Marsupials have monoallelic MEST expression with a conserved antisense lncRNA but MEST may not be imprinted.

Marilyn Renfree (m.renfree@unimelb.edu.au)
The University of Melbourne https://orcid.org/0000-0002-4589-0436

Teruhito Ishihara
THe Babraham Institute

Shunsuke Suzuki
Shinshu University

Jane Fenelon
The University of Melbourne

Oliver Griffith
Macquarie University

Geoff Shaw

Article

Keywords:

Posted Date: February 17th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2577479/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

The imprinted isoform of the Mest gene in mice is involved in key mammalian traits such as placental and fetal growth, maternal care and mammary gland maturation. MEST has a distinct promoter differentially methylated region (DMR) in eutherian mammals but in marsupials, while MEST was thought to be imprinted, it had no DMR. In this study, we examined similarities and differences in the MEST gene locus across mammals using a marsupial, the tammar wallaby, a monotreme, the platypus, and a eutherian, the mouse, to investigate how MEST imprinting evolved in mammals. By confirming the presence of the short isoform in all mammalian groups (which is imprinted in eutherians), this study suggests that an alternative promoter for the short isoform has evolved at the MEST gene locus in the common ancestor of mammals. In the tammar, the short isoform of MEST shared the putative promoter CpG island with an antisense IncRNA previously identified in humans and an isoform of a neighbouring gene CEP41. The antisense IncRNA was expressed in tammar sperm, as seen in humans. This suggested that the conserved IncRNA might play an important part in the establishment of MEST imprinting in therian mammals, but it was not imprinted in the tammar. In contrast to previous studies, this study showed that MEST is non-imprinted and mono-allelically expressed in marsupials. This suggests that selection of MEST imprinting in eutherians must have occurred after the marsupial-eutherian split with the acquisition of a key epigenetic imprinting control region, the DMR.

Introduction

Genomic imprinting, the expression of genes of parental origin, occurs in therian mammals (marsupials and eutherians) among vertebrates (Ferguson-Smith, 2011; Renfree et al., 2009). This epigenetic process is essential for normal development in therians as disruption of imprinting causes many developmental abnormalities (Ferguson-Smith, 2011; Lefebvre et al., 1998; Reik and Walter, 2001; Renfree et al., 2009). However, how and why imprinting has evolved in therian mammals is currently unknown. It is certain that genomic imprinting arose very early in mammals, but the absence of imprinting in monotremes suggests that its evolution occurred after the split between monotremes and therian mammals (Pask et al., 2009; Renfree et al., 2009). Therefore, to understand the ancestral condition and the evolution of imprinting in mammals, gene loci that are imprinted in both marsupials and eutherians need to be compared amongst all three mammalian groups, including monotremes.

There are more than 200 imprinted genes in eutherians (Santini et al., 2021; Wang et al., 2011; Wang and Clark, 2014). In marsupials, 36 genes that are imprinted in eutherians have been examined so far (Stringer et al., 2014; Edwards et al., 2019; Ishihara et al., 2022). However, of these, only 9 genes have been confirmed to be imprinted in marsupials (Ager et al., 2007; Das et al., 2012; Ishihara et al., 2022; Killian et al., 2000; O’Neill et al., 2000; Smits et al., 2008; Suzuki et al., 2007, 2005). In addition, there are five marsupial-specific imprinted genes (Douglas et al., 2014; Grant et al., 2012; Mahadevaiah et al., 2020; Stringer et al., 2012; Suzuki et al., 2018). These common imprinted genes could provide insights into the evolutionary conserved and/or lineage specific imprinting mechanisms between marsupials and eutherians. Mesoderm specific transcript (MEST, also known as PEG1) is an important candidate for this
comparison between marsupials and eutherians because it is critical for successful reproduction. MEST is thought to be imprinted in both eutherians and marsupials (Das et al., 2012; Eggermann et al., 2012; Huntriss et al., 2013; Li et al., 2015; Mayer et al., 2000; Reule et al., 1998; Riesewijk et al., 1997; Suzuki et al., 2005). However, the marsupial MEST lacks a DMR (Das et al., 2012; Suzuki et al., 2005) whereas eutherian MEST has a distinct DMR (Li et al., 2015; Riesewijk et al., 1997). This suggests that marsupial MEST imprinting may be regulated by another mechanism such as differential histone modification-based mechanisms. Although its exact function is unknown, disruption of MEST imprinting causes developmental defects in placental and fetal growth and abnormal maternal behaviour in mice (Hiramuki et al., 2015; Lefebvre et al., 1998; Mayer et al., 2000). The MEST imprinted isoform also regulates mammary gland maturation in mice (Yonekura et al., 2019). Since the gene is associated with many classical mammal-specific traits, characterising the evolution of this gene locus across mammals could shed light on how imprinting evolved in the common ancestor of therian mammals.

