The Evolution of the \textit{DLK1-DIO3} Imprinted Domain in Mammals

Carol A. Edwards\textsuperscript{1}, Andrew J. Mungall\textsuperscript{2}, Lucy Matthews\textsuperscript{2}, Edward Ryder\textsuperscript{3}, Dianne J. Gray\textsuperscript{1}, Andrew J. Pask\textsuperscript{4}, Geoffrey Shaw\textsuperscript{4}, Jennifer A.M. Graves\textsuperscript{5}, Jane Rogers\textsuperscript{2}, the SAVOIR consortium, Ian Dunham\textsuperscript{2}, Marilyn B. Renfree\textsuperscript{4}, Anne C. Ferguson-Smith\textsuperscript{1}\textsuperscript{*}

\textsuperscript{1}Department of Physiology, Development, and Neuroscience, University of Cambridge, Cambridge, United Kingdom, \textsuperscript{2}Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, \textsuperscript{3}Department of Genetics, University of Cambridge, Cambridge, United Kingdom, \textsuperscript{4}Department of Zoology, University of Melbourne, Victoria, Australia, \textsuperscript{5}Research School of Biological Sciences, The Australian National University, Canberra, Australia

A comprehensive, domain-wide comparative analysis of genomic imprinting between mammals that imprint and those that do not can provide valuable information about how and why imprinting evolved. The imprinting status, DNA methylation, and genomic landscape of the \textit{Dlk1-Dio3} cluster were determined in eutherian, metatherian, and prototherian mammals including tammar wallaby and platypus. Imprinting across the whole domain evolved after the divergence of eutherian from marsupial mammals and in eutherians is under strong purifying selection. The emergence of \textit{Dlk1-Dio3} imprinting in eutherians has occurred on the maternally inherited chromosome and is associated with region-specific resistance to expansion by repetitive elements and the local introduction of noncoding transcripts including microRNAs and C/D small nucleolar RNAs. A recent mammal-specific retrotransposition event led to the formation of a completely new gene only in the eutherian domain, which may have driven imprinting at the cluster.

Introduction

Genomic imprinting is a process that causes genes to be expressed according to their parental origin and is evident in plants and mammals. Many imprinted genes are located in clusters regulated by a single imprinting control element, whose function across the whole imprinted domain depends on DNA methylation acquired differentially in the male and the female germines \cite{1}. It is not known how or why mammalian imprinting evolved; however, its emergence is associated with the evolution of a placenta \cite{2,3}, and the correct dosage of imprinted genes is important in prenatal growth, postnatal metabolism \cite{4}, and neurodevelopment \cite{5}. Where tested, the majority of imprinted genes are expressed and imprinted, sometimes specifically, in the placenta \cite{6}, suggesting that even distantly related placental mammals such as metatherians (marsupials) will have imprinting, while oviparous mammals, the prototherians (monotremes), will not. Assessment of the imprinting status of a few individual mammalian imprinted genes is consistent with these data. The orthologues of four genes imprinted in mouse and human are clearly imprinted in marsupials \cite{7–10}, and no evidence of imprinting has been found in monotremes, although only three genes have been tested to date \cite{8,11,12}.

The \textit{Dlk1-Dio3} imprinted domain in eutherian mammals contains the protein-coding genes \textit{Delta-like homologue 1} (\textit{Dlk1}), \textit{Retrotransposon-like gene 1} (\textit{Rtl1}), and the type 3 deiodinase (\textit{Dio3}) expressed from the paternally inherited chromosome, and multiple long and short non–protein coding RNAs including microRNAs (\textit{miRNA}s) and C/D small nucleolar RNA (\textit{snRNA}) genes expressed solely from the maternally inherited chromosome (Figure 1A). Seven imprinted \textit{miRNA}s are located within \textit{anti-Rtl1}, and over forty are located further downstream including within the \textit{miRNA}-containing gene \textit{Migf} (Figure 1A). All of the genes in the domain are developmentally regulated and expressed in a range of embryonic and extraembryonic cells types with postnatal expression being found predominantly in the brain \cite{13–15}. In mouse, imprinting is regulated by an intergenic differentially methylated region (\textit{IG-DMR}), located 75 kb downstream of \textit{Dlk1}, that becomes methylated during spermatogenesis but remains unmethylated in the maternal germine \cite{16,17}. When a targeted deletion of the \textit{IG-DMR} is inherited maternally, an epigenetic switch occurs causing the maternally inherited chromosome to behave like the paternal chromosome; no effect is seen when the deletion is paternally inherited. The \textit{IG-DMR} is also differentially methylated in human \cite{17}, and recently identified patients with deletions and epimutations in the \textit{DLK1-DIO3} region indicate that this element likely acts as the imprinting control element in human \cite{18}. Tight linkage and strong conservation of \textit{Dlk1} and \textit{Dio3} is maintained in all vertebrates.


Academic Editor: Chris P. Ponting, Oxford University, United Kingdom

Received February 8, 2008; Accepted April 18, 2008; Published June 3, 2008

Copyright: © 2008 Edwards et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abbreviations: ECR, evolutionary conserved region; IG-DMR, intergenic differentially methylated region; LTR, long terminal repeat; \textit{miRNA}, microRNA; \textit{snRNA}, small nucleolar RNA; UTR, untranslated region

* To whom correspondence should be addressed. E-mail: afsmith@mole.bio.cam.ac.uk
The two genes are located 10.5 kb apart in Takifugu rubripes, approximately 370 kb apart in chicken, and 830 kb in human and mouse (Figure 1B).

