Regulation of alternative polyadenylation by genomic imprinting

Citation for published version:

Digital Object Identifier (DOI):
10.1101/gad.473408

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Genes & Development

Publisher Rights Statement:
Copyright © 2008, Cold Spring Harbor Laboratory Press

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Regulation of alternative polyadenylation by genomic imprinting

Andrew J. Wood, Reiner Schulz, Kathryn Woodfine, Katarzyna Kolotowska, Colin V. Beechey, Jo Peters, Deborah Bourc’his, and Rebecca J. Oakley

1Department of Medical and Molecular Genetics, King’s College London, Guy’s Hospital, London SE1 9RT, United Kingdom; 2Medical Research Council Mammalian Genetics Unit, Harwell OX11 0RD, United Kingdom; 3Institut National de la Santé et de la Recherche Médicale (INSERM) U741, Institut Jacques Monod, 75251 Paris, Cedex 05, France

Maternally and paternally derived alleles can utilize different promoters, but allele-specific differences in cotranscriptional processes have not been reported. We show that alternative polyadenylation sites at a novel murine imprinted gene \( H13 \) are utilized in an allele-specific manner. A differentially methylated CpG island separates polyA sites utilized on maternal and paternal alleles, and contains an internal promoter. Two genetic systems show that alleles lacking methylation generate truncated \( H13 \) transcripts that undergo internal polyadenylation. On methylated alleles, the internal promoter is inactive and elongation proceeds to downstream polyadenylation sites. This demonstrates that epigenetic modifications can influence utilization of alternative polyadenylation sites.

Received January 28, 2008; revised version accepted March 10, 2008.

Transcription by RNA polymerase II (Pol II) requires multiple linked steps, including the assembly of an initiation complex, promoter release, elongation followed by splicing, polyadenylation [polyA], and dissociation of the polymerase complex from the template DNA. Each step provides an opportunity for the cell to regulate gene expression, either by changing the abundance or availability of mRNAs or by generating variant gene products. Epigenetic modifications to DNA and histones can influence transcription at the initiation stage by altering promoter sequences or initiating complex components [Kass et al. 1997a,b]. Following initiation, regions of heterochromatin can subtly impede the progress of an elongating polymerase complex, repressing transcriptional output [Lorincz et al. 2004]. However, it is not clear whether epigenetic modifications downstream from a promoter can lead to variant gene products through alternative polyadenylation or splicing.

Polyadenylation, the addition of multiple adenyl residues to the 3’ end of a newly synthesized transcript, confers stability and is required for nuclear export [Huang and Carmichael 1996; Jacobson and Peltz 1996]. This occurs following cleavage of the newly synthesized pre-mRNA, typically 15–30 nucleotides downstream from a conserved hexamer motif [usually AAUAAA or AUUAAA] [Proudfoot and Brownlee 1976; Zhao et al. 1999]. A large proportion of human genes utilize more than one polyadenylation [polyA] site [Tian et al. 2005], making alternative polyadenylation a major source of transcriptional diversity. The mechanisms governing alternative polyA site selection in mammalian cells have been studied in detail at only a small number of loci; notably the immunoglobulin heavy chain [Takagaki and Manley 1998] and calcitonin/\textit{CGRP} genes [Lou et al. 1998]. In both cases, polyA site selection is cell type-specific and is regulated by changes in the concentration of diffusible RNA processing factors [Lou et al. 1998; Takagaki and Manley 1998].

Imprinted genes are differentially expressed on maternally and paternally derived alleles and are estimated to comprise 1%–2% of the mammalian transcriptome [Barlow 1995; Wood and Oakley 2006]. These genes are typically clustered into regions of the genome containing multiple imprinted transcripts, which share common \textit{cis}-regulatory elements. Such elements undergo differential CpG methylation when passing through the male and female germline. Allele-specific markings are maintained during embryonic development and can influence transcription of adjacent genes via \textit{cis}-acting regulatory RNAs such as \textit{Air} [Sleutels et al. 2002] or insulator factors such as CTCF [Hark et al. 2000].