There is a clear difference between the MEST locus in eutherians and marsupials: the presence or absence of a DMR. This suggests that the regulation of imprinting at the MEST locus may have evolved independently with very different mechanisms or that there may be common factors other than the DMR. Conserved imprinting loci between marsupials and eutherians have been found to associate with IncRNAs that may have been conserved or evolved independently (Smits et al., 2008; Suzuki et al., 2018). In the human MEST gene locus, there is an antisense IncRNA, MESTIT1 (PEG1-AS) (Li et al., 2002; Nakabayashi, 2002). This antisense IncRNA is predominantly expressed in the human testis and mature spermatozoa (Li et al., 2002). If this IncRNA plays an important part in the establishment of MEST imprinting, it is possible that marsupials also have a similar antisense transcript. More detailed comparison of the MEST gene locus across mammals is necessary to elucidate conserved features and to determine how MEST became imprinted during mammalian evolution.

In this study, we examined similarities and differences in the MEST gene locus across mammals. First, we asked whether the shorter isoform, which is imprinted in eutherians, is present in all mammalian groups. Next, we examined whether the antisense IncRNA was present in the tammar wallaby. We also examined the gene structure and genomic element of MEST and its neighbouring gene, centrosomal protein 41 (CEP41) in all three mammalian groups. Here we report that the MEST isoform evolved in the common ancestor of mammals. We confirmed the presence of an antisense IncRNA from the MEST locus in tammar sperm. However, despite the presence of the IncRNA, we found that marsupial MEST was non-imprinted and mono-allelically expressed, whereas in eutherians, the evolution of MEST must have been accompanied by the acquisition of a DMR and imprinting after the marsupial-eutherian split.

**Material And Methods**

**Animals**

Tammar wallaby (Macropus eugenii) samples of Kangaroo Island, South Australia origin, were collected from either wild animals or captive animals from our colony maintained by the University of Melbourne.
Adults were killed humanely and tissues were collected as previously described (Renfree, 1973a, 1973b; Stringer et al., 2012). Adult testes were snap-frozen immediately after dissection. Sperm was collected post-mortem from adult male wallabies after euthanasia. The cauda epididymis was cut into small pieces in sterile saline and incubated in 15 ml tubes for 1h at 37°C to allow the spermatozoa to swim up. The sperm pellets were collected by centrifugation of supernatant and immediately snap-frozen. Both the avascular bilaminar (BOM) and vascular trilaminar (TOM) regions of the choriovitelline placentas were collected from fetuses in the final third of gestation (day 19-25 days of the 26.5 day active pregnancy (n=29)) from adult females post mortem as previously described (Renfree 1973a; Renfree 1973b; Ager et al. 2007; Stringer et al. 2012) and snap-frozen immediately after dissection. Endometrial tissues were also collected from adult females and snap-frozen immediately after dissection. All tammar animal handling and husbandry was in accordance with the National Health and Medical Research Council of Australia (2013) guidelines (National Health and Medical Research Council (Australia), 2013) and approved by the University of Melbourne Animal Experimentation Ethics committees. Platypus (Ornithorhynchus anatinus) tissues were collected from wild-caught animals under permits from NSW Parks and Wildlife and ethically approved by the University of Melbourne Animal Ethics committees. Testes of mouse (Mus musculus) were kindly provided by the Pask laboratory at the University of Melbourne, Melbourne, Australia as a secondary use from ethically approved projects.

**RNA extraction and cDNA synthesis**

Snap-frozen adult testes of each species and the tammar placenta tissues consisting of the separated avascular bilaminar omphalopleure (BOM) and the vascular trilaminar omphalopleure (TOM) were used for RNA extraction using the GenElute Mammalian total RNA Miniprep Kit (Sigma-Aldrich, Missouri, USA) following the manufacturer's instructions. Snap-frozen tammar sperm pellets were used for RNA extraction using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The extracted RNA was treated with the DNA free DNase treatment and removal kit (Thermo Fisher scientific, Massachusetts, USA) to remove residual genomic DNA. Two hundred (200) ng of total RNA was used as a template for cDNA synthesis using SuperScript IV First strand Synthesis System (Invitrogen, Carlsbad, USA).

**Transcriptome analysis**

To identify the potential antisense IncRNA in adult tammar testes, tammar transcriptome data sets derived from adult testis were analysed. Publicly available tammar raw RNA-seq data sets (DRP001145) were downloaded from NCBI SRA (https://www.ncbi.nlm.nih.gov/sra). All RNA-seq reads were trimmed using TrimGalore! (v0.6.5) (https://github.com/FelixKrueger/TrimGalore) with default settings to eliminate adaptor sequences, poor quality reads and very short (<20 bp) reads. The trimmed reads were aligned to the wallaby genome.v3 (https://wallabase.science.unimelb.edu.au) using HISAT2 (v2.1.0) (Kim et al., 2019) with a parameter −rna-strandness FR to reflect the strandedness of sequenced RNA. The mapped reads were assigned to each strand by Samtools (v1.9) (Li et al., 2009). The output file was visualised on Integrative genome viewer (IGV) (Robinson et al., 2011; Thorvaldsdóttir et al., 2013).
5’ and 3’ Rapid amplification of cDNA ends (RACE)