Results

To determine the sequence and organization of the region in marsupial and monotreme mammals, we cloned and sequenced the region between DLK1 and DIO3 in the platypus, Ornithorhynchus anatinus, and the tammar wallaby, Macropus eugenii. Bacterial artificial chromosome (BAC) clones containing the orthologous DLK1 and DIO3 genes were identified [19]. Thirteen overlapping wallaby BACs and seven overlapping platypus BACs were isolated from genomic libraries, then initially characterized using a parallel landmark content mapping and fingerprinting strategy [20], and sequenced (Figure S1 and Table S1). This genomic sequence represents complete coverage of the domain in both species and was generated independently of the whole-genome sequencing projects for these organisms. The wallaby sequence is 1,510.8 kb and slightly smaller than that of the South American marsupial Monodelphis domestica (1,637.8 kb plus 26 gaps). The marsupial region is therefore approximately twice as long as its eutherian orthologue (Figure 1B). The region in platypus is 594.8 kb, which is 28% smaller than in mouse.

For both wallaby and platypus, DLK1 and DIO3 genes were identified, cDNAs characterized, and the genes subjected to imprinting analysis (Figure 2 and Figures S2 and S3). For wallaby, fetal tissues, yolk sac placenta, and pouch young samples were dissected. Platypus fetal material is unavailable, so the analysis was conducted on primary adult skin fibroblasts cultured from two male and one female platypus; therefore the analysis in that species is limited. Several single nucleotide polymorphisms (SNP) (Figure 2A) were identified for DLK1 from wallaby tissues that included one sample (2386) from a homozygous mother allowing allele-specific segregation of the variant alleles. Direct sequencing of cDNAs amplified across both polymorphisms indicated there was preferential expression from the G/-CTT allele (Figure 2D).

Comparative sequence analysis of the Dlk1-Dio3 genomic landscape between eutherian and other noneutherian mammals can identify the dynamic changes that are associated with and have the potential to contribute to imprinting. Figure 3A and Table 1 show the relative GC and repeat sequence content of the region in seven genomes; three eutherian species (human, mouse, dog), two marsupials (opossum and tammar wallaby), one monotreme (platypus), and one bird (chicken). The eutherian GC content, %CpG and number of CpG islands was significantly higher than the genome average (p < 0.01 using Chi-squared test) in contrast to marsupial and monotreme mammals, and chicken, that all lack imprinting at this domain. Repeat content was analysed using the most recent previously unreleased platypus repeat database (kindly provided by R Hubley, Repeatmasker). Eutherian LINE content is consistent with the genome-wide average; however, there is a paucity of LINEs in the region between Dlk1 and Mirt (miRNA-containing gene) in the eutherians (Table 1). The majority of repeats identified in the DLK-DIO3 domain in the marsupials are LINE1 repeats. This is consistent with the high number of LINEs identified in the opossum genome and suggests that expansion in the DLK-DIO3 region, as in the marsupial genome as a whole, is due to LINE1 insertion. The opossum region has a slightly larger proportion of SINEs than expected from the genome average. The SINE content is also greater in the tammar wallaby, although the whole-genome sequence for this species is not currently available for comparison. The relative repeat content in platypus is greater than eutherians despite the region being smaller in this species (Figure 3A). The majority of repeats in the platypus DLK1-DIO3 region are SINEs and the more ancient LINE2s. Interestingly, there is a notable absence of long terminal repeat (LTR) elements at this locus in platypus (Figure 3A). The chicken region is devoid of any SINE elements which is consistent with the whole genome analysis of this species. Hence platypus and marsupials have greater SINE content in the domain than do the eutherian mammals with imprinting. This is consistent with the SINE depletion previously reported when comparing imprinted with nonimprinted domains in mouse and human [21,22]. Together, these findings indicate that selection against SINE repeats is an evolutionary feature of imprinted domains (see Discussion).

Detailed comparative sequence analysis was conducted...
between the Dlk1-Dio3 domain in the seven vertebrates. Using a threshold of 55% nucleotide sequence identity over 80 bp, which recognizes the Dlk1 exons in all seven sequences, 141 evolutionary conserved regions (ECRs) were identified across sub-groups representing eutherians, marsupials, platypus, and chicken (Table S2). Of the 141 ECRs found, 22.7% (31) were common to all seven vertebrates, 15.6% (22) were common to all mammals, and another 16 were found in all therian mammals. Six were found only in platypus and chicken. Figure 3B illustrates the number of ECRs arranged according to the sub-classes of vertebrates in which they are identified. In mammals, 27.7% were identified in at least one eutherian, one marsupial and platypus, whereas 24.8% were found in at least one species representing each therian infraclass. Although the greatest number of ECRs is found within the mammalian species, the more ancestral ECRs (the 31 found in all species studied) are on average larger, having a mean length of 494 bp compared with the mean length of all ECRs at 340 bp and suggesting greater functional constraint.

We used the 31 ECRs found in all vertebrates to align the Dlk1-Dio3 domain and subdivide it into 30 inter-ECR zones for further comparative analysis (Figure 4A). Exons of Dlk1 and Dio3 are represented by vertebrate ECRs 1–3 and 30–31, respectively. The intergenic distribution of the ECRs is not uniform throughout the domain with two-thirds being located in the 3' half of the domain. One of the ECRs, approximately 3 kb upstream of Dio3, contains a highly conserved putative CTCF binding site in all therian species. The amount of sequence in each of the 30 inter-ECR zones relative to the overall size of the domain was quantified for each vertebrate (Figure 4B). This provides a measure of the overall expansions/contractions between species. The regional changes between marsupial, monotreme, and eutherian mammals across the domain are not uniform. The most striking differences between the mammals lie in zone 3 (between vECRs 3 and vECR4), zone 6, and zone 7. Zone 3 which is located between the last exon of Dlk1 and the conserved intron region of Gt2, is expanded in eutherians (Figure 4B). This expansion does not appear to be caused by LINEs, because LINE1 and LINE2 repeats are equivalently
represented in eutherians and marsupials (Figure 4). As shown, this zone contains a higher proportion of SINE elements than reported for the whole genome and compared with the entire domain. However, the increased SINE content does not explain the expansion of zone 3, which is due to the acquisition of unique sequence, including the imprinting control region (the IG-DMR) and presumably other eutherian specific regulators. In contrast, eutherian zone 6, located between \textit{Gtl2} and \textit{Rtl1}, is smaller than in marsupials, platypus, and chicken, implying either that it contracted or that it is resistant to expansion. This latter explanation is favoured, because in marsupials expansion is predominantly due to \textit{LINE1s}, and in platypus to \textit{LINE2} repeats and \textit{SINEs}. This exclusion suggests an important previously unrecognized eutherian specific function for that zone (Figure 4).

