

Human–Ovine Comparative Sequencing of a 250-kb Imprinted Domain Encompassing the Callipyge (*clpg*) Locus and Identification of Six Imprinted Transcripts: *DLK1*, *DAT*, *GTL2*, *PEG11*, *antiPEG11*, and *MEG8*

Carole Charlier,¹ Karin Segers,¹ Danny Wagenaar,¹ Latifa Karim,¹ Stéphane Berghmans,¹ Olivier Jaillon,² Tracy Shay,³ Jean Weissenbach,² Noelle Cockett,³ Gabor Gyapay,² and Michel Georges^{1,4}

¹Department of Genetics, Faculty of Veterinary Medicine, University of Liège (B43), 4000-Liège, Belgium; ²Genoscope, Centre National de Séquençage, CP 5706, 91057 EVRY Cedex, France; ³Department of Animal, Dairy, and Veterinary Sciences, College of Agriculture, Utah State University, Logan, Utah 84322-4700, USA

Two ovine BAC clones and a connecting long-range PCR product, jointly spanning ~250 kb and representing most of the *MULGE5–OY3* marker interval known to contain the *clpg* locus, were completely sequenced. The resulting genomic sequence was aligned with its human ortholog and extensively annotated. Six transcripts, four of which were novel, were predicted to originate from within the analyzed region and their existence confirmed experimentally: *DLK1*, *DAT*, *GTL2*, *PEG11*, *antiPEG11*, and *MEG8*. RT-PCR experiments performed on a range of tissues sampled from an 8-wk-old animal demonstrated the preferential expression of all six transcripts in skeletal muscle, which suggests that they are under control of common regulatory elements. The six transcripts were also shown to be subject to parental imprinting: *DLK1*, *DAT*, and *PEG11* were shown to be paternally expressed and *GTL2*, *antiPEG11*, and *MEG8* to be maternally expressed.

[The sequence data described in this paper have been submitted to the GenBank data library under accession no. AF354168.]

The callipyge [from *καλι* (beautiful) and *πυγε* (buttocks)] phenotype reported in sheep is a muscular hypertrophy that affects primarily the hindquarters, is due to an increase in the proportion and diameter of fast twitch muscle fibers, and manifests itself at ~1 mo of age (Carpenter et al. 1996; Jackson et al. 1997a,b,c; Freking et al. 1998b,c). It was shown to be inherited and fully controlled by an autosomal locus (*clpg*) mapping to the telomeric end of sheep chromosome 18 (Cockett et al. 1994; Freking et al. 1998c). Subsequent marker-assisted segregation analysis revealed that the callipyge phenotype is subject to a unique parent-of-origin effect, referred to as polar overdominance, in which only heterozygous individuals having inherited the CLPG mutation from their sire express the muscular hypertrophy (Cockett et al. 1996).

As part of a positional cloning effort, we and others have mapped the *clpg* locus by linkage analysis to the 4 cM *IDVGA30–OY3* interval on distal OAR18q

(Fahrenkrug et al. 2000; Shay et al. 2001), constructed an ovine BAC contig spanning that interval (Segers et al. 2001), and refined the map position of the *clpg* locus by breakpoint analysis to a ~400 kb chromosome segment flanked by microsatellite markers *MULGE5* and *OY3* (Berghmans et al. 2001). The same interval was shown to contain the *DLK1* and *GTL2* genes, which were recently demonstrated to be reciprocally imprinted in the mouse and human: *DLK1* being expressed from the paternal allele and *GTL2* from the maternal allele (Miyoshi et al. 2000; Schmidt et al. 2000; Takada et al. 2000; Wylie et al. 2000). Therefore, these results point toward the existence of a novel, evolutionary conserved imprinted domain containing at least two imprinted genes.

Here, we contribute to the characterization of this imprinted domain by performing a human–ovine comparative analysis of ~250 kb of genomic sequence covering most of the marker interval containing the *clpg* locus. We identify and characterize the ovine *DLK1* and *GTL2* genes within this genomic sequence, as well as four novel transcripts (*DAT*, *PEG11*, *antiPEG11*, and *MEG8*) and show all of these to be strongly

⁴Corresponding author.

E-MAIL michel.georges@ulg.ac.be; FAX 32-04-366-41-22.