To determine the transcription start site (TSS) of MEST in the tammar and platypus, RACE (rapid amplification of cDNA ends) experiments were performed using the SMARTer RACE 5’/3’ kit (Clontech, California, USA). The TSS of CEP41, a neighbouring gene of MEST, was also confirmed by 5’ RACE reactions in the tammar, platypus and mouse. The transcription start position of a putative antisense novel IncRNA was also confirmed by 5’ RACE reaction in the tammar and mouse. The first round 5’ RACE reaction was performed with adult testis cDNA from each species using SeqAmp DNA Polymerase (Clontech, California, USA) with gene specific primers (Table 1). Up to 1 µg of adult testis RNA was used to synthesise the testis cDNA for RACE reactions by the SMARTer RACE 5’/3’ kit. The nested 5’ RACE was performed by GoTaq DNA polymerase (Promega, Wisconsin, USA) and the RACE products were cloned using pGEM-T Easy Vector (Promega, Wisconsin, USA) and JM109 competent cells or Stellar Competent Cells (Clontech, California, USA). Plasmids were extracted using Wizard Plus SV Minipreps DNA Purification System (Promega, Wisconsin, USA) and sequenced. To complete full sequence of a putative antisense novel IncRNA in the tammar, 3’ RACE experiments were performed. The first round RACE reaction was performed with adult testis cDNA or placenta tissue (BOM) cDNA using SeqAmp DNA Polymerase (Clontech, California, USA) with gene specific primers (Supplementary Table 1). The nested 3’RACE was performed by GoTaq DNA polymerase (Promega, Wisconsin, USA) and the RACE products were cloned using pGEM-T Easy Vector (Promega, Wisconsin, USA) and JM109 competent cells. Plasmids were extracted using Wizard Plus SV Minipreps DNA Purification System (Promega, Wisconsin, USA) and sequenced by the Sanger sequencing method with M13 primers (Supplementary Table 1).

Identification and expression analysis of the antisense IncRNA MESTIT1 in tammar sperm

Tammar MESTIT1 specific primers were designed based on the sequences determined by RACE reactions using Primer3Plus (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (Supplementary Table 1). PCR reactions were performed using adult tammar testis cDNA with GoTaq Green master mix (Promega Corporation, USA) to confirm PCR amplification with the MESTIT1 specific primers and sequencing. To determine whether the IncRNA was present in the tammar sperm, PCR using sperm cDNA and DNase treated sperm RNA was used (n=3).

Comparative analysis of mammalian MEST and CEP41 gene locus.

DNA sequences of human MESTA, human MESTB, mouse Mesta, mouse Mestb, mouse Cep41, platypus putative MEST and platypus putative CEP41 were obtained from NCBI (https://www.ncbi.nlm.nih.gov). Amino acid sequences retrieved from DDBJ/EMBL/GenBank/RefSeq database were used for generating alignments using CLC sequence viewer 8 (https://clc-sequence-viewer.software.informer.com). Accession numbers: Homo sapiens MESTA, NM_177525.2; H. sapiens MESTB, NM_177524.2; M. musculus Mesta, NM_008590.2; M. musculus Mestb, NM_001252293.1; M. musculus Cep41: NM_031998.3; O. anatinus putative MEST: XM_001511283.6; O. anatinus putative CEP41: XM_001511256.6. Genomic sequences of the MEST-CEP41 flanking region for each mammalian species (human, mouse, African savanna elephant (Loxodonta africana), tammar and platypus) were obtained from either NCBI
CpG island predictions were performed by MethPrimer (http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi) (Li and Dahiya, 2002). Graphs of CpG dinucleotide ratio were made by GENETYX software (https://www.genetyx.co.jp). Genomic elements within the intergenic region between mammalian MEST and CEP41 genes were investigated by BLAST searches (https://blast.ncbi.nlm.nih.gov/Blast.cgi).

**Genomic DNA extraction**

Snap-frozen BOM, TOM, and endometrial tissues were used for genomic DNA (gDNA) extraction. BOM and TOM were the fetal gDNA source and endometrium was the maternal gDNA source. DNA extraction was performed with Wizard Genomic DNA purification kit (Promega, Wisconsin, USA) following the manufacturer's instructions.

**Allelic Expression analysis**

Extracted genomic DNAs were used as a template for PCR amplification to find a combination of maternal homozygous and fetal heterozygous single nucleotide polymorphisms (SNPs). PCR reactions were performed using gene specific primers (Supplementary Table S1) with Ex-Taq polymerase (Takara, Shiga, Japan) or Go-Taq polymerase (Promega, Wisconsin, USA) under the following cycle conditions: 95°C 30s, 65°C 30s, and 72°C 45s. To analyse sequences of the transcript, cDNA of placenta tissues (BOM and TOM) was used as a template for PCR amplification under the following cycle conditions: 95°C 30s, 65°C 30s, and 72°C 45s. After performing gel electrophoresis, confirmed PCR products from gDNA and cDNA were extracted and directly sequenced by Sanger sequencing to confirm SNP sites and allele specific expression.