The eutherian specific expansion of zone 7, as for zone 3, is not associated with the insertion of repetitive sequences, compared with marsupials. Rather, zone 7 represents the region located between \textit{Rtl1} and \textit{Mirg}, which, in eutherians, contains approximately 50 miRNA genes and three clusters of C/D snoRNA genes, all expressed from the maternally inherited chromosome [23]. With one exception (see below), our analysis failed to find homologous sequences in marsupials, platypus, or chicken. Instead, the zone contains LINE1s, and in platypus to \textit{LINE2} repeats and \textit{SINEs}. This exclusion suggests an important previously unrecognized eutherian specific function for that zone (Figure 4).

The eutherian specific expansion of zone 7, as for zone 3, is not associated with the insertion of repetitive sequences, compared with marsupials. Rather, zone 7 represents the region located between \textit{Rtl1} and \textit{Mirg}, which, in eutherians, contains approximately 50 miRNA genes and three clusters of C/D snoRNA genes, all expressed from the maternally inherited chromosome [23]. With one exception (see below), our analysis failed to find homologous sequences in marsupials, platypus, or chicken. Instead, the zone contains LINE1s, and in platypus to \textit{LINE2} repeats and \textit{SINEs}. This exclusion suggests an important previously unrecognized eutherian specific function for that zone (Figure 4).

The eutherian specific expansion of zone 7, as for zone 3, is not associated with the insertion of repetitive sequences, compared with marsupials. Rather, zone 7 represents the region located between \textit{Rtl1} and \textit{Mirg}, which, in eutherians, contains approximately 50 miRNA genes and three clusters of C/D snoRNA genes, all expressed from the maternally inherited chromosome [23]. With one exception (see below), our analysis failed to find homologous sequences in marsupials, platypus, or chicken. Instead, the zone contains LINE1s, and in platypus to \textit{LINE2} repeats and \textit{SINEs}. This exclusion suggests an important previously unrecognized eutherian specific function for that zone (Figure 4).

The eutherian specific expansion of zone 7, as for zone 3, is not associated with the insertion of repetitive sequences, compared with marsupials. Rather, zone 7 represents the region located between \textit{Rtl1} and \textit{Mirg}, which, in eutherians, contains approximately 50 miRNA genes and three clusters of C/D snoRNA genes, all expressed from the maternally inherited chromosome [23]. With one exception (see below), our analysis failed to find homologous sequences in marsupials, platypus, or chicken. Instead, the zone contains LINE1s, and in platypus to \textit{LINE2} repeats and \textit{SINEs}. This exclusion suggests an important previously unrecognized eutherian specific function for that zone (Figure 4).
repeats in marsupials and as before, LINE2s and SINEs in platypus (Figure S4). Therefore eutherians acquired transcribed non–protein coding RNAs in a zone that appears resistant to expansion by LINEs and SINEs. Interestingly, the acquisition of snoRNA genes in the imprinted Prader Willi-Angelman syndrome locus also corresponds to the acquisition of imprinting [12].

In the mouse, all the imprinted non–protein coding transcripts in the domain require the imprinting control element and sequences 5’ to Gtl2 for their activity on the maternally inherited chromosome. They are all expressed in the same orientation, and data suggest that they are at least in part associated with a single long transcription unit [17,24]. ECRs specifically associated with Gtl2 were identified by phylogenetic footprinting (Figure 5A). Two approaches were undertaken to determine whether GTL2 and other non–protein coding transcripts were present within the domain; expression analysis of DLK1-DIO3 intergenic ECRs and the amplification from cDNA of randomly selected sequences from the wallaby region (Figure S5A and Table S3). Five mammalian ECRs were found in the vicinity of Gtl2, of which three were common to all vertebrates; one corresponds to exon 5 of NM_144513 (ECR19), and the remainder appear to be intronic. One of the intronic ECRs (ECR18) was previously identified in intron 8 of Y13832 [22]. An additional ECR (ECR14) located close to exon 1 was identified and found to be inverted in eutherians (Figure 5A). This and the three vertebrate ECRs were expressed at very low levels in wallaby tissues, with no transcriptional activity from the other two. RT-PCR analysis of 29 additional, randomly selected sequences in wallaby located between Gtl2 and Mirg identified weak transcriptional activity from five sequences, including one mammalian Mirg–specific ECR (Figure S5). Quantitative RT-PCR comparing the relative expression of ECR19 and one of the random sequences (Ran3) with DIO3 expression in the same samples confirms expression from the GTL2-like locus in marsupials is between 1.1 ± 10^-3 and 4.2 ± 10^-4 lower in fetal head and pouch young brain (Figure 5). Polymorphisms located in ECR19 and the MIRG–ECR were used to demonstrate that this low level of transcription is biallelic (Figure 5B and Figure S4).