Article and publication are at www.genome.org/cgi/doi/10.1101/gr.172701.

expressed in skeletal muscle and imprinted in the sheep.

RESULTS

Genomic Sequencing of a 250-kb Region Spanning the Ovine *clpg* Locus

Two BAC clones (359E3 and 229G11) and a 7.5-kb long-range PCR product connecting these two BACs were predicted to jointly cover the central 250 kb of the *MULGE5-OY3* interval shown to contain the *clpg* locus (Berghmans et al. 2001). The two BAC clones were subjected to shotgun sequencing at threefold redundancy by use of standard procedures, whereas the cloned PCR product was sequenced at sixfold redundancy by use of the EZ::TN Transposon Insertion System (Epicentre Technologies). The resulting sequence traces were assembled by use of the PHRED-PHRAP-CONSED suite of programs (Ewing and Green 1998; Ewing et al. 1998; Gordon et al. 1998). A 70-kb segment spanning the *GTL2* gene (see below) was finished at 10-fold redundancy.

This approach generated 55 contigs with an average size of 4520 bp (range, 444–68,800 bp) for a total of 248,602 bp. Error probabilities computed from the quality value, q , as defined by Ewing & Green (1998) averaged 59 corresponding to an estimated error rate of $1/(1.2 \times 10^{-6})$.

Aligning the Orthologous Ovine and Human Sequences

The ovine contigs were ordered and oriented by performing BLAST searches (Altschul et al. 1990) against the finished orthologous human genomic sequence obtained from GenBank (accession nos. AL132711, AL117190, and AL132709). This approach allowed us to predict the position of 30 of the 55 contigs jointly spanning 214,466 bp of ovine sequence (Fig. 1).

To confirm these predictions, we designed outward pointing primers at the ends of all 30 contigs and performed connecting PCR reactions on genomic and BAC DNA with the right and left primers of adjacent contigs, respectively. These experiments confirmed the predicted contig order and orientation (data not shown) and indicated that none of the gaps between adjacent contigs exceeds 2100 bp (data not shown).

We then performed a matrix comparison of the two sequences with the COMPARE and DOTPLOT programs (GCG Wisconsin Package) with a window of 49 bp. Figure 1 illustrates the resulting graph, clearly indicating extensive sequence conservation between the two species in this region.

Next, we identified 49-bp windows in the human sequence exhibiting $\geq 98\%$ similarity with a supposedly orthologous window in the ovine sequence. In total, 6125 such highly similar windows were identi-