Imprinted loci represent an attractive model for studying epigenetic mechanisms of transcriptional control, as both active and silent alleles are present in the same cellular environment. In the presence of an identical complement of \textit{trans}-regulatory factors, allelic differences in transcription must be the result of epigenetic regulation. While the allele-specific use of alternative promoters has been well documented at imprinted loci [Blagitko et al. 2000; Holmes et al. 2003], allele-specific differences in cotranscriptional events such as splicing and polyadenylation have not been reported previously.

We undertook a computational screen for imprinted genes, identifying a novel imprinted domain on mouse Chromosome 2 containing at least two protein-coding genes: \textit{H13} and \textit{Mcts2} [Wood et al. 2007]. \textit{Mcts2} is a duplicate of the X-linked \textit{Mcts1} gene that originated by retrotransposition of an mRNA from the X-linked locus into the fourth intron of \textit{H13} [Wood et al. 2007]. Unlike most retrotransposed gene copies, \textit{Mcts2} has maintained the capacity to encode a protein and shares 94% amino acid identity with the X-linked paralog >65 million years after the gene duplication event. The promoter of \textit{Mcts2} is associated with a CpG island that acquires DNA methylation in the female germline, and the retrogene is expressed primarily from paternally derived alleles in the developing embryo [Wood et al. 2007].

We identify five polyA sites for \textit{H13} and show that at least three are utilized in an allele-specific manner. This is likely to depend on allele-specific DNA methylation...
at the internal Mcts2 promoter, which is established during maternal gametogenesis and serves to promote functional H13 products from maternally derived alleles in the embryo. These findings provide strong evidence that epigenetic modifications can influence alternative polyadenylation, and pave the way for future studies to determine precisely how this influence can be exerted.

Results and Discussion

Alternative polyadenylation at the H13 locus

H13 encodes signal peptide peptidase, an intramembrane aspartic protease responsible for the dislocation of signal peptides from the endoplasmic reticulum membrane (Weihofen et al. 2002). The ORF is spread over 12 exons (Fig. 1A) and encodes a protein consisting of seven transmembrane helices (Fig. 1B). EST evidence suggests a number of transcript variants. Northern blots (Fig. 1C) and 5′ and 3′ RACE were performed in material extracted from whole brain tissues of neonatal mice, revealing one promoter and five distinct polyA sites (Fig. 1A). Alternative H13 promoters cannot be absolutely excluded, all transcript species identified by Northern analysis (Fig. 1C) can be accounted for by different 3′ RACE products. Three polyA sites are clustered within a single block of 3′ untranslated region (UTR) situated 20 kb downstream from the monoclonal Mcts2 gene [H13a–c]. The alternatively skipped penultimate exon of the 2-kb H13a transcript contains a premature stop codon that removes a C-terminal KKXX membrane retention signal from full-length signal peptide peptidase proteins, altering subcellular localization patterns [Ury et al. 2006]. Two internal polyA sites lie 0.5 kb (H13d) and 3 kb (H13e) upstream of the Mcts2 transcriptional start site (Fig. 1A). Biochemical characterization of the protein product of H13 has previously demonstrated that the catalytic residues essential for signal peptide peptidase activity lie within the region of the ORF downstream from the H13d and H13e polyA sites [Weihofen et al. 2002]. Hence, only transcripts extending beyond Mcts2 and the CpG island [H13a–c] can encode functional signal peptide peptidase proteins (Fig. 1B).