**Results**

**Re-evaluating the transcription start site of tammar MEST.**

Although MEST imprinting has been characterised in both the South American grey short-tailed opossum (Monodelphis domestica) (Das et al., 2012) and the tammar (Suzuki et al., 2005), the transcription start sites (TSSs) of the MEST isoforms are not yet well defined in marsupials. To characterise the tammar MEST gene locus, the putative tammar MEST gene was searched using the wallaby genome database (Wallabase: https://wallabase.science.unimelb.edu.au/) and compared with mouse Mest protein (Accession number: NP_032616.1). 2859 bp of putative tammar MEST was identified (Figure 1A). Since the MEST gene has isoform dependent imprinted expression in eutherians (Kosaki et al., 2000; Li et al., 2015; Reule et al., 1998; Riesewijk et al., 1997; Yonekura et al., 2019), we examined isoforms of MEST by 5’RACE experiments using the adult tammar testis. As MESTA shares the protein coding sequence with the longer isoform of MEST, MESTB in eutherians (Figure 1A: green coloured boxes), 5’RACE reaction was performed using primers designed for the protein coding sequences of the putative tammar MEST (Figure 1A: green coloured boxes). After sequencing the 5’ RACE products, we confirmed that there are three isoforms of MEST (Figure 1B and C). The longer isoform of MEST was expressed from a more
upstream CpG island that was distinct from the other two isoforms (Figure 1C). We renamed the shortest isoform of MEST as MESTA (DDBJ accession number: LC747011) and the longer isoform of MEST as MESTB (DDBJ accession number: LC747012) in accordance with eutherian MEST isoforms. The intermediate size isoform was renamed MESTC. Although each isoform had a different TSS, the translation start sites were common to each other.

**Identification of two isoforms of MEST in monotremes.**

To ask whether the short isoform of MEST is therian mammal specific, the presence of MEST isoforms in monotremes was examined by 5′RACE experiments using adult platypus testis. To characterise the monotreme MEST gene locus, the platypus orthologue of MEST was searched by NCBI Blast (https://blast.ncbi.nlm.nih.gov/Blast.cgi) comparing with the tammar MEST amino acid sequences, and 2374 bp of putative platypus MEST (Accession number: XM_001511283.6) was identified (Figure 2A). Next, 5′RACE reactions were performed using primers designed for the protein coding region of the putative platypus MEST (Figure 2A and B: green coloured boxes). After sequencing the RACE products, we confirmed that platypus has two isoforms of MEST (Figure 2B and C). Their TSSs were located at different CpG islands from each other (Figure 2C). We renamed the shortest isoform of MEST as MESTA (DDBJ accession number: LC747014) and the longer isoform of MEST as MESTB (DDBJ accession number: LC747015) in accordance with eutherian and marsupial MEST isoforms. Although each isoform had a different TSS, the translation start sites were common to each other.

**Identification of an orthologue of MESTIT1 in the tammar wallaby.**

Since the human IncRNA MESTIT1 is expressed from the DMR at the promoter of human MESTA (Li et al., 2002; Nakabayashi, 2002), we asked whether similar antisense transcripts are present around the TSS of the tammar MESTA by analysing stranded transcriptome data sets. In adult tammar testis transcriptome data, antisense mapping reads were present in the CpG island near the TSS of the tammar MESTA (Figure 3A). After isolating a partial transcript by PCR with an antisense transcript candidate-specific primer, 5′ and 3′RACE experiments were performed to confirm the full length of the transcript (Figure 3B). While the 5′ RACE reaction yielded one band, the 3′RACE reaction resulted in three different bands. Of these, two 3′RACE products were non-coding transcripts with an alternative poly A signal and distinct poly A tail (DDBJ accession numbers; MESTIT1 isoform1: LC746974; MESTIT1 isoform 2: LC746975) (Figure 3B: black asterisks). Surprisingly, the largest 3′RACE product encoded an isoform of the neighbouring gene of MEST, CEP41 (Later renamed CEP41A, DDBJ accession number: LC747013) (Figure 3B: red asterisks). The neighbouring genes of the putative MESTIT1 orthologue in the tammar showed synteny to corresponding genes in human genome, suggesting that the MESTIT1 gene is conserved between the tammar and the human genomes.

To ask whether the IncRNA MESTIT1 is present in the tammar sperm as seen in human, RT-PCR analysis was performed. PCR amplification was observed in all sperm samples only after reverse-transcription (Figure 3C).
Tammar has two isoforms of CEP41.

During the RACE reaction to identify MESTIT1 in tammar, an isoform of CEP41 was identified. To determine whether isoforms of CEP41 other than the isoform identified by the RACE experiment of MESTIT1, the presence of CEP41 isoforms was examined using the tammar wallaby. In the wallaby genome database (Wallabase: https://wallabase.science.unimelb.edu.au), there was a putative CEP41, but its exon structure was not the same as the CEP41 identified in this study (Figure 4A). Therefore, we renamed the isoform identified by 3’RACE as CEP41A (Figure 4A). The putative CEP41 and CEP41A shared several exons. To confirm the presence of other isoforms of CEP41, 5’ RACE reactions were performed using primers designed for the common exons of CEP41A and the putative CEP41 (Figure 4B: green coloured boxes). After sequencing the 5’RACE products, two TSSs were identified for the tammar wallaby CEP41. CEP41A shared a CpG island with marsupial MESTA. The other isoform was found to share a CpG island with MESTB. Because of the different exon structures, the newly identified CEP41 isoform was named CEP41B. Furthermore, we confirmed that the possible protein encoded by CEP41B differs in its C-terminal region from the amino acid sequence encoded by CEP41A (Figure 4C).