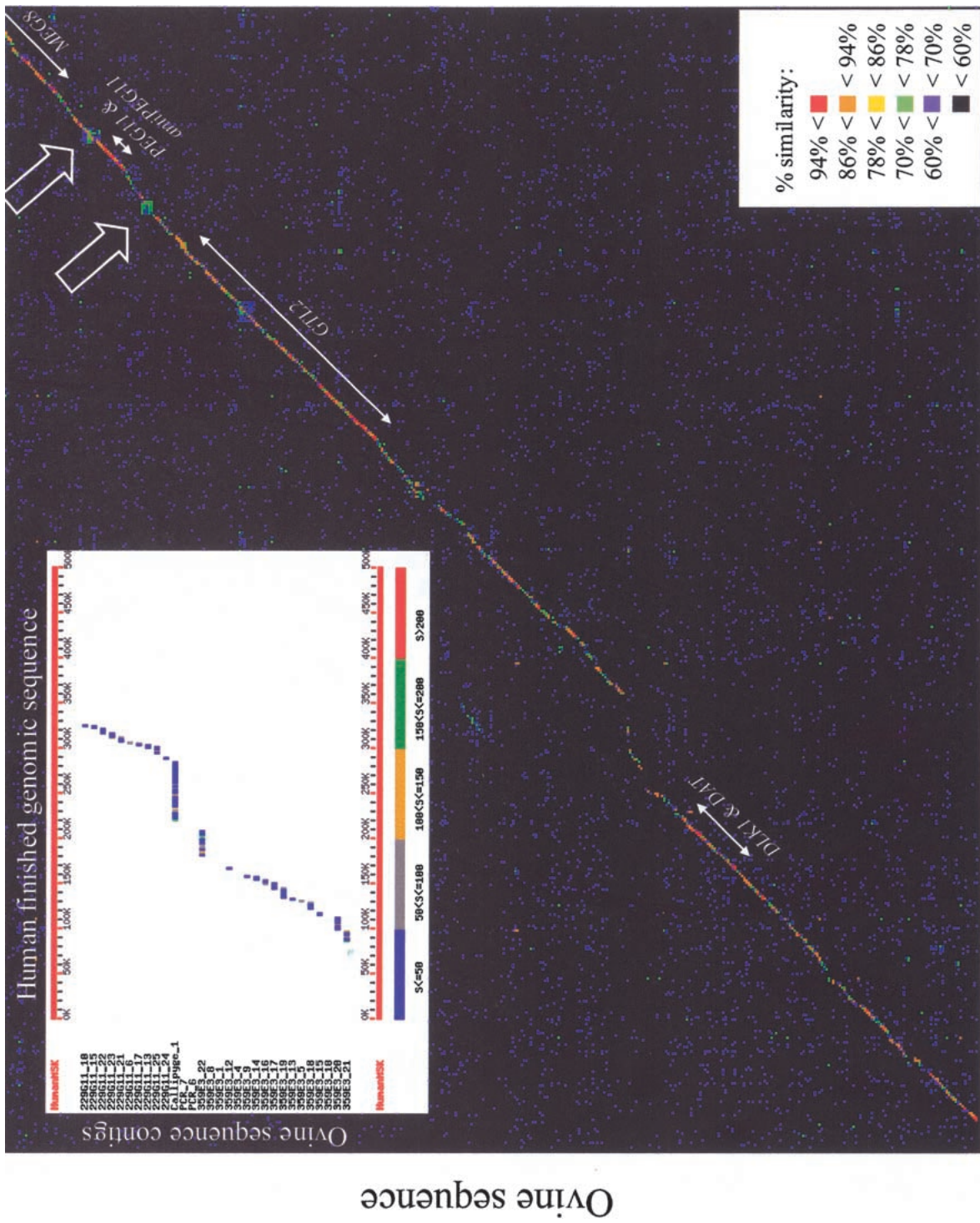
fied and marked as orthologous anchor points connecting the two species. Sequences between adjacent anchor points (average distance, 37 bp; range, 0–19,283) were aligned by use of the algorithm of Needleman and Wunsch (1970) implemented with the BESTFIT program (GCG Wisconsin Package). This approach resulted in the 271,963 bp aligned sequences shown in Figure 2. According to this alignment, the ovine sequences would have undergone 1730 deletions when compared with the human sequence amounting to a total of 46,100 bp (average basepair per deletion, 26.6; range, 1–1438), whereas the human sequences would have undergone 1520 deletions when compared with the ovine sequence amounting to a total of 21,963 bp (average basepair per deletion, 14.4; range, 1–750).

A similarity profile was generated along this aligned sequence by sliding a 100-bp window through the aligned sequences and computing for each window the percentage similarity between the two sequences (Fig. 2). The overall similarity score was 44.72%, ranging from 0% to 99%.

(G + C) Content and Identification of CpG Islands

The average (G + C) content of the examined human and ovine sequences was 52.26% and 54.31%, respectively, and, therefore, substantially higher than the average 42 (G + C) % across the human genome. Analysis of the moving average (G + C) content computed across the analyzed human and ovine sequence with a 200-bp sliding window (Fig. 2) indicates that this chromosome segment belongs to an isochores (Bernardi 1995) with (G + C) content superior to 50%, with the exception of (1) an interspersed repeat-rich insertion in the human sequence (positions 94–108 kb), (2) a shared interspersed repeat-rich region (positions 140–150 kb), and (3) a shared interspersed repeat-poor region (positions 240–244 kb), the latter three regions being comparatively (A + T) rich.