![Figure 1](image-url)

**Figure 1.** Gene structure and CpG methylation at the murine H13/Mcts2 locus. |A| Transcript map of H13 isoforms, determined by 5′ and 3′ RACE, Northern, and GenBank EST evidence. The bar indicates nucleotide position on mouse Chromosome 2 (Mm. build 37). Representative GenBank accession numbers are as follows: H13a 1.8 kb, AJ345032; H13b 2.0 kb, AK049101; H13b, NM_010376; H13c, F830002G13; H13d, 4021402J09; H13e, AK004690; Mcts2, NM_025543. Exons are black rectangles and splice patterns are indicated. Vertical arrows indicate polyA sites, and horizontal arrows show the direction of transcription. The greater abundance of H13a prevented the simultaneous detection of H13a, H13b, and H13c in a single 3′ RACE reaction. Consequently, we did not determine whether the alternatively skipped penultimate exon is also included in H13b or H13c transcripts. The positions of Northern probes for C are shown at the bottom, as are the three CpG islands at which methylation was investigated in D. |B| Predicted structure of proteins translated from H13 transcript isoforms. Cylinders represent transmembrane (TM) helices. Residues critical for signal peptide peptidase activity are indicated within transmembrane helices 4 and 5. Distinct C termini for H13a–c protein isoforms arise from the inclusion or exclusion of a penultimate exon (see the text). Membrane topology for the larger protein was determined by Weihofen et al. (2002). The proteins putatively translated from H13d and H13e have not been described previously; hence, membrane topology is not shown. |C| Northern blots performed with neonatal brain mRNA identify transcripts utilizing all five polyA sites identified by 3′ RACE. (D) DNA methylation at the CpG islands: the H13 promoter, the Mcts2 promoter, and the H13 downstream. Horizontal lines denote individual DNA strands on which filled circles represent methylated CpG dinucleotides; open circles are unmethylated CpGs. Genomic DNA was extracted from whole brain of neonatal mice carrying uniparental partial duplications for distal Chromosome 2 [MatDp(dist2) and PatDp(dist2)].
Allele-specific utilization of H13 polyA sites. (A) Enlarged schematic showing the fourth intron of H13 including Mcts2, the differentially methylated CpG island, and the H13d polyA site. Exons are black rectangles, and splice patterns are indicated with arrows depicting the direction of transcription. The fourth exon of H13a-c is shown entirely contained within the 3'UTR of H13d. The 3' RACE primer used in B anneals within this region. (B) 3' RACE was performed using a gene-specific primer within the fourth exon of H13a-c, in material from F1 hybrids of the C57BL/6J (B6) and Mus musculus castaneus (cast) substrains. The major products utilized either the H13a (1.1 kb) or H13d (0.6 kb) polyA sites. Products corresponding to H13b and H13c are not seen, presumably due to their lower abundance and larger size reducing the efficiency of amplification. Additional splice variants of H13a-c are seen in whole newborn brain but not in ES cells. In all figures, the maternal strain is birthed brain but not in ES cells. In all figures, the maternal strain is.

Allele-specific polyadenylation in intersubspecies hybrid mice

The differentially methylated region (DMR) at the Mcts2 promoter lies ~200 base pairs [bp] downstream from the H13d polyA site [Fig. 2A], and ~3.5 kb downstream from H13e [Supplemental Fig. 1]. The proximity of these two sites to the DMR raised the possibility that they could be differentially utilized on maternally and paternally derived alleles. To address this, 3' RACE was performed in neonatal brain material from intersubspecies hybrid mice. Following gel purification, 3' RACE products were sequenced over regions containing SNPs between the parental strains, revealing the relative utilization of each polyA site by each parental allele [Fig. 2B,C, Supplemental Fig. 1]. Transcripts terminating upstream of the Mcts2 DMR [H13d and H13e] originate preferentially from paternally derived alleles [Fig. 2C, Supplemental Fig. 1]. In contrast, the H13a polyA site downstream from the DMR is utilized predominantly on maternally derived alleles [Fig. 2C]. This is observed in cDNA prepared from whole neonatal brains and from embryonic stem (ES) cells [Fig. 2C]. RT–PCR amplification of exons 1–3 yields products that originate from both parental alleles [Fig. 2D], consistent with biallelic transcriptional initiation from the unmethylated H13 promoter [Fig. 1C]. The allele-specific utilization of polyA sites within a single cell type indicates that polyadenylation can be regulated by epigenetic mechanisms.