It was of particular interest to determine whether the protein-coding, retrotransposon-like gene Rtl1 (also known as Peg11/Mart1) was present in non-eutherian mammals. Rtl1 is a member of the Ty3-Gypsy family of LTR retrotransposons with closest similarity to the Sushi-ichi class [25]. In mouse

Table 1. GC and Repeat Content of the Dlk1-Dio3 Region in Vertebrates

<table>
<thead>
<tr>
<th>Genomic Features</th>
<th>Sub-Class</th>
<th>Human</th>
<th>Mouse</th>
<th>Dog</th>
<th>Wallaby</th>
<th>Opossum</th>
<th>Platypus</th>
<th>Chicken</th>
</tr>
</thead>
<tbody>
<tr>
<td>GC content</td>
<td>%GC in region</td>
<td>48.0</td>
<td>47.0</td>
<td>50.5</td>
<td>38.1</td>
<td>37.0</td>
<td>44.6</td>
<td>38.9</td>
</tr>
<tr>
<td></td>
<td>%GC in genome</td>
<td>40.9</td>
<td>41.8</td>
<td>41.1</td>
<td>37.3</td>
<td>37.7</td>
<td>45.9</td>
<td>41.5</td>
</tr>
<tr>
<td></td>
<td>%CpG in region</td>
<td>3.5</td>
<td>2.5</td>
<td>5.2</td>
<td>0.7</td>
<td>0.7</td>
<td>2.1</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>%CpG in genome</td>
<td>2.0</td>
<td>1.7</td>
<td>2.2</td>
<td>0.9</td>
<td>0.9</td>
<td>2.1</td>
<td>2.1</td>
</tr>
<tr>
<td>Cpg islands &gt;200 bp</td>
<td>41</td>
<td>22</td>
<td>153</td>
<td>12</td>
<td>10</td>
<td>27</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>Number of ECRs</td>
<td>Total</td>
<td>109</td>
<td>81</td>
<td>103</td>
<td>122</td>
<td>125</td>
<td>101</td>
<td>56</td>
</tr>
<tr>
<td>SINEs Total</td>
<td>8.27</td>
<td>6.33</td>
<td>6.45</td>
<td>13.17</td>
<td>12.12</td>
<td>21.36</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Alu/B1</td>
<td>5.60</td>
<td>1.43</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>B2-B4</td>
<td>4.24</td>
<td>0.11</td>
<td>0.55</td>
<td>2.38</td>
<td>10.38</td>
<td>10.47</td>
<td>20.28</td>
<td></td>
</tr>
<tr>
<td>LINEs Total</td>
<td>19.57</td>
<td>13.09</td>
<td>18.12</td>
<td>42.82</td>
<td>39.32</td>
<td>25.16</td>
<td>6.48</td>
<td></td>
</tr>
<tr>
<td>LINE1</td>
<td>17.11</td>
<td>12.87</td>
<td>15.88</td>
<td>26.33</td>
<td>28.21</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>LINE2</td>
<td>2.12</td>
<td>0.16</td>
<td>1.91</td>
<td>7.03</td>
<td>4.72</td>
<td>22.13</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>L3/CR1</td>
<td>0.14</td>
<td>0.01</td>
<td>0.16</td>
<td>5.51</td>
<td>4.16</td>
<td>2.29</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>RTE</td>
<td>0.00</td>
<td>0.00</td>
<td>0.17</td>
<td>3.94</td>
<td>2.22</td>
<td>0.61</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>LTR elements</td>
<td>Total</td>
<td>8.53</td>
<td>8.42</td>
<td>3.94</td>
<td>11.92</td>
<td>10.67</td>
<td>0.02</td>
<td>0.73</td>
</tr>
<tr>
<td>MaLRs</td>
<td>3.39</td>
<td>4.54</td>
<td>2.19</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>ERVL</td>
<td>2.02</td>
<td>1.01</td>
<td>1.45</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>ERV_classI</td>
<td>3.12</td>
<td>0.17</td>
<td>0.28</td>
<td>0.16</td>
<td>10.01</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>ERV_classII</td>
<td>0.00</td>
<td>2.69</td>
<td>0.00</td>
<td>0.99</td>
<td>0.49</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>DNA elements</td>
<td>Total</td>
<td>2.01</td>
<td>1.09</td>
<td>1.43</td>
<td>1.28</td>
<td>1.56</td>
<td>0.89</td>
<td>0.73</td>
</tr>
<tr>
<td>MER1_type</td>
<td>0.78</td>
<td>0.87</td>
<td>0.71</td>
<td>0.32</td>
<td>0.25</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>MER2_type</td>
<td>0.83</td>
<td>0.11</td>
<td>0.32</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Unclassified repeats</td>
<td>0.08</td>
<td>0.35</td>
<td>0.08</td>
<td>0.00</td>
<td>0.00</td>
<td>0.53</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Total interspersed repeats</td>
<td>38.46</td>
<td>29.28</td>
<td>30.02</td>
<td>58.38</td>
<td>63.66</td>
<td>47.96</td>
<td>7.94</td>
<td></td>
</tr>
<tr>
<td>Percent interspersed repeats from vECR1-vECR8(Dlk1-Mirg)</td>
<td>24.83</td>
<td>19.34</td>
<td>19.93</td>
<td>58.12</td>
<td>64.79</td>
<td>46.9</td>
<td>10.78</td>
<td></td>
</tr>
<tr>
<td>SINEs</td>
<td>8.6</td>
<td>5.38</td>
<td>6.55</td>
<td>10.54</td>
<td>10.2</td>
<td>20.36</td>
<td>0.06</td>
<td></td>
</tr>
<tr>
<td>LINEs</td>
<td>10.74</td>
<td>10.5</td>
<td>10.18</td>
<td>45.42</td>
<td>40.33</td>
<td>25.94</td>
<td>8.01</td>
<td></td>
</tr>
<tr>
<td>LTR elements</td>
<td>3.87</td>
<td>2.53</td>
<td>2</td>
<td>0.91</td>
<td>12.64</td>
<td>0.00</td>
<td>1.73</td>
<td></td>
</tr>
<tr>
<td>DNA elements</td>
<td>1.61</td>
<td>0.9</td>
<td>1.26</td>
<td>1.62</td>
<td>0.25</td>
<td>0.99</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>Percent interspersed repeats in genome</td>
<td>45.5</td>
<td>40.9</td>
<td>35.5</td>
<td>52.2</td>
<td>9.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINEs</td>
<td>12.6</td>
<td>7.2</td>
<td>10.2</td>
<td>10.4</td>
<td>0.00</td>
<td>6.5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LINEs</td>
<td>20.0</td>
<td>19.6</td>
<td>18.2</td>
<td>29.2</td>
<td>6.5</td>
<td>0.00</td>
<td>6.5</td>
<td></td>
</tr>
<tr>
<td>LTR elements</td>
<td>8.1</td>
<td>9.8</td>
<td>3.7</td>
<td>10.6</td>
<td>1.3</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>DNA elements</td>
<td>2.8</td>
<td>0.8</td>
<td>1.9</td>
<td>1.7</td>
<td>0.8</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
</tbody>
</table>