CEP41A isoform is not present in either mouse or platypus

To ask whether the CEP41 isoform, CEP41A, is a marsupial specific isoform, the presence of CEP41 isoforms in mouse and platypus was examined by 5’RACE experiments using their adult testes (Figure 5). First, the sequence of the mouse CEP41 gene was obtained from NCBI (Accession number: NM_031998.3). Since CEP41A shared several protein-coding exons with CEP41B in the tammar, 5’ RACE reactions were performed using primers against the conserved region of mammalian CEP41 (Figure 5A: green coloured boxes). A single transcript was isolated from adult mouse testis. This transcript was identical to the known mouse CEP41 and had a genetic structure similar to the tammar CEP41B (Figure 5A). Similar experiments were performed with adult platypus testes after obtaining a putative platypus CEP41 gene from NCBI (Accession number: XM_001511256.6). A single transcript was isolated from an adult platypus testis. The transcripts also had high homology to the tammar marsupial CEP41B (Figure 5B).

Genomic analysis of the MEST and CEP41 flanking regions

To investigate CpG island locations and CEP41 and MEST isoforms in mammals, the genomic structures of the MEST and CEP41 flanking region in human, tammar and platypus were compared with each other. In both tammar and platypus, two large domains of CpG islands exist in close proximity (Figure 6A: yellow highlighted regions). The two major MEST isoforms, MESTA and MESTB, were expressed from the two large CpG island domains, respectively (Figure 6A). However, there is expansion of the intergenic region between CEP41 and MEST in the human genome, and the human MESTB did not share the same CpG island with human CEP41 (Figure 6A). NCBI BLAST searches of the human MEST and CEP41 flanking region identified a LINE1 ORF1 in the vicinity of MEST. Similar LINE1 elements were also found in the mouse and elephant genomes by NCBI BLAST searches (Figure 6B). However, the elephant LINE1 element was located close to CEP41 (Figure 6B).
**Neighbouring transcripts of MESTA are not imprinted in the tammar placenta tissues**

Since human *MESTIT1* is imprinted (Nakabayashi, 2002), it was possible that the tammar *MESTIT1* is also imprinted. To confirm whether *MESTIT1* is imprinted in the tammar wallaby, allelic expression of the gene was performed using tammar placenta tissues. First, SNP sites were examined by direct sequencing of genomic DNA with RT-PCR (**Figure 7A: black arrows**). After examining 29 samples, one SNP site was identified in the common region of the two isoforms (**Figure 7A**). In the shorter isoform, specific primers could not be designed because the SNP site was too close to the poly-A tail. However, for the longer isoform, we could detect it from cDNA using the same primers as used for the SNP search (**Figure 7A**). Using these primers, the imprinting status of the tammar lncRNA was determined by direct sequencing of the PCR products that contained the SNP site (**Figure 7A**). 29 samples were examined, and three samples were a combination of maternal homozygous and fetal heterozygous SNPs (**Figure 7B**). In contrast to the gDNA PCR data in which the two peaks at the SNP site have almost the same signal strength, the cDNA PCR products showed that the lncRNA is preferentially expressed from either one of the two alleles (**Figure 7B**). Individuals #1 and #2 showed paternally skewed expression, but individual #3 showed maternally skewed expression (**Figure 7B**). Therefore, in the tammar placenta, the lncRNA was not imprinted.

Although *CEP41* is not imprinted in mouse (Yamada et al., 2002), the tammar *CEP41A* shares the same CpG island with the shorter *MEST* isoform, *MESTA*. It was therefore possible that marsupial *CEP41A* is imprinted with *MESTA*. To confirm whether *CEP41A* is imprinted in the tammar wallaby, allelic expression of the gene was examined in the tammar placenta tissues. Since each isoform has a unique TSS and unique exons, allelic expression analysis was performed using *CEP41A* specific primers (**Figure 7C and D**). First, SNP sites in 3'UTR were examined by direct sequencing of genomic DNA with PCR (**Figure 7C: blue arrows**). After examining 18 samples, two SNP sites were identified in the 3'UTR of the tammar *CEP41A* (**Figure 7D: Arrowheads**). Allelic expression was performed using *CEP41A* specific primers (**Figure 7D**). Of these 18 samples, two animals were a combination of maternal homozygous and fetal heterozygous. Fortunately, the maternal homozygous SNPs in these two samples were different from each other (**Figure 7E**). In animal #1, *CEP41A* was preferentially expressed from the maternal allele (C) in the BOM and TOM tissues. However, in animal #2, *CEP41A* was preferentially expressed from the paternal (C) allele in the TOM tissues (**Figure 7E**). *CEP41A* was not detectable in the BOM of animal #2 so we could not determine its allelic expression (**Figure 7E**). In animal #3, *CEP41A* showed a clear bi-allelic expression in the BOM placenta and skewed expression in the TOM placenta (**Figure 7E**). Therefore, in the tammar placenta, *CEP41A* was not imprinted.

