLKe5.F1* 5'-GAC AAT GGC CTT GCT TCT TA and reverse primer *CS-DLKe5.R2* 5'-AAT GCC AAG TAC AGT CGT CT using 1.5 U Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, CA), 10 pmol primers, 1.5 mM MgCl_2 , and 10 mM dNTPs in a 25- μl PCR reaction volume ($94^{\circ}\text{C} \times 30$ sec, $60^{\circ}\text{C} \times 30$ sec, and $72^{\circ}\text{C} \times 30$ sec for 35 cycles).

Following identification of heterozygous polymorphic individuals, total RNA was isolated from brain, liver, and kidney of three young opossum offspring by homogenization in RNA-Stat 60 (Tel-Test, Friendswood, TX); subsequent processing was performed as recommended by the manufacturer. The polymorphisms identified and used for imprint analysis (nucleotide 471, G/A and nucleotide 623, C/G, GenBank accession No. AY972866) are in exons 4 and 5, respectively. Although *oDlk1.01*, *oDlk1.02*, and *oDlk1.03* contain exons 4 and 5, we were able to discern between the three alternatively spliced transcripts by amplification of entire cDNA sequences since they varied in length (Fig. 1B). First-strand cDNA was oligo-dT primed and synthesized from DNaseI-treated RNA using Superscript II as recommended by the manufacturer (Invitrogen, Carlsbad, CA). *oDlk1* primers were used to amplify cDNA by nested PCR using exon 1 forward primer *DLKF1* 5'-ATG TTC GCC TTG GGA TCC TT and exon 5 reverse primer *CS-DLKe5.R2* 5'-AAT GCC AAG TAC AGT CGT CT for round 1 followed by exon 1 forward primer *DLKF2* 5'-GGG ATC CTT TCT GCT GTT G and exon 5 reverse primer *CS-DLKR1* 5'-GAA TTC TCT GGA CCC ACT TG) for round 2. We used 1.5 U Platinum *Taq* DNA polymerase (Invitrogen, Carlsbad, CA), 10 pmol primers, 1.5 mM MgCl_2 , and 10 mM dNTPs in a 25- μl PCR reaction volume ($94^{\circ}\text{C} \times 30$ sec, $65^{\circ}\text{C} \times 30$ sec or $72^{\circ}\text{C} \times 1$ min 10 sec for 35 cycles). RT-PCR products were separated by electrophoresis on a 1.5% agarose gel, and appropriately sized fragments of cDNA were excised and gel-extracted (GenElute, Sigma Chemical Co., St. Louis, MO). Products were sequenced (ABI 377 sequencer, Applied Biosystems, Foster City, CA) and analyzed for expression using EditSeq (DNASTAR, Inc., Madison, WI).

Global comparisons of *DLK1*-*GTL2*/*MEG3* genomic regions. NCBI and Ensembl were used for Blast searches (<http://www.ncbi.nlm.nih.gov/BLAST/>; <http://www.ensembl.org/Multi/blastview>). Translational alignments of *Dlk1* sequences were performed with ClustalW (<http://www.ebi.ac.uk/>

clustalw/). CpG islands and gene predictions were identified using GrailEXP (<http://compbio.ornl.gov/grailexp/>) and WebGene (<http://www.itb.cnr.it/sun/webgene/>) under default parameters. RepeatMasker (<http://www.repeatmasker.org/>) was used to identify simple and complex repeats. Sequence alignments and ORF predictions were performed using GeneJockey (Biosoft, Cambridge, United Kingdom).

Results and discussion

***Dlk1* transcripts in the opossum.** A search of the Ensembl genome assembly of *M. domestica* initially identified an opossum homolog to the final coding exon of mouse *Dlk1* (exon 5). 5' RACE experiments were subsequently conducted to determine the entire mRNA sequence of opossum *Dlk1* (*oDlk1*) (GenBank accession No. AY972866). A Blast search (NCBI) with *oDlk1* aligned to *Dlk1* sequences of other species but not to other members of the delta/notch/serrate family of signaling molecules. The genomic sequence of *oDlk1* encompasses 33,728 bp. Examination of *Dlk1* in all eutherian species available demonstrates that the overall exon/intron structure of this gene has been conserved for over 200 million years and is organized into five exons (Fig. 1A).