doi:10.1371/journal.pbio.0060135.t001
Figure 4. Comparative Analysis of Inter-ECR Zones in the DLK1-DIO3 Domain

(A) ECRs identified in all seven species are shown as blue (Dlk1 and Dio3 exons) or black lines. The ECRs are linked to their orthologues in the neighbouring species in order to illustrate the repeat content and relative expansions/contraction within each sequence.

(B) The length of each inter-ECR zones from vECR1 (DLK1 exon 3) to vECR31 (DIO3) as a proportion of the length of the domain in eutherians, marsupials, platypus, and chicken. Zone 1 = vECR1–vECR2, zone 2 = vECR2–vECR3, etc. Mean ± standard error for the three eutherians and the two marsupials are shown.

doi:10.1371/journal.pbio.0060135.g004
and human, it has lost its LTRs, encodes a protein essential for normal placental development and fetal growth and viability (M. Ito, A. Ferguson-Smith, unpublished data, and [26]), and is expressed from the paternally inherited chromosome. Its levels are regulated by miRNAs processed from an antisense transcript on the maternally inherited chromosome that are 100% complementary to the \textit{Rtl1} mRNA (Figure 1A) [17,27,28]. Another member of this family, \textit{Peg10} located on mouse Chromosome 6, was recently shown to be imprinted in wallaby fetus and placenta (but is absent in the platypus), and its repression on the maternally inherited chromosome is associated with differential methylation in the body of the gene [9]. We could not demonstrate \textit{RTL1} sequences in the platypus or chicken domain. However, we did find sequences related to \textit{Rtl1} in the appropriate position in marsupials but, interestingly, it is extensively degraded with very few regions of homology remaining (Figure 5D). No expression of the most highly conserved region was found in fetal and pouch young tissues (Figure 6A). This suggests that \textit{Rtl1} retrotransposed into the locus prior to the divergence of marsupial and eutherian mammals and, in the absence of functional selection, it degraded in marsupials but acquired a growth regulatory function in eutherians coincident with the evolution of imprinting.

A number of miRNAs that are antisense to \textit{Rtl1} are transcribed from the maternal chromosome in eutherians.

Using the miRNA prediction programme miR-abela [29], no miRNAs were found to be conserved between all vertebrates, and none were conserved between eutherians and marsupials. A single predicted miRNA was conserved between the marsupials (74\% identity) (Figure S6B and S6C). Interestingly, this was located in the vicinity of the eutherian miR127, which is transcribed antisense to \textit{Rtl1} and along with seven others, contributes to the stability of the \textit{Rtl1} mRNA through an RNAi-dependent mechanism [27]; a function that would not be evident in marsupials that lack this gene. The sequence of the predicted processed miRNA from marsupial miR127 through both mammals is shown to be conserved between the two lineages. RT-PCR analysis failed to amplify the primary transcript or predicted hairpin from wallaby fetal head or pouch young brain cDNAs (Figure S6D). These data suggest

Figure 5. Assessment of Noncoding RNA Transcription

(A) Identification of ECRs in the \textit{Gtl2} region in noneutherians. mLAGAN and zPicture alignments of mouse \textit{Gtl2} with human, dog, wallaby, opossum, platypus, and chicken are shown. Four intronic ECRs are identified and one (ECR19) aligns within exon 5 of NM_144513. ECR14 is inverted in the eutherians and was only identified using the zPicture alignment. Weak expression was identified for ECRs14, 15, 18 and 19. RT-PCR for ECR19 in fetal head and pouch young body is shown.

(B) Weak expression from ECR19 in tammar wallaby fetal head and pouch young. An A/G SNP was identified in ECR19 and biallelic expression was observed.

(C) The expression ratio of ECR19 and Random Primer set 3 relative to \textit{DIO3} in fetal head and pouch young brain as calculated by quantitative RT-PCR.

(D) A region orthologous to the retrotransposon-derived gene \textit{Rtl1} was identified in marsupials. The mLAGAN algorithm was used to align human \textit{RTL1} with mouse, dog, wallaby, and opossum. Regions with homology of >75\% over 80 bp are shown in blue. Regions of human \textit{RTL1} with homology to the Sushi-ichi domains are highlighted. Homology between the eutherian and marsupial regions indicates that \textit{RTL1} inserted into the region before the divergence of eutherians and metatherians.

doi:10.1371/journal.pbio.0060135.g005
Figure 6. Methylation Analysis of DLK1 Exon 5 and DIO3 Promoter in Wallaby and Platypus

(A) Hypermethylation was observed in both the wallaby and platypus DLK1 exon 5 regions. Wallaby genomic DNA from d23 RPY fetus (gDNA) and wallaby sperm gDNA was digested with XbaI (Xb), further digested with HpyCH4IV (Hy) and analysed by Southern blot hybridisation using MeDLK1Ex5 as a probe. Platypus gDNA was digested with StuI (St) and further with MspI (Ms), HpaII (Hp), and HhaI (Hh) and analysed by Southern blot hybridisation using OaDLK1Ex5 as a probe.