**Re-evaluating allelic expression of MEST in the tammar placenta tissues**

Since *CEP41A* and *MESTIT1* were not imprinted in the tammar placenta even though they shared the same CpG island with the *MEST* gene, we re-evaluated *MEST* imprinting in the tammar placenta. The two major isoforms identified in this study were expressed from the two different CpG islands, respectively (**Figure 8A**). Since each isoform has a unique TSS and unique exons, allelic expression analysis was
performed using each isoform specific primer and the shared reverse primer (Figure 8B). Our RACE experiments could not identify the isoform previously described (Suzuki et al., 2005), but using the same primers as Suzuki et al. (2005) we re-examined its allelic expression (Figure 8A and B). The reverse primer was designed to detect the previously described C/A SNP site at the 3'UTR region (Suzuki et al., 2005) (Figure 8B: Arrowhead). Seventeen samples were examined, and six animals had a heterozygous SNP at the SNP site. All of the six samples showed monoallelic expression (Figure 8C). The mothers of four of the six animals were not homozygous, but fortunately, the maternal homozygous SNPs in the remaining two samples were different from each other (Figure 8C). In animal #1, all isoforms were expressed from the paternal allele (A), exclusively. However, in animal #2, all isoforms were expressed from the maternal (A) allele (Figure 8C). Therefore, all isoforms in the two animals were not imprinted but all were mono-allelically expressed in the tammar TOM tissues (n=6).

Discussion

By characterising MEST isoforms in two of the three mammalian lineages, the marsupials and monotremes, this study confirmed the presence of the short isoform in all mammalian groups. We identified a conserved antisense IncRNA, MESTIT1, in the tammar. This antisense transcript was present in the tammar sperm as seen in human (Li et al., 2002). Further comparison of the MEST and CEP41 gene loci amongst mammals showed that there is a marsupial specific CEP41 isoform sharing a CpG island with the shorter isoform of MEST. The MEST-CEP41 flanking region in eutherians acquired a DMR and a retrotransposon insertion. In contrast to previous studies which suggested that MEST was imprinted in marsupials (Das et al., 2012; Suzuki et al., 2005), while the MEST isoforms examined in this study were mono-allelically expressed, they were not imprinted in the tammar placenta. Although MEST, CEP41A, MESTIT1 shared the same CpG island, only MEST showed the apparent monoallelic expression in the placenta. This study suggests that monoallelic expression of MEST is conserved in therian mammals but is imprinted only in eutherian mammals and with the acquisition of a promoter DMR.

The short isoform of MEST was confirmed in both marsupials and platypus. Since this short isoform associates with mammary gland maturation in mice (Yonekura et al., 2019), it is possible that the evolution of the MEST isoform occurred with the evolution of the mammary gland in the common ancestor of mammals. However, the exact function of the MEST short isoform is currently unknown in any mammal. To confirm its conserved role in mammary gland maturation, it would be interesting to characterise transcript localisation of the isoform in developing marsupial or monotreme mammary glands in subsequent studies.

Detailed comparisons of the flanking regions of MEST-CEP41 have provided insight into the evolution of this locus in mammalian evolution. For example, there was a conserved antisense IncRNA, MESTIT1, in the tammar. This antisense IncRNA is expressed from the known DMR in human and the CpG island where the MESTA transcription occurs in the tammar. These data suggest that the CG rich region has conserved bidirectional promoter activity in therian mammals. Importantly, the tammar MESTIT1 shares the same CpG island and transcriptional orientation with a marsupial specific CEP41, CEP41A. However,
in eutherians, there was a retrotransposon insertion between CEP41 and MEST so that this insertion physically separated the two genes in eutherians. Our RACE experiment could not detect any CEP41A-like transcript in mice. Even in the platypus, despite the physical proximity of MEST and CEP41, we were unable to identify a transcript similar to the tammar CEP41A by the 5’ RACE reaction. Thus, the bidirectional promoter activity at the CG rich region where MESTA transcription occurs must have evolved in the common ancestor of therian mammals.

Previously, allelic expression of tammar MEST was examined without reciprocal SNP combinations in homozygous mothers (Suzuki et al., 2005). In the present study, we used a reciprocal SNP combination in homozygous mothers and analysed allelic expression of the tammar MEST in placenta. This showed that the tammar MEST is a non-imprinted mono-allelically expressed gene in the placenta. In the opossum study, the MEST transcript was mono-allelically expressed but there was no parental origin information (Das et al., 2012) but this is consistent with our data to suggest that MEST is not imprinted in marsupials. It is currently unclear whether the mono-allelic expression of MEST is caused by random mono-allelic expression or by a non-random mechanism. This needs to be clarified in subsequent studies.

Imprinted expression can only evolve when gene expression levels affect inclusive fitness in the ancestor. It is thought that imprinted genes might have evolved from genes with pre-existing gene dosage sensitivity which led to the feature of parental origin specific expression (Haig, 2000; Patten et al., 2014). In this way, marsupials and the MEST gene locus may offer an opportunity to study the key steps involved in how imprinting evolved from a normal bi-allelic expressed gene to a dosage sensitive/mono-allelically expressed gene to an imprinted gene. More detailed analysis of the expression mechanism and function of MEST is needed to clarify the differences between genes that are imprinted and those that are not. Further characterisation of monotreme MEST allelic expression is also necessary to confirm whether this gene is mono-allelically expressed or not, but obtaining maternal-young material in monotremes is extremely difficult.