Using a forward primer in exon 1 and a reverse primer in exon 5, RT-PCR amplification of total RNA from opossum tissues of endodermal (liver), ectodermal (brain), and mesodermal (kidney) origin identified three distinct *oDlk1* transcripts in the representative tissues from all three germ layers. RT-PCR products obtained were too small to originate from genomic cross-contamination since genomic DNA would have produced a band size of around 33 kb. Instead, these bands range from 800 to 1100 bp, consistent with the sizes expected for cDNA products (Fig. 1A,B). Sequence analysis of these bands revealed that the largest (*oDlk1.01*) contains the five exons normally present in eutherian mammals, whereas the alternatively spliced product *oDlk1.02* comprised exons 1, 4, and 5, and *oDlk1.03* contained only exons 1 and 5 (Fig. 1A). Analysis of the *oDlk1* cDNA intron/exon boundaries by alignment to opossum genomic DNA (Ensembl scaffold_5934) identified consensus splice sites flanking each exon. Therefore, opossum *Dlk1* is unique relative to that found in eutherian mammals because *oDlk1* is spliced into multiple isoforms. A single alternate transcript of *Dlk1* involving splice sites within exon 5 has been identified in both the mouse (Ensembl gene ENSMUSG00000040856) and human (Ensembl gene ENSG00000185559) EST databases. However, a search of NCBI EST and Ensembl genomic databases, including sequences from

all available cDNA libraries, identified no isoforms homologous to *oDlk1.02* or *oDlk1.03* in any eutherian mammals to date. Thus, our analysis of marsupial *Dlk1* provides the first description of novel *Dlk1* isoforms.

Comparisons between translated *oDlk1.01* and human *DLK1* (GenBank accession No. P80370) revealed a number of conserved features, including a signal peptide, six EGF-like repeats, 18 disulfide bonds, and putative transmembrane and cytoplasmic regions (Fig. 1C). Only four of these EGF-like domains are conserved in *oDlk1.02*, and this isoform contains 12 of the 18 disulfide bonds found in full-length, functional *DLK1* (Fig. 1C). *oDlk1.03* has a stop codon at the exon 1/5 junction that would result in the translation of only a signal peptide. Thus, this alternatively spliced transcript is predicted to be nonfunctional.

The minimum number of EGF-like repeats required for full *DLK1* function is presently unknown (Fleming 1998). Significant Notch inhibition is associated only with full-length, folded *Dlk1* peptides, suggesting that the conformational state of the full-length *Dlk1* molecule is necessary for protein function. Yeast two-hybrid experiments designed to investigate interactions between specific *Dlk1* EGF-like repeats and Notch1 have demonstrated that *Dlk1* EGFs 1/2 and 5/6 are the primary sources of interaction (Baladron et al. 2005). *oDlk1.02* contains only EGFs 3, 4, 5, and 6. Thus, the protein encoded by the *oDlk1.02* transcript (Fig. 1A,C) may not be fully functional without the complete complement of EGF-like domains and disulfide bonds or it may have a different function than full-length *oDlk1.01*.

***Dlk1* imprint status in the opossum.** Previous studies concerning the evolution of *IGF2* and *M6P/IGF2R* imprinting provided evidence that imprinting evolved with the advent of live birth in therian mammals (O'Neill et al. 2000; Killian et al. 2001a, 2001b). In eutherians, *DLK1/MEG3* forms a JRG motif similar to that of *IGF2/H19* (Wylie et al. 2000). Although *Igf2* is imprinted in the opossum (O'Neill et al. 2000; Killian et al. 2001b), it was unknown if *oDlk1* expression is similarly regulated. In human, mouse, and sheep, *DLK1* is monoallelically expressed from the paternal allele (Schmidt et al. 2000; Wylie et al. 2000; Charlier et al. 2001). In contrast, when we analyzed the cDNA of the three alternatively spliced *Dlk1* transcripts identified in the opossum, each isoform was biallelically expressed in tissues from all three germ layers as indicated by retention of the polymorphism present in the DNA. Thus, none of the *oDlk1* transcripts are imprinted, including the eutherian homolog *oDlk1.01* (Fig. 2).