(B) A map depicting the HpaII and HhaI sites and methylation status in Dlk1 exon 5. Black circles indicate methylated sites, white circles unmethylated sites, and half black circles indicate partial methylation. CpG islands in the region are shown as grey boxes.

(C) The Dio3 promoter region is unmethylated in both wallaby and platypus. Wallaby fetal head gDNA was digested with HindIII and further with MspI (Ms), HpaII (Hp), and HhaI (Hh) and hybridised with MeDIO3CpG. The methylation-sensitive HpaII and HhaI tracks exhibited full digestion indicating the region is unmethylated. Platypus gDNA was digested with XbaI (Xb) then with MspI (Ms), HpaII (Hp), HhaI (Hh), or SmaI (Sm) and hybridised with OaDIO3CpG. High CG content results in many HpaII and HhaI fragments, which are unmethylated and too small to be resolved on this filter. The smallest SmaI site expected was identified, showing the platypus Dio3 promoter is unmethylated. Control hybridisation with a OaDIO3 promoter proximal probe identified a fully methylated XbaI fragment of ~3 kb in the HpaII, HhaI, and SmaI tracks, confirming the integrity of the genomic DNA in these tracks (unpublished data).

doi:10.1371/journal.pbio.0060135.g006
that this is not a functional miRNA, and sequence similarity is due to miR127 being located within RTL1.

A small number of conserved CpG islands and CpG-rich regions were found to be shared between eutherians, marsupials, and platypus and their methylation status was determined. They included the promoters of Dlk1 and Dio3 and the differentially methylated region in the last exon of eutherian Dlk1, known as the Dlk-DMR [16,30]. Each region was analysed by methylation-sensitive Southern blots with genomic DNA from platypus and wallaby and from wallaby sperm. Results are shown in Figure 6. The ECR at intron 5 in Gt2 (Y13832) is CpG-rich, and this too was analysed. As in eutherians, the DLK1 and Dio3 CpG-island promoters are completely unmethylated on both parental chromosomes. The Gt2 ECR is partially methylated on both parental chromosomes in mouse, and has the same pattern in platypus and wallaby. In mouse, the Dlk-DMR is hypermethylated on the paternally inherited chromosome and in sperm, and hypomethylated on the maternally inherited chromosome [16,30]. Platypus and wallaby genomic DNA showed hypomethylation of the locus similar to that seen on the paternal chromosome in the mouse. Wallaby sperm was also hypermethylated. This suggests that the methylation state of the mouse paternal chromosome resembles the methylation state of the mammalian domain prior to the emergence of imprinting and implies that hypomethylation of the maternal chromosome evolved with imprinting.

Discussion

In eutherians, Dlk1 and Dio3 are developmentally important genes that are expressed in numerous embryonic and extraembryonic tissues. Here we have shown that DLK1 and Dio3 are both biallelically expressed in marsupial fetus, placenta, and neonatal pouch young. DLK1 was recently shown to be expressed biallelically in adult brain, liver, and kidney in the South American marsupial, Monodelphis domestica; however, analysis of imprinting in embryonic and extraembryonic tissues was not conducted in that study [31]. We also demonstrate biallelic expression of both genes in platypus. Because fetal material is not available, biallelic expression of these genes during platypus development can only be inferred. Together, our results indicate that imprinting of the whole DLK1-DIO3 domain evolved after the divergence of metatherian and eutherian mammals.

Comparative sequence analysis of the DLK1-DIO3 region in seven different amniote vertebrates (representing Eutheria, Metatheria, Protheria, and Aves) demonstrates that the overall genomic landscape in this region is GC-rich in eutherians but not in the other species studied. It has previously been postulated that GC-rich isochores in eutherians were once located on GC-rich microchromosomes in the ancestral amniote [32]. The elevated GC content in eutherians but not in the noneutherian species suggests that the increase occurred in eutherians rather than existing as an ancient isochore.

A number of results suggest that the DLK1-DIO3 is a recombination hot spot and under purifying selection in eutherian species where it is imprinted. First, elevated GC content correlates with increased levels of recombination [32]. Second, the introns of DLK1 are shorter in the eutherians than in the noneutherian species (Figure S2B), and decreased intron length is associated with high recombination rates [33]. Third, the reduced SINE content in the eutherian indicates the region is under purifying selection, especially because SINEs are usually associated with GC-rich regions. Interestingly, the region between vertebrate ECR1 and ECR8, which encompasses Dlk1, Gt2, Rtl1, snoRNAs, and miRNAs, is particularly devoid of LINEs, indicating that this region is under even greater constraint (Table 1 and Figure 4A). Finally, the eutherian DLK1-DIO3 regions are also all located close to the telomeres, whereas in noneutherian species, they are located mid-chromosome [19]. A correlation of elevated recombination levels at sub-telomeric regions has previously been reported [34–36]; however, it is possible that this sub-telomeric position is the result of increased breakage in GC-rich regions [37]. Imprinted domains have previously been shown to be associated with elevated GC content [38–41], short introns [42], and reduced SINE content when compared to nonimprinted regions in eutherians [21,43]. Our finding that this comparison can be extended to the same domain between mammals that imprint and those that do not strongly suggests that imprinted domains are under purifying selection perhaps to constrain domain size such that cis-acting elements can function correctly.

None of the ECRs maps to the position of the eutherian imprinting control element. Whether any of the ECRs plays a functional role in the regulation of the domain is currently under investigation. Those specific to subgroups such as oviparous vertebrates, or the sixteen ECRs specific to therian mammals, might relate to the regulation of specific functions such as the development of extraembryonic structures in therians.