Disrupted polyadenylation following uniparental inheritance

To confirm these findings in an independent genetic system, polyadenylation at H13 was studied in mice carrying uniparental partial disomies for distal Chromosome 2. This allowed the expression of H13 transcripts to be compared in mice that inherit two copies of the H13 locus from their mothers [MatDp(dist2)] versus mice that inherit two paternal copies of this genomic region [PatDp(dist2)]. Northern blots were generated from total brain RNA and hybridized with a probe to the common 5' exons [Fig. 1A, probe #3]. This confirmed the preferential maternal expression of transcripts utilizing the H13d and H13c polyA signals, but H13d transcripts were not detectable on this genetic background [C3H × 101[F1]] [Fig. 3A]. Transcripts utilizing both the H13d and H13e polyA signals are present at greater levels following pa-
Collectively, these data indicate that polyA sites situated 0.2 kb \([H13d]\) and 3 kb \([H13e]\) upstream of the Mcts2 DMR are utilized primarily on unmethylated paternal alleles. Maternal alleles preferentially generate H13 transcripts that elon-gate through the methylated and transcriptionally inactive Mcts2 promoter to the three polyA sites located 20 kb downstream \([H13a–c]\).

Alternative polyadenylation is controlled by maternal germine methylation

To further investigate the role of allele-specific methylation in this phenomenon, embryos lacking maternally derived methylation imprints were generated by crossing Dmnt31\(^{-/-}\) mothers with cast fathers \(\text{(Bourchhis et al. 2001)}\). Heterozygous offspring lack allele-specific DNA methylation at the Mcts2 promoter \(\text{(Fig. 4A,)}\) and this is associated with activation of the normally silent maternal allele \(\text{(Fig. 4B)}\). Qualitative 3’ RACE for the H13 gene demonstrates that loss of intronic methylation is associated with increased utilization of the H13d polyA site in Dmnt31\(^{-/-}\) embryos \(\text{(Fig. 4C)}\). This is entirely consistent with PatDp\(\text{(dist2)}\) individuals, which also exhibit biallelic hypomethylation at this region \(\text{(Fig. 3, data not shown)}\), despite up-regulation of H13d relative to H13a.

Figure 3. Disrupted polyadenylation in mice carrying uniparental partial disomies. \[A\] Northern generated with total RNA extracted from neonatal brains of mice carrying uniparental partial duplications for distal Chromosome 2. Nonduplicated mice are either homozygous, heterozygous, or noncarriers of the T11H translocation but in all cases will inherit distal Chromosome 2 biparentally. The blot was hybridized with probe \#3 from Figure 1A. One biological replicate was performed. Note that H13a transcripts are not completely absent following paternal inheritance, but H13d transcripts are not detected in MatDp\(\text{(dist2)}\). \[B\] Qualitative 3’ RACE assay, using the same gene-specific primer employed in Figure 2B. Two biological replicates gave similar results \(\text{(data not shown).} \[C\] Combined expression of transcripts extending through the DMR \([H13a–c]\) in neonatal brains measured by QRT–PCR. Data are presented as ratios of the mean Ct value in relation to the mean Ct value for \(\beta\)-actin for each genotype. Microarray expression data confirmed that this endogenous control gene did not change significantly between MatDp\(\text{(dist2)}\) and PatDp\(\text{(dist2)}\) samples \(\text{(Supplemental Fig. 2).} \[D\] Expression of H13d transcripts measured by QRT–PCR. The amplified region lies within 500 bp of the H13d polyA site. Data are presented as in \(C\).