Fig. 1. Opossum *Dlk1* structure and expression. **(A)** Alternatively spliced opossum *Dlk1* transcripts. Genomic organization of the five exons of *Dlk1* (E1–E5) are marked by gray boxes; horizontal arrow above the exons shows the orientation of gene transcription. The *oDlk1.0* transcript contains all five exons; the *oDlk1.02* transcript contains exons 1, 4, and 5; and the *oDlk1.03* transcript contains exons 1 and 5. **(B)** Agarose gel of opossum alternatively spliced *Dlk1* transcripts in liver, kidney, and brain. Opossum alternatively spliced transcripts *oDlk1.01*, *oDlk1.02*, and *oDlk1.03* are marked by white arrowheads at approximately 1100, 900, and 800 bp, respectively. Observed smearing is presumably a result of oligo-dT priming of total RNA. **(C)** Conserved protein domains for alternatively spliced opossum *Dlk1* transcripts. Transcripts *oDlk1.01* and *oDlk1.02* contain conserved protein domains, whereas *oDlk1.03* does not translate into a significant open reading frame. Conserved protein features were identified by amino acid alignment to human *DLK1* (GenBank accession No. P80370). Exons are shown as gray boxes and labeled E1–E5; percent homology to human *DLK1* is shown.

Fig. 2. Opossum *Dlk1* expression and imprint status. The C/G polymorphism at nucleotide 623 (GenBank accession No. AY972866) and the G/A polymorphism at nucleotide 471 (GenBank accession No. AY972866) in *oDlk1* transcripts were used to determine their imprint status in liver (endoderm), kidney (mesoderm), and brain (ectoderm). Tissues from three different animals were used, and all *oDlk1* transcripts were biallelically expressed in all tissues. Representative genomic and cDNA sequences from one individual are shown.

Phylogenetic comparison of DLK1/MEG3 orthologs. Because opossum *Dlk1* is not imprinted, we compared the genomic region containing *oDlk1* to orthologous imprinted regions in humans and mice (Wylie et al. 2000; Charlier et al. 2001; Paulsen et al. 2001). It has been suggested that imprinted expression of the *Dlk1/Meg3* domain in mice is regulated by a control element residing in a differentially methylated CpG island that lies at the 5' end of *Meg3* (Croteau et al. 2003) and contains CTCF-binding site consensus sequences (Wylie et al. 2000). Deletion of this differentially methylated region results in specific biallelic expression of maternally transcribed genes within the entire 1-Mb imprinted domain on mouse Chromosome 12, consistent with the CpG island serving as an imprint control element for the maternally imprinted genes in this region (Lin et al. 2003).

Comparison of human and mouse *DLK1/MEG3* with the orthologous region in the opossum revealed significant changes in domain structure concomitant with the evolution of imprinting at *DLK1*. In particular, a search of the sequence 130-kb 3' of *oDlk1* (Ensembl scaffold_5934.15000001-20000000.4000000-

3120001, nucleotides 73,667–203,667) failed to identify opossum *Meg3*, indicating an absence of this gene from the opossum *Dlk1* domain (Fig. 3). Furthermore, analysis of this region 3' of *oDlk1* with GrailEXP and WebGene did not detect any predicted transcript that could function in a manner similar to that of *Meg3*. We also conducted Blastn and Blastx searches of the *M. domestica* genome assembly and found no sequences with homology to *Meg3* throughout the entire genome (http://pre.ensembl.org/Monodelphis_domestica/). Finally, the CpG island and CTCF-binding sites 5' to *Meg3* that are differentially methylated in imprinted eutherian mammals (Takada et al. 2000; Wylie et al. 2000) are missing from the sequences 130 kb downstream of *oDlk1*.

To extend our search for *Meg3* sequences beyond the opossum, we conducted several additional Blast searches. The most highly conserved *Meg3* sequences (GenBank accession No. NM_144513, nucleotides 44–908) were blasted as both mRNA and translated sequence against the NCBI nonredundant, month, chromosome, and EST databases and Ensembl genomic databases for all available species outside placental mammals. No homology to

Fig. 3. Alignment of the most highly conserved elements between the human (GenBank accession Nos. AL117090, AL132711) and mouse *DLK1* domain (GenBank accession No. AJ320506) (Paulsen et al. 2001) to opossum (Ensembl scaffold_5934.15000001–20000000). Exons of *DLK1* and *MEG3* are marked by black rectangles; horizontal arrows above the exons show the direction of gene transcription. The open diamond with black bars represents an L1 retrotransposon with GA repeats and is a putative primordial *Meg3* insertion site.

conserved *Meg3* sequences was identified in chicken (*Gallus gallus*), fish, (*Danio rerio*, *Fugu rubripes*, *Tetraodon nigroviridis*), frog (*Xenopus tropicalis*), nematode (*Caenorhabditis elegans*), fly (*Drosophila melanogaster*), or yeast (*Saccharomyces cerevisiae*). These results support our finding that the imprinted noncoding RNA *Meg3* sequences are eutherian-specific. Indeed, sequences homologous to *Meg3* were identified in the following eutherian mammals: human (*Homo sapiens*), chimp (*Pan troglodytes*), rat (*Rattus norvegicus*), mouse (*Mus musculus*), sheep (*Ovis aries*), and dog (*Canis familiaris*).