Following Gardiner-Garden and Frommer (1987) and Larsen et al. (1992), we looked for CpG islands defined as regions >200 bp with a moving average (C + G) % >50% and a moving average of observed over expected CpG content >0.6. This analysis revealed four CpG islands or clusters of CpG islands common to both human and sheep: (1) a small CpG island at position 49 kb; (2) a larger one at position 67.5 kb (coinciding with the 5' end of *DLK1*; see below); (3) a cluster of CpG islands stretching from position 174 to 180 kb (corresponding to the 5' end of *GTL2*; see below); and (4) a cluster of CpG islands at positions 236–238 kb. It is noteworthy that each of these shared CpG islands coincides with phylogenetic footprints exhibiting similarity scores superior to 80%. In addition to these four conserved elements, we detected three ovine-specific CpG islands at the following positions: (1) 31



Human sequence

Figure 1 Dot-matrix comparison of the human and ovine genomic sequences. The percent similarity represented by a given dot is indicated according to the color code shown. Positions and names of the identified genes are shown. Simple sequence repeats flanking the *PEG11* gene are marked by arrows. (Inset) Alignment and ordering of 30 of the 55 ovine sequence contigs against the orthologous human genomic sequence by use of BLAST. (Altschul et al. 1990). The ovine contigs are shown sorted according to their relative positions predicted on the basis of the observed human-ovine homology. Ovine contigs showing no hits with the human sequence in this figure were positioned on the basis of BLAST searches performed at lower stringency (data not shown).

Table 1. Interspersed Repeats in the Callipyge Imprinted Domain

Human (250,000 bp analyzed)			Ovine (214,466 bp analyzed)		
Family*	No. of elements	Percentage of sequence	Family*	No. of elements	Percentage of sequence
SINEs	132	11.01%	SINEs	83	5.58%
ALUs	70	7.84%	rum.spec.	40	3.39%
MIRs	62	3.17%	MIRs	33	1.74%
LINEs	56	10.27%	LINEs	63	10.30%
LINE1	22	7.96%	LINE1	34	6.76%
LINE2	31	2.06%	LINE2	17	1.25%
L3/CR1	3	0.25%	L3/CR1	2	0.24%
			BovB/Art2	10	2.05%
LTR elements	19	4.77%	LTR elements	13	1.70%
MaLRs	8	0.96%	MaLRs	5	0.36%
ERVL	7	3.26%	ERVL	5	1.08%
ERV-Class I	1	0.20%	ERV-Class I	2	0.11%
ERV-Class II	0	0.00%	ERV-Class II	0	0.00%
DNA elements	13	1.46%	DNA elements	9	0.66%
MER1 type	10	0.70%	MER1 type	6	0.45%
MER2 type	2	0.68%	MER2 type	1	0.03%
Unclassified	0	0%	Unclassified	1	0.04%
Total interspersed repeats		27.51%	Total interspersed repeats		18.29%
Simple repeats	44	1.04%	Simple repeats	33	0.91%
Low complexity	31	0.84%	Low complexity	32	0.51%

*Repeats are labeled as defined in Smit (1999) and Smit and Green (2000).

kb; (2) 76 kb (coinciding with the 3' end of *DLK1*; see below); and (3) 187 kb.

Identification of DNA Sequence Repeats

Repetitive sequences were identified in both the human and sheep sequence by use of RepeatMasker (Smit and Green 2000) and a modified version of the RepBase Database (RepBase Update 1999; A.F.A. Smit and P. Green, unpubl.). Table 1 summarizes the results that were obtained.

Interspersed repeats account for 27.51% of the human sequence versus 18.29% of the ovine sequence.

An obvious explanation for this observation would be that repetitive sequences are better characterized in the human compared with the ovine. To verify this hypothesis, we performed BLAST searches with the masked ovine sequence against itself to detect potentially new ruminant-specific repetitive sequences. No evidence was found for the existence of such as-yet-uncharacterized repeats.

Another noticeable observation is the high proportion of interspersed repeats that are located in orthologous positions in human and sheep (Fig. 2), which indicates that the acquisition of the corresponding repeats most likely predates the time of human-ovine divergence. On the other hand, a substantial number of the more sizeable gaps as defined by

BESTFIT either in the human or ovine sequence coincide with the occurrence of interspersed repeats in the other species. Approximately 13 kb of a patchwork of different interspersed repeats seem to have been inserted en bloc within the human sequence (positions 95–108 kb). Traces from more modest repeat-mediated insertions are noticeable throughout the analyzed sequence.