Expression analysis has provided evidence that Gt2 and other noncoding transcripts existed throughout amniote evolution, suggesting that Gt2 did not arise from an eutherian-specific retrotransposition event that triggered imprinting at the domain as has been previously suggested [31]. Our results show that weak regional non–protein coding transcriptional activity can occur in some places across the domain in noneutherian mammals and suggest that the process repressing the protein-coding genes on the maternal chromosome in eutherians (driven by the imprinting control region upstream from Gt2) facilitated stronger expression from these non–protein coding transcripts. The appearance of functional miRNAs and C/D snoRNAs within the locus may therefore have been a consequence of the acquisition of imprinting with the strongly expressed Gt2 gene, providing an ideal host transcript. It is not known whether the duplications that gave rise to the miRNA clusters occurred before or after evolution of imprinting at the locus. Interestingly, a role for these miRNAs in the trans-regulation of neural and placental processes has been inferred [44]. A functional role for these transcripts in the regulation of the neighbouring imprinted protein-coding genes also cannot be ruled out. Furthermore, the emergence of a regulatory relationship between RTL1 and its reciprocally imprinted miRNA-containing antisense transcript is also intriguing. In contrast to the more distal miRNA clusters to which they are not related, these seven anti-RTL1 miRNAs are not likely to have arisen through duplication/divergence events. Rather, these may have evolved as a host defence mechanism associated with the retrotransposon properties of RTL1.
During the course of evolution, the genomic landscape of the Dlk1-Dio3 region has undergone a number of changes (Figure 7). Most significantly, the region has become imprinted. This analysis has proven that imprinting in this domain emerged after the divergence of marsupials and eutherian mammals. This provides evidence that mammalian imprinting evolved at different loci at different times in eutherian mammals. This gain of function may indeed have driven imprinting at the domain, conferred through the acquisition of the imprinting control element. Gtl2 and associated transcripts became up-regulated on the maternal chromosome in eutherians, and miRNAs and C/D snoRNAs specifically evolved in the region. Once imprinted, gene expression was fixed in the region it underwent purifying selection, correlating with an increase in GC content, reduction in Dlk1 intron size, and selection against SINE and LINE insertions. Comparison of these results with similar detailed analyses on domains acquiring imprinting prior to the divergence of marsupials and eutherians will provide further insight into the relationships between dynamic changes in genomic landscape and the evolution of imprinting.

Materials and Methods

Expression analysis. RNA was extracted using the GenElute mammalian total RNA miniprep kit (Sigma) following the manufacturer’s protocol. cDNA was synthesized using Superscript III Reverse Transcriptase (Invitrogen) following the manufacturer’s instructions. The RT-PCRs were primed using random hexamer primers or the following gene-specific primers; platypus DLK1 5’-GAACGTTTACAAACAGATAATGCT-3’, wallaby DIO3 5’-CGGGGCATCAGAGTTACA-3’, and platypus DIO3 5’-GACTCCGTCTCCGAGAACAT-3’, and 5’-TGAACATCTTACAAAACAAGAAACAT-3’. cDNA was amplified using Hot Start KOD polymerase (Novagen), PCR conditions are as described in [19]. For particularly GC-rich regions (e.g., platypus DIO3) 1X Polymate (Bioline) was also added to the PCR reaction. Primer sequences and annealing temperatures can be found in Table S4.

Allelic discrimination quantitative RT-PCR. Custom TaqMan assays were produced using the Assays-by-Design facility at Applied Biosystems. 1 µl of cDNA was amplified in a 12.5-µl reaction 1X TaqMan Universal PCR Master Mix (Applied Biosystems) and 1X specific assay as per the manufacturer’s instructions. C_{T} (threshold cycle) values for both the VIC and FAM probes were recorded and the difference between them (ΔC_{T}) was calculated. Samples were analysed in triplicate. Genomic DNA from homozygous individuals was used as controls to ensure no cross hybridisation occurred between the two probes. The ΔC_{T} of cDNAs was compared with a standard curve of ΔC_{T} values from two homozygous gDNAs mixed at different ratios (9:1, 9:1, 4:1, 7:3, 3:2, 1:1, 2:3, 3:7, 1:4, 9:1, and 1:49), and the percentage expression from each allele was extrapolated. This
method was adapted from [45]. The primers used were as follows: MeDIO3UTR-F, 5'-CTTCTCCCTCCCTCCCAAATTTCT-3'; MeDIO3UTR-R, 5'-TGCACTGCAAAACTTTGGAGGA-3'; and – allele probe, 5'-VIC-TTCTCCCTCCCTCCCTCCCTC-3', and – allele probe, 5'-FAM-TTTTCTCTCCCTCCCTCCCTC-3'. SYBR green qRT-PCR assays were performed using the SensiMix kit (Quanta). The amplification of each primer pair was determined using a serial dilution of cDNA (1, 1/25, 1/125, 1/625). Reactions were performed in triplicate, and the average C_{T} value of each dilution was used to generate a standard curve. The slope of the curve when plotted to \log_{10} was used to determine the efficiency of amplification (E) for each primer set using the following equation: 
\[ E = \frac{\text{slope}}{\text{slope} - 1} \]
and the relative fold expression calculation 
\[ \frac{\text{CT(sample)}}{E(\text{sample})} \]
was achieved through sequencing of BAC clones indicated in red. The 13 sequenced clones span 1,674,705 bp. BAC details are summarised in Table S1A. BACs identified from library screening were amplified to confirm presence of probe sequences, and HindIII digests and fingerprinting allowed them to be aligned and assembled into a contig. One BAC from each end was selected for shotgun sequencing. Using end sequence or fully sequenced BACs, new probes were generated and the process repeated until complete coverage was achieved.

Supporting Information

**Figure S1.** Overlapping BAC Clones Identified for Sequencing.

(A) Complete coverage of the Dlk1-Dio3 region in Macrobus eugenii was achieved through sequencing of BAC clones indicated in red. The thirteen sequenced clones span 1,674,705 bp. BAC details are summarised in Table S1A.

(B) Complete coverage of the Dlk1-Dio3 region in Ornithorhyncus anatinus was achieved through sequencing BAC clones indicated in red. Seven clones spanning 795,257 bp were sequenced. BAC details are summarised in Table S1B.