Figure 3 shows the elements most highly conserved between mouse and human within this domain (Paulsen et al. 2001). These include a CpG island at the fifth exon of *DLK1*, *DAT* (*DLK1-Associated Transcript*) sequences extending from the 3' coding region of *DLK1*, the *CLPG1* gene (Freking et al. 2002), intergenic CG-rich repeats, repetitive sequences overlapping ancestral LINE elements, and the *MEG3* gene. Orthologous sequences in the opossum do not contain intergenic CpG islands or CG-rich repeats that are absent from the eutherian domain. However, within the 122-kb 3' of *oDlk1* we have identified a region (Ensembl contig_86684.17914–18270) that is 66% homologous to the *Bcnt* (*Bucentaur*) pseudogene (Iwashita et al. 2003). In addition, there is a GA-rich region juxtaposed to a full-length LINE-1 reverse transcriptase element (Ensembl scaffold_5934.15727506–15725767) approximately 82 kb downstream from

the final coding exon of *oDlk1*. Both the LINE-1 element and *Bcnt* pseudogene region are not found in orthologous mouse and human sequences. The only defined element conserved between marsupial and eutherian mammals is a CpG island that lies at the promoter and first exon of *Dlk1* (Fig. 3). Thus, the absence of the necessary imprinting factors within the mouse and human JRIG motif from within the opossum *Dlk1* domain is consistent with our finding that *oDlk1* is biallelically expressed.

The relationships between the components of the Notch signaling pathway is crucial in determining cell fate. Biallelic expression of *Dlk1* relative to that of other pathway components with which it interacts has deleterious consequences in eutherian mammals as evidenced by the severity and lethality of pUPD for Chromosome 14 (Georgiades et al. 1998, 2000; Sutton et al. 2003). Thus, imprinting of *Dlk1* is postulated to be a novel mechanism for maintaining appropriate dosage levels for this gene (Wylie et al. 2000).

The necessity for tight regulation of *Dlk1* may also explain the presence of multiple *oDlk1* transcripts. Given the divergence of protein architecture between the two *oDlk1* transcripts that contain open reading frames, biallelic expression of *oDlk1.01* may not result in a doubling of available protein product if the truncated *oDlk1.02* is functionally deficient and competes for Notch receptors. Thus, in the absence of imprinting machinery for *oDlk1*, marsupials may have used dominant negative regulation in

order to control *oDlk1* expression levels. The introduction of *Meg3* and its associated sequences then precluded the need for alternatively spliced transcripts to downregulate *Dlk1* function. Therefore, the acquisition of imprinting at the *Dlk1* locus in placental mammals may have changed both its expression and transcription patterns.

Putative genome reorganization site in opossum *Dlk1* domain. An extensive analysis of the 130 kb of sequence downstream *oDlk1* revealed a genomic feature of particular interest. Approximately 82 kb downstream from the final coding exon of *oDlk1* is a single GA-rich region that shows homology to exon 5 of human *MEG3* and intron 5 of mouse *Meg3*. Juxtaposed to this conserved motif is a full length LINE-1 reverse transcriptase element (Ensembl scaffold_5934.15727506–5725767) that contains two long open reading frames (Fig. 3). Interestingly, the chicken genome also contains a repeat retrotransposon (chicken repeat 1, CR1-F2) in an orthologous position (Ensembl *Gallus gallus* Chromosome 5, nucleotides 45654369–5654586). A region of conservation between mouse and human lies in an orthologous position between *DLK1/MEG3* and has been identified as a partial ancestral LINE element (Paulsen et al. 2001).