Our data shows that the MEST gene locus evolved differently in both eutherians and marsupials after the monotreme-therian split. This study demonstrates that monoallelic expression of the MEST gene occurs in marsupials without imprinting. In contrast, selection of imprinting of MEST in eutherians must have occurred after the marsupial-eutherian split with the acquisition of a DMR and imprinting.

**Declarations**

**Data availability**


**Acknowledgements**
We thank Corinne van den Hoek for assistance with the animals and all members of the wallaby research group for help with the tammar wallabies and the dissections in Melbourne. We thank Professor Peter Temple-Smith and Dr Kath Handasyde for help with the collection of platypus material. Confocal microscopy was done with the support and infrastructure of Melbourne Advanced Microscopy. TI acknowledges support provided by a writing up award from the Albert Shimmins fund.

Authors contributions: TI, SS, JCF, GS, OWG and MBR discussed and designed the research; TI performed most of the research; JCF, GS and MBR collected platypus tissues; MBR, GS, TI collected the tammar tissues; JCF extracted RNA from platypus tissue; SS, OWG and MBR provided supervision; TI, SS, JCF, OWG, GS and MBR discussed the data and TI, SS, and MBR wrote the paper. All authors approved the final manuscript.

Conflict of Interest: The authors have no conflict of interest to declare

References


23. National Health and Medical Research Council (Australia), 2013. Australian code for the care and use of animals for scientific purposes.


the H19 noncoding RNA and H19-IGF2 imprinting mechanism in therians. Nat Genet 40, 971–976. https://doi.org/10.1038/ng.168


Tables

Table 1 and Supplementary Table 1 are not available with this version.

Figures
Identification of \textit{MEST} isoforms in the tammar wallaby. \textbf{A.} Exon structures of eutherian \textit{MEST}s and putative tammar \textit{MEST}. In both human and mouse, the protein coding region was common between the two isoforms, \textit{MESTA} and \textit{MESTB}. The highly conserved protein coding regions are represented by green coloured boxes. White and grey coloured boxes represent protein-coding exon and UTRs, respectively. 

\textbf{B.} Identification of the TSS of tammar \textit{MEST} isoforms. 5'RACE and nested RACE primers are represented by the black arrows. Asterisks represent RACE products encoding partial tammar \textit{MEST} sequences. Boxes represent exons identified by sequencing of the RACE products. Black boxes represent CpG islands determined by the Methprimer programme.
Identification of \textit{MEST} isoforms in platypus. \textbf{A.} Exon structures of eutherian \textit{MEST}s, the tammar \textit{MEST}s and putative platypus \textit{MEST}. In human, mouse and the tammar wallaby, the protein coding region was common between the two isoforms, \textit{MESTA} and \textit{MESTB}. The highly conserved protein coding regions are represented by green coloured boxes. White and grey coloured boxes represent protein-coding exon and UTRs, respectively. \textbf{B.} Identification of the TSS of the platypus \textit{MEST} isoforms. 5’RACE and nested RACE.
primers are represented by the black arrows. Asterisks represent RACE products encoding partial platypus MEST sequences. Boxes represent exons identified by sequencing of the RACE products. Black boxes represent CpG islands determined by the Methprimer programme.
Identification of the marsupial IncRNA in adult testis and mature sperm. A. Identification of an antisense transcript at the TSS of the tammar MESTA with transcriptome data. Tammar testis transcriptome data was visualised with IGV in a strand specific manner. There was an antisense transcript candidate at the TSS of MESTA in the tammar testis. B. Characterisation of the full-length sequence of the antisense IncRNA by RACE reactions. 5’ and 3’ RACE and nested RACE were performed using adult testis with primers for the putative antisense transcript. 5’RACE generated a single amplicon whereas 3’RACE reactions resulted in three different products. The two 3’RACE products (black asterisks) encoded a non-coding transcript with a polyadenylation signal and poly-A tail. The largest RACE product encoded an isoform of CEP41 with a protein coding sequence (909bp). C. RT-PCR analysis of the marsupial IncRNA in sperm. Sperm samples were collected from three adult males, #1, #2, and #3, and total RNAs were purified from them. cDNAs were synthesized with (RT+) or without reverse transcription (RT−) using the oligo(dT) primer, and PCR was performed with primer pairs for the tammar IncRNA. Presence of the transcript in sperm was confirmed in all biological replicates. Abbreviations; M: Molecular marker; P: Positive control (adult testis); WT: Water.
Identification of two isoforms of **CEP41** in the tammar wallaby. **A.** Exon structures of **CEP41A**, which was identified by our 3’RACE experiment, and a putative **CEP41** gene obtained from the wallaby genome database (Wallabase). These two isoforms shared some exons but had different gene structures. **B.** Identification of TSS of the tammar **CEP41** isoforms. 5’RACE and nested RACE primers are represented by the black arrows. Asterisks represent race products coding partial tammar **CEP41** sequences. The longest
The isoform had the same sequence with the *CEP41A*. The smallest isoform encoded a different transcript with two additional exons. Since the two transcripts had a different exon structure from each other, we renamed the newly identified transcript as *CEP41B*. **C.** Protein alignment of the two CEP41 isoforms, CEP41A and CEP41B. The putative CEP41B protein had additional amino acids in its C-terminal region.