Figure 4. Disrupted polyadenylation in maternal imprint-free embryos. \[A\] Methylation status of the Mcts2 CpG island in pooled embryos \(\text{(n = 5)}\) derived from Dmnt31\(^{-/-}\) dams and wild-type cast sires \(\text{(Dmnt31}\)^{+/+}\) and control embryos, determined by sequencing bisulphite-modified genomic DNA. Parental strains were distinguished on the basis of a SNP between Mus domesticus \(\text{(B6)}\) and Mus musculus \(\text{(cast)}\) alleles. Filled circles are methylated, and closed circles are unmethylated. Strands derived from the same PCR amplification are joined by braces on the right. \[C\] Allele-specific RT-PCR sequencing assay for Mcts2 in Dmnt31\(^{-/-}\) hybrid embryos \(\text{[pooled, n = 5]}\). \[D\] Qualitative 3’ RACE in cDNA from pooled Dmnt31\(^{-/-}\) hybrid embryos. \[D\] 3’ RACE products from \(C\) were excised and sequenced over regions containing SNPs between the two parental strains.
approximately equal levels on each allele in sites show reciprocal allele-specific biases in wild-type H13a at levels sufficient to assess their allele-specific expression. The imprinted H13 locus has the potential to provide novel insights into this important transcriptional phenomenon. Further work is now required to elucidate the precise mechanism by which identical DNA sequences can generate alternatively polyadenylated transcripts within a single cellular environment.

Materials and methods

Tissue sources

Mice with unparental partial disomies (or unparental duplications) were generated by intercrossing heterozygotes for the Tj2-911H reciprocal translocation (Cattanach and Beechey 1997). Mice carrying maternal duplications for distal Chromosome 2 MatDp(dist2) also carry paternal duplications for distal Chromosome 9, and vice versa. Maternal imprint-free embryos were generated as described previously (Bourc'his et al. 2001). RNA from the hybrid ES cell line used in Figure 2 was a gift from Wolf Reik.

DNA methylation

RNA and DNA were prepared using the RNaseasy and DNaseasy kits (Qia-gen) and quantified using an Agilent 2100 bioanalyzer. Bisulphite conversion was performed as described (Wood et al. 2007). All strands presented in Figures 1 and 4 differed either in epigenotype or in the presence of unconverted cytosines, or were derived from separate PCR amplifications. All strands showed >95% conversion.

Gene expression

5′ and 3′ RACE-ready cDNAs were synthesized using the SMART RACE kit (Clontech), with 0.5 µg of polyA+ mRNA per 100 µL of final volume. Details of all RACE and PCR primers, cycling conditions, and SNPs are listed in the Supplemental Material. Northern blots were performed using Affymetrix U430_v2) were hybridized according to standard Affymetrix protocols, and data were analyzed using the GCOS/MAS5 algorithm.

QRT-PCR was performed using custom TaqMan assays designed by Applied Biosystems. cDNA was prepared using the SuperScript II kit (Invitrogen) primed with oligo DT. Three microliters of total RNA were used for a 50-µL final RT reaction volume, of which 2 µL were used per 10 µL of QPCR reaction. Expression of H13 isoforms was measured relative to
$\beta$-actin, which did not change significantly between MatDp(dist2) and PatDp(dist2) samples measured by microarray [Supplemental Fig. 2]. Sample values were determined using the relative expression ratio mathematical model [Pfaffl 2001], with Ct values transformed into expression ratios according to the formula:

$$\text{Ratio} = 2^{\left[\frac{\text{Ct}_{\text{control baseline}} - \text{Ct}_{\text{target baseline}}}{\text{Ct}_{\text{target sample}} - \text{Ct}_{\text{control sample}}}\right]}$$

where $\text{Ct}_{\text{target baseline}}$ denotes the mean of the Ct values measured for the target gene in the baseline sample, $\text{Ct}_{\text{control baseline}}$ is the equivalent value for the endogenous control gene, and $\text{Ct}_{\text{target sample}}$ and $\text{Ct}_{\text{control sample}}$ are the mean Ct values determined in the nonbaseline sample.

**Acknowledgments**

We thank Wolf Reik and En Li for hybrid ES cell RNA. We thank Nick Proudfoot for useful discussions. This work was supported by the Wellcome Trust (R.J.O.), the Generation Trust (A.J.W.), the Biotechnology and Biological Sciences Research Council (R.J.O.), and The European Molecular Biology Organization (R.S.).

**References**