LINE 1 elements encode for proteins with endonuclease activity critical for retrotransposition (Feng et al. 1996) and reverse transcriptase activity involved in DNA replication, recombination, and repair, as well as an exonuclease III domain. In addition to providing sites for recombination events within the host genome, active LINE elements are thought to provide reverse transcriptase activity that can result in the movement of genes and pseudogenes (Casavant et al. 2000; Bushman 2004). It has also been proposed that the arrangement of imprinted genes and clusters may have been triggered by recombination through such retrotransposed elements (Walter and Paulsen 2003). For example, imprinting of *U2af1-rs1* appears to have originated from a rodent-specific retrotransposition event (Nabetani et al. 1997). In addition, pseudogenes are often indicative of ancestral retrotranspositional events (Walter and Paulsen 2003). *Mkrn1-p1* and *Nnat* are imprinted in eutherian mammals, and murine pseudogene copies of each have been identified (Evans et al. 2005; Hirotsune et al. 2003). *Nnat* and its pseudogene each contain an L1 element (Evans et al. 2005). Furthermore, the imprinted gene *Znf127* appears to be derived from the *Mkrn1* gene by a retrotransposition event predating the divergence of human and rodent lineages (Gray et al. 2000). Thus, the apparent presence of the *Dlk1*

LINE-1 element in a common ancestor of placental mammals could have served as a reorganization point for the introduction of *Meg3* sequences into the genome of a eutherian base species and triggered imprinting of *Dlk1/Meg3*.

Evolution of imprinted genes. We have recently shown that *Nnat* is also specifically imprinted in eutherians, but, in contrast to *Dlk1*, there is no evidence of *Nnat* being present in noneutherian mammals (Evans et al. 2005). Therefore, the eutherian-specific imprint status of *Dlk1* is unique compared with that of *Nnat* because even though it is present in all therian mammals, it is imprinted only in eutherians. The biallelic expression of *oDlk1* does not necessarily dispute the kinship theory, which states that evolution of genomic imprinting is a consequence of parental conflict based in viviparity (Haig 1997, 2004). Rather, it indicates that the machinery necessary for the imprinting of *Dlk1* is simply not available in the marsupial orthologous domain.

Nnat, *p57(KIP2)*, and *Dlk1* are not imprinted in the marsupial, yet human and mouse orthologs of these genes are expressed with parental bias (Evans et al. 2005; Suzuki et al. 2005). Consequently, we propose that after the original subset of genes was imprinted in therian base species during the Jurassic era (Killian et al. 2000, 2001a,b; O'Neill et al. 2000), a second wave of imprinting occurred in the eutherian base species during the Cretaceous era. This supports the postulate that following the initial evolution of genomic imprinting with the emergence of intrauterine gestation in mammals, other genes continued to acquire imprint control through selective force or even chance (Kaneko-Ishino et al. 2003).

Our hypothesis also supports the postulate that the rapid evolution of imprinted genes contributes significantly to mammalian speciation (Vrana et al. 1998). Many imprinted genes such as *IGF2* show evolutionary conservation of imprinting throughout therian mammal radiation (Murphy and Jirtle 2003). However, genes such as *DLX5* (Okita et al. 2003; Kimura et al. 2004; Horike et al. 2005) and *M6P/IGF2R* (Killian et al. 2001b) demonstrate divergent imprint status between mouse and human. It is possible that when two species hybridize, genes whose imprint status is species-dependent may block the formation of hybrid zygotes through hybrid sterility or inviability, or even provide a selective advantage to surviving hybrids that then become newly evolving species. Accordingly, we propose that a second wave of genomic imprinting contributed to the formation

of a larger placenta capable of meeting the increased gestational demands incurred during eutherian development.

This study of opossum *Dlk1* demonstrates for the first time that the JRIG domains identified in eutherian mammals did not arise in a single common ancestor. We have described multiple, biallelically expressed *oDlk1* transcripts, in contrast to the single, imprinted transcript previously identified in eutherian mammals. Phylogenetic comparisons of orthologous regions in human, mouse, and opossum revealed that the nonimprinted *oDlk1* lies within a genomic domain that is markedly different than eutherian *DLK1* imprinted domains. These comparisons indicate that without *Meg3*, noncoding RNAs, and the sequences surrounding this gene, insufficient signals exist to imprint opossum *Dlk1*. We have also identified a candidate region that may have resulted in the transposition of *Meg3* sequences into the *Dlk1* domain of a eutherian base species, thus facilitating the establishment of imprinting at this locus. Together our findings indicate that after the initial evolution of imprinted genes in therian base species during the Jurassic era, a second wave of genomic imprinting occurred within the eutherian base species.

Acknowledgments

The authors thank Heather Evans for critical review of the manuscript and Michelle Clamp of the Broad Institute for her help in accessing updated *M. domestica* sequence data. This study was supported by NIH grants CA25951, ES08823, and ES13053.