**A**. Identification of the TSS of mouse *CEP41* isoforms. In the tammar, most of the protein coding exon was common between the two isoforms, *CEP41A* and *CEP41B*. **B**. Protein alignment of the two CEP41 isoforms, CEP41A and CEP41B. The putative CEP41B protein had additional amino acids in its C-terminal region.

**Figure 5**

*CEP41A* was a marsupial specific isoform. **A**. Identification of the TSS of mouse *CEP41* isoforms. In the tammar, most of the protein coding exon was common between the two isoforms, *CEP41A* and *CEP41B*. **B**. Protein alignment of the two CEP41 isoforms, CEP41A and CEP41B. The putative CEP41B protein had additional amino acids in its C-terminal region.
The highly conserved protein coding region is represented by green coloured boxes. 5'RACE and nested RACE primers are represented by the black arrows. Asterisks represent RACE products coding partial mouse \textit{CEP41} sequences. The RACE product had the similar genetic structure with the tammar \textit{CEP41B}. 

B. Identification of TSS of platypus \textit{CEP41} isoforms. The highly conserved protein coding region is represented by green coloured boxes. 5’RACE and nested RACE primers are represented by the black arrows. Asterisks represent RACE products coding partial platypus \textit{CEP41} sequences. The RACE product had a similar genetic structure with tammar \textit{CEP41B}. 

---

**Diagram**

\textbf{A. Human}

\textbf{Tammar}

\textbf{Platypus}

\textbf{B.} Diagram showing LINE1 relocation and insertion.
**Figure 6**

The evolution of *MEST-CEP41 flanking region in mammals*. **A.** Comparison of the genomic structures between *MEST* and *CEP41*. The blue graphs show CpG contents in the genomic sequences and the yellow highlight represent CpG islands predicted by the Methprimer programme. **B.** The timing of the LINE1 element emergence during mammalian evolution. Red boxes represent the location of the LINE1 element. In eutherians such as elephant, human and mouse, *MEST* and *CEP41* are physically separated by the LINE1 element. However the location of the LINE1 element in mouse and human is completely different from that of the elephant. In both marsupials and monotremes, *CEP41* and *MEST* are not physically separated.
Figure 7

Allelic expression analysis of the tammar MESTIT1 and CEP41A. A. Isoforms of MESTIT1 and primers for allelic expression analysis. A SNP site was found in the common region of the two isoforms (arrowhead). In the shorter isoform, specific primers could not be designed because the SNP site was close to the poly-A tail. However, primers used for the SNP search could detect the longer isoform. B. Allelic expression of the longer isoform of MESTIT1 in the tammar placentas. The imprinting status of the tammar MESTIT1
was determined by direct sequencing of PCR products that contained a SNP site. In contrast to the genomic DNA data in which double peaks at the SNP site have almost the same signal strength, the cDNA data clearly showed that the IncRNA is predominantly expressed from either one of the two alleles. 

**C.** Exon structures of the two CEP41 isoforms. Orange and aqua coloured boxes represent the CEP41A-specific and CEP41B-specific exons, respectively. Blue arrows represent primers used for detecting the genomic SNP site.  

**D.** Primer design for the allelic expression analysis of the CEP41A transcript. Allelic expression analysis of CEP41A was performed using CEP41A specific forward primer and a reverse primer designed for the 3' UTR common amongst CEP41 isoforms. Arrowheads represent SNP sites.  

**E.** Allelic expression of the tammar CEP41A by direct sequencing with PCR amplification. In animal #1, skewed maternal expression was observed in the placenta tissues. However, in animal #2, skewed paternal expression was observed in the TOM tissue. Due to low expression levels, it was not possible to determine allelic expression of CEP41A in the BOM of Animal #2. In animal #3, CEP41A was bi-allelically expressed in the BOM but skewed expression was observed in the TOM.
Figure 8

Allelic expression of the MEST isoforms in the tammar placenta. A. Exon structures of the two MEST isoforms identified in this study and the previously described isoform. Orange, aqua and green coloured boxes represent MESTA-specific, MESTB-specific and the previously described isoform specific exons, respectively. B. Primer design for the allelic expression of the MEST isoforms. The previously confirmed SNP site (Suzuki et al., 2005) was used for this study. An arrowhead represents the SNP site. Allelic expression analysis of each isoform was performed using each isoform specific forward primer.
and a common reverse primer. C. Allelic expression of tammar MEST isoforms by direct sequencing with PCR amplification. In animal #1, clear paternal expression was observed in the TOM tissue. However, in animal #2, all MEST isoforms were expressed mono-allelically from the maternal genome in TOM tissue. MEST was not sufficiently highly expressed in BOM tissue to determine its allelic expression.