The Macrobus eugenii genomic BAC library was from the Arizona Genomics Institute (average insert size 166 kb) covering 11.36 genome equivalents and cloned into the HindIII site of pCU6GBC1. The Ornithorhyncus anatinus BAC library was from Clemson University Genomics Institute (average insert size 143 kb) covering 11 genome equivalents. It was constructed in the HindIII site of pCU6GBC1. Libraries were screened with probes for Dlk1 and Dio3. A Dlk1 probe was generated for wallaby by searching the Monodelphis domestica trace archive with human DLK1 with a probe designed against the most conserved sequence. DLK1 sequence for platypus was identified by searching the platypus trace archive with chicken Dlk1 sequence and a probe was generated to this platypus sequence. Dio3 probes amplified from wallaby and platypus were generated after identification of conserved primers from alignments of human, mouse, rat, and chicken Dio3. BACs identified by library screening were amplified to confirm presence of probe sequences, and HindIII digests and fingerprinting allowed them to be aligned and assembled into a contig. One BAC from each end was selected for shotgun sequencing. Using end sequence or fully sequenced BACs, new probes were generated and the process repeated until complete coverage was achieved.

![Found at doi:10.1371/journal.pbio.0060135.sg001 (1.6 MB TIF).](image)

**Figure S2.** Exon Sequence, Structure, and Polymorphism for Wallaby and Platypus DLK1

(A) mRNA sequence of wallaby and platypus DLK. The protein coding region (CDS) is highlighted turquoise and is in uppercase. The 3' UTR is highlighted in yellow. The full extent of the 3' UTR was not established and is illustrated up to a predicted Polyadenylation signal (green). All of the splice sites have an intronic GT in the donor site and intronic AG in the acceptor sites (red). The polymorphisms identified within the genes are in bold and surrounded by square parentheses.

(B) Schematic alignment of the intron-exon structure of the gene in human, mouse, wallaby, opossum, platypus, and chicken indicates that the two 3' murine introns are greatly expanded compared to those of the other vertebrates. In eutherians, the Dkl1 gene span is less than 1% of the entire domain (from the start of Dkl1 to the stop codon of Dio3). However, in the noneutherian species analysed, the Dkl1 gene span is greater than 2% of the region.

(C) Sequence traces of genomic DNA and cDNA from heterozygous individuals confirm that the gene is biallelically expressed in all species.

![Found at doi:10.1371/journal.pbio.0060135.s002 (25.5 MB TIF).](image)

**Figure S3.** Wallaby and Platypus Dio3

Dio3 in all vertebrates contains a single exon. The sequence of the Dio3 gene in wallaby and platypus is shown. The coding sequence is highlighted turquoise and is in uppercase. The 3' UTR is highlighted in yellow. The full extent of the 3' UTR is illustrated up to a predicted Polyadenylation signal (green). The polymorphisms identified within the genes are in bold and surrounded by square parentheses.

![Found at doi:10.1371/journal.pbio.0060135.s003 (2.6 MB PDF).](image)

**Figure S4.** Comparative Repeat Content of Regions Showing Expansions/Contractions between Therians

Inter-zone repeat content of regions which show expansions and lack of expansion (contractions) in eutherian and marsupial mammals. LINE elements are subdivided into four families (L1, L2, L3, and RTE) to illustrate the different contributions made to overall LINE content.

![Found at doi:10.1371/journal.pbio.0060135.s004 (739 KB TIF).](image)
Figure S5. Assessment of Noncoding RNA Transcription in Wallaby
(A) Schematic representation of the Dlk1-Dio3 domain indicating locations of ECRs and randomly selected loci assessed for expression. Position of the primer pairs used for the RT-PCR analysis are indicated by black bars below the horizontal line. The positions of primer used to analyse ECR expression are shown in red. Above the horizontal line are ECRs. Turquoise, ECRs conserved in all vertebrates studied; pink, ECRs conserved in all non-eutherians studied; blue, other ECRs. The extent of the expression is indicated by a broken green line.

(B) RT-PCR with primers designed to ECR36 (located within an intron of Migt in mouse). Expression was seen from wallaby fetal head.

(C) A G/T SNP was identified in one wallaby PY sample. Sequence trace data demonstrate biallelic expression of ECR36.

Table S1. Wallaby and Platypus Sequenced Clones
(A) List of sequenced wallaby BACs including accession numbers and finished lengths. (B) List of sequenced platypus BACs including accession numbers and finished lengths.

Table S2. The ECRs Identified in the Dlk1-Dio3 Region in Human, Mouse, Dog, Wallaby, Opossum, Platypus, and Chicken
Each ECR is given a number from 1 to 141. In addition, the 31 vertebrate ECRs are indicated by V1 to V31. The coordinates of each ECR are given for their position in the analysed sequences. ECRs, which are inverted in at least one species are indicated by (–). One ECR (ECR127) is conserved in all non-eutherians studied; pink, ECRs conserved in all non-eutherians studied; blue, other ECRs. The extent of the expression is indicated by a broken green line.

Table S3. Summary Expression Analysis Indicating Absence or Presence of Expression
Found at doi:10.1371/journal.pbio.0060135.s003 (17 KB XLS).

Table S4. List of PCR Primers Used for Expression Analysis and Southern Hybridisation Probes
The PCR conditions used are indicated.

Table S5. List of Wallaby Samples Used in This Analysis
By manipulating the reproductive cycle of the tammar wallaby, it is possible to recover animals at specific stages of development. This is because lactating mothers have a second fertilised embryo arrested in diapause. During the wallaby breeding season, pregnancy can be initiated by the removal of the pouch young (RPY), which reactivates the blastocyst from diapause. Pregnancies are dated relative to the day of pouch young removal and are prefixed with d (e.g., d23). Pouch young are dated from days post parturition and are prefixed by a D (e.g., D10).

References
Evolution of an Imprinted Domain in Mammals