Website References

- http://pre.ensembl.org/Monodelphis_domestica/, Ensembl
- <http://www.ensembl.org/Multi/bblastview,Ensembl>
- <http://www.ebi.ac.uk/clustalw/>, ClustalW
- <http://compbio.ornl.gov/grailxp/>, GrailEXP
- <http://www.repeatmasker.org/>, RepeatMasker
- <http://www.ncbi.nlm.nih.gov/BLAST/>, NCBI
- <http://www.itb.cnr.it/sun/webgene/>, WebGene

References

1. Amarger V, Nguyen M, Van Laere AS, Braunschweig M, Nezer C, et al. (2002) Comparative sequence analysis of the INS-IGF2-H19 gene cluster in pigs. *Mamm Genome* 13, 388–398
2. Astuti D, Latif F, Wagner K, Gentle D, Cooper WN, et al. (2005) Epigenetic alteration at the *DLK1*-*GTL2* imprinted domain in human neoplasia: analysis of neuroblastoma, pheochromocytoma and Wilms' tumour. *Br J Cancer* 92, 1574–1580
3. Baladron V, Ruiz-Hidalgo MJ, Nueda ML, Diaz-Guerra MJ, Garcia-Ramirez JJ, et al. (2005) *dlk* acts as a negative regulator of Notch 1 activation through interactions with specific EGF-like repeats. *Exp cell Res* 303, 343–359
4. Bushman F (2004) Gene regulation: selfish elements make a mark. *Nature* 429, 253–255
5. Carlsson C, Tornehave D, Lindberg K, Galante P, Billestrup N, et al. (1997) Growth hormone and prolactin stimulate the expression of rat preadipocyte factor-1/delta-like protein in pancreatic islets: molecular cloning and expression pattern during development and growth of the endocrine pancreas. *Endocrinology* 138, 3940–3948
6. Casavant NC, Scott L, Cantrell MA, Wiggins LE, Baker RJ, et al. (2000) The end of the LINE?: lack of recent L1 activity in a group of South American rodents. *Genetics* 154, 1809–1817
7. Charlier C, Segers K, Wagenaar D, Karim L, Berghmans S, et al. (2001) Human-ovine comparative sequencing of a 250-kb imprinted domain encompassing the calpyge (*clpg*) locus and identification of six imprinted transcripts: *DLK1*, *DAT*, *GTL2*, *PEG11*, *antiPEG11*, and *MEG8*. *Genome Res* 11, 850–862
8. Croteau S, Charron MC, Latham KE, Naumova AK (2003) Alternative splicing and imprinting control of the *Meg3/Gtl2-Dlk1* locus in mouse embryos. *Mamm Genome* 14, 231–241
9. Davis E, Jensen CH, Schroder HD, Farnir F, Shay-Hadfield T, et al. (2004) Ectopic expression of *DLK1* protein in skeletal muscle of padummal heterozygotes causes the calpyge phenotype. *Curr Biol* 14: 1858–1862
10. Evans HK, Wylie AA, Murphy SK, Jirtle RL (2001) The neuronatin gene resides in a "micro-imprinted" domain on human chromosome 20q11.2. *Genomics* 77, 99–104
11. 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
12. Falls JG, Pulford DJ, Wylie AA, Jirtle RL (1999) Genomic imprinting: implications for human disease. *Am J Pathol* 154, 635–647
13. Feng Q, Moran JV, Kazazian HH Jr, Boeke JD (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87, 905–916
14. Ferguson-Smith AC (2000) Genetic imprinting: silencing elements have their say. *Curr Biol* 10, R872–R875
15. Fleming RJ (1998) Structural conservation of Notch receptors and ligands. *Semin Cell Dev Biol* 9, 599–607
16. Freking BA, Murphy SK, Wylie AA, Rhodes SJ, Keele JW, et al. (2002) Identification of the single base change

- causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res* 12, 1496–1506
17. Garcés C, Ruiz-Hidalgo MJ, Bonvini E, Goldstein J, Laborda J (1999) Adipocyte differentiation is modulated by secreted delta-like (*dlk*) variants and requires the expression of membrane-associated *dlk*. *Differentiation* 64: 103–114
 18. Georgiades P, Chierakul C, Ferguson-Smith AC (1998) Parental origin effects in human trisomy for chromosome 14q: implications for genomic imprinting. *J Med Genet* 35, 821–824
 19. Georgiades P, Watkins M, Surani MA, Ferguson-Smith AC (2000) Parental origin-specific developmental defects in mice with uniparental disomy for chromosome 12. *Development* 127, 4719–4728
 20. 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
 21. Haig D (1997) Parental antagonism, relatedness asymmetries, and genomic imprinting. *Proc R Soc Lond B Biol Sci* 264, 1657–1662
 22. Haig D (2004) Genomic imprinting and kinship: how good is the evidence? *Annu Rev Genet* 38, 553–585
 23. Haig D, Graham C (1991) Genomic imprinting and the strange case of the insulin-like growth factor II receptor. *Cell* 64, 1045–1046
 24. Hansen LH, Madsen B, Teisner B, Nielsen JH, Billestrup N (1998) Characterization of the inhibitory effect of growth hormone on primary preadipocyte differentiation. *Mol Endocrinol* 12, 1140–1149
 25. Hirotsune S, Yoshida N, Chen A, Garrett L, Sugiyama F, et al. (2003) An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene. *Nature* 423, 91–96
 26. Horike S, Cai S, Miyano M, Cheng JF, Kohwi-Shigematsu T (2005) Loss of silent-chromatin looping and impaired imprinting of *DLX5* in Rett syndrome. *Nat Genet* 37, 31–40
 27. Iwashita S, Osada N, Itoh T, Sezaki M, Oshima K, et al. (2003) A transposable element-mediated gene divergence that directly produces a novel type bovine Bcl2 protein including the endonuclease domain of RTE-1. *Mol Bio Evol* 20, 1556–1563
 28. John RM, Aparicio SA, Ainscough JF, Arney KL, Khosla S, et al. (2001) Imprinted expression of neuronatin from modified BAC transgenes reveals regulation by distinct and distant enhancers. *Dev Biol* 236, 387–399
 29. Kaneko-Ishino T, Kohda T, Ishino F (2003) The regulation and biological significance of genomic imprinting in mammals. *J Biochem (Tokyo)* 133: 699–711
 30. Kaneta M, Osawa M, Sudo K, Nakauchi H, Farr AG, et al. (2000) A role for *pref-1* and *HES-1* in thymocyte development. *J Immunol* 164, 256–264
 31. Killian JK, Byrd JC, Jirtle JV, Munday BL, Stoskopf MK, et al. (2000) *M6P/IGF2R* imprinting evolution in mammals. *Mol Cell* 5, 707–716
 32. Killian JK, Nolan CM, Stewart N, Munday BL, Andersen NA, et al. (2001a) Monotreme *IGF2* expression and ancestral origin of genomic imprinting. *J Exp Zool* 291, 205–212
 33. Killian JK, Nolan CM, Wylie AA, Li T, Vu TH, et al. (2001b) Divergent evolution in *M6P/IGF2R* imprinting from the Jurassic to the Quaternary. *Hum Mol Genet* 10, 1721–1728
 34. Kimura MI, Kazuki Y, Kashiwagi A, Kai Y, Abe S, et al. (2004) *Dlx5*, the mouse homologue of the human-imprinted *DLX5* gene, is biallelically expressed in the mouse brain. *J Hum Genet* 49, 273–277
 35. Kobayashi S, Wagatsuma H, Ono R, Ichikawa H, Yamazaki M, et al. (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
 36. Laborda J (2000) The role of the epidermal growth factor-like protein *dlk* in cell differentiation. *Histol Histopathol* 15, 119–129
 37. Li L, Forman SJ, Bhatia R (2005) Expression of *DLK1* in hematopoietic cells results in inhibition of differentiation and proliferation. *Oncogene* 27, 4472–4476
 38. Lin SP, Youngson N, Takada S, Seitz H, Reik W, et al. (2003) Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1-Gtl2* imprinted cluster on mouse chromosome 12. *Nat Genet* 35, 97–102
 39. Lyle R, Watanabe D, de Vrochte D, Lerchner W, Smrzka OW, et al. (2000) The imprinted antisense RNA at the *Igf2r* locus overlaps but does not imprint *Mas1*. *Nat Genet* 25, 19–21
 40. Moon YS, Smas CM, Lee K, Villena JA, Kim KH, et al. (2002) Mice lacking paternally expressed *Pref-1/Dlk1* display growth retardation and accelerated adiposity. *Mol Cell Biol* 22, 5585–5592
 41. Murphy SK, Jirtle RL (2003) Imprinting evolution and the price of silence. *Bioessays* 25, 577–588
 42. Murphy WJ, Eizirik E, Johnson WE, Zhang YP, Ryder OA, et al. (2001) Molecular phylogenetics and the origins of placental mammals. *Nature* 409, 614–618
 43. Murphy SK, Freking BA, Smith TP, Leymaster K, Nolan CM, et al. (2005) Abnormal postnatal maintenance of elevated *DLK1* transcript levels in callipyge sheep. *Mamm Genome* 16, 171–183
 44. Nabetani A, Hatada I, Morisaki H, Oshimura M, Mukai T (1997) Mouse *U2af1-rs1* is a neomorphic imprinted gene. *Mol Cell Biol* 17, 789–798
 45. O'Neill MJ, Ingram RS, Vrana PB, Tilghman SM (2000) Allelic expression of *IGF2* in marsupials and birds. *Dev Genes Evol* 210, 18–20
 46. Okita C, Meguro M, Hoshiya H, Haruta M, Sakamoto YK, et al. (2003) A new imprinted cluster on the human chromosome 7q21-q31, identified by human-

- mouse monochromosomal hybrids. *Genomics* 81, 556–559
47. Onyango P, Miller W, Lehoczy J, Leung CT, Birren B, et al. (2000) Sequence and comparative analysis of the mouse 1-megabase region orthologous to the human 11p15 imprinted domain. *Genome Res* 10, 1697–1710
 48. Paulsen M, Takada S, Youngson NA, Benchaib M, Charlier C, et al. (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
 49. Reik W, Murrell A (2000) Genomic imprinting: silence across the border. *Nature* 405, 408–409
 50. Reik W, Walter J (2001) Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2, 21–32
 51. Schmidt JV, Matteson PG, Jones BK, Guan XJ, Tilghman SM (2000) The *Dlk1* and *Gtl2* genes are linked and reciprocally imprinted. *Genes Dev* 14, 1997–2002
 52. Sleutels F, Zwart R, Barlow DP (2002) The non-coding *Air* RNA is required for silencing autosomal imprinted genes. *Nature* 415, 810–813
 53. Smas CM, Sul HS (1993) Pref-1, a protein containing EGF-like repeats, inhibits adipocyte differentiation. *Cell* 73, 725–734
 54. Smit M, Segers K, Carrascosa LG, Shay T, Baraldi F, et al. (2003) Mosaicism of Solid Gold supports the causality of a noncoding A-to-G transition in the determinism of the callipyge phenotype. *Genetics* 163, 453–456
 55. Sutton VR, McAlister WH, Bertin TK, Kaffe S, Wang JC, et al. (2003) Skeletal defects in paternal uniparental disomy for chromosome 14 are re-capitulated in the mouse model (paternal uniparental disomy 12). *Hum Genet* 113, 447–451
 56. Suzuki S, Renfree MB, Pask AJ, Shaw G, Kobayashi ST, et al. (2005) Genomic imprinting of *IGF2*, *p57(KIP2)* and *PEG1/MEST* in a marsupial, the tammar wallaby. *Mech Dev* 122, 213–222
 57. Takada S, Tevendale M, Baker J, Georgiades P, Campbell E, et al. (2000) Delta-like and *gtl2* are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. *Curr Biol* 10, 1135–1138
 58. Takada S, Paulsen M, Tevendale M, Tsai CE, Kelsey G, et al. (2002) Epigenetic analysis of the *Dlk1*-*Gtl2* imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with *Igf2*-*H19*. *Hum Mol Genet* 11, 77–86
 59. Vrana PB, Guan XJ, Ingram RS, Tilghman SM (1998) Genomic imprinting is disrupted in interspecific *Peromyscus* hybrids. *Nat Genet* 20, 362–365
 60. Walter J, Paulsen M (2003) The potential role of gene duplications in the evolution of imprinting mechanisms. *Hum Mol Genet* 12 Spec No 2, R215–R220
 61. Weidman JR, Murphy SK, Nolan CM, Dietrich FS, Jirtle RL (2004) Phylogenetic footprint analysis of *IGF2* in extant mammals. *Genome Res* 14, 1726–1732
 62. Wylie AA, Murphy SK, Orton TC, Jirtle RL (2000) Novel imprinted *DLK1/GTL2* domain on human chromosome 14 contains motifs that mimic those implicated in *IGF2/H19* regulation. *Genome Res* 10, 1711–1718