The 27.51% of total interspersed repeats in the human sequence are considerably lower than the 41% expected on average (Smit 1999) for human sequences with G + C content between 50% and 54%. This low

Figure 2 Ovine-human comparative sequence annotation of the *callipyge* imprinted domain. (Gaps) White bars correspond to the gaps defined by the BESTFIT program (GCG Wisconsin Package) in, respectively, the ovine (O) and human (H) sequence when aligning the orthologous sequences between anchorpoints defined as in Results. These gaps are reported in the background in dark grey throughout the rest of the figure. The red bars in the "Gaps: O" line report the limits between adjacent ovine sequence contigs. (Repeats) Reports the positions of repetitive sequences identified with RepeatMasker (Smit & Green 2000) in, respectively, the ovine (O) and human (H) sequences, and labeled according to the color code shown at the bottom of the figure. (CpG islands/(G + C) content) The moving average (G + C) content computed using a 200-bp sliding window is shown for the human sequence (H) only, as the profiles were virtually identical in both species; the red line corresponds to 50% average (G + C) content. CpG islands defined according to Gardiner-Garden and Frommer (1987) and Larsen et al. (1992) are represented as yellow bars for the ovine (O) and human (H) sequences respectively. (Similarity) Similarity profile obtained by sliding a 100-bp window through the BESTFIT alignment and plotting the corresponding percent identity between the two species (H and O). The bottom line corresponds to 50% identity, and the red line to 80% identity. (ESTs) Genomic sequences aligned with at least one human (H) or bovine (O) EST when performing high-stringency BLAST searches with the corresponding repeat-masked genomic sequence against dbEST, are indicated by the thick yellow lines. The thin yellow lines connect noncontiguous genomic sequences which are either matching contiguous segments of the same EST (corresponding therefore to introns), or matching different ESTs belonging to the same cDNA clone. (GENSCAN "+") (GENSCAN "-") output of the GENSCAN software (Burge and Karlin 1997) applied to the "+" and "-" strand respectively. Predicted transcription start sites and polyadenylation signals are symbolized by white ◀ and ▶, respectively. Exons are shown as green ("+" strand) or red ("- strand) boxes, respectively. Conserved exons with a "forward-backward" probability value ≥ 0.9 in at least one of the two species are shown in bright colors; other exons in darker colors. The positions of three previously-described microsatellite markers (Berghmans et al. 2001): *MULGE6*, *BULGE33*, and *BMS1561* are marked by blue asterisks on top of the sequence alignment. The position of the identified and confirmed genes or transcripts (*DLK1*, *DAT*, *GTL2*, *PEG11*, *antiPEG11*, and *MEG8*) are shown at the bottom of the sequence alignment.

abundance seems primarily due to a conspicuous shortage of Alu repeats representing only 7.84% of the analyzed sequence versus the expected 22.2% for sequences with 50% to 54% G + C content (Smit 1999).

Gene Prediction

Ab Initio Gene Prediction

The repeat-masked human and bovine sequences were analyzed with the GENSCAN (version 1.0) ab initio gene prediction program (Burge and Karlin 1997). The obtained results are illustrated in Figure 2. For the human sequence, GENSCAN predicted 33 exons corresponding to seven genes on the + strand, and 24 exons corresponding to five genes on the – strand. For the ovine sequence, GENSCAN predicted 18 exons corresponding to five genes on the + strand, and 56 exons corresponding to seven genes on the – strand.

A subset of the predicted exons was considered to contain highly reliable predictions on the basis of the fact that exons were shared between the two species and were associated with a forward-backward probability $P \geq 0.90$ in at least one of these. The forward-backward probability P estimated by GENSCAN can be used as an approximation of the likelihood that the predicted exon is a true exon (Burge and Karlin 1997).

Five such highly reliable exons, predicted by GENSCAN to be part of the same gene, were identified on the arbitrarily defined + strand (corresponding to genes transcribed in the *cen* to *ter* direction in Fig. 2) between positions 67 and 77 kb. Each of these exons coincides with an evolutionary footprint in which the sequence similarity between the two species exceeds 80%. The position of the 5' exon also coincides with the position of one of the conserved CpG islands. It is now known that these exons correspond to the *DLK1* gene (see below).

An additional highly reliable short exon was found on the + strand at position 187 kb, coinciding with the position of a CpG island detected only in the ovine sequence (see above). Apart from this conserved exon, the respective ovine and human genes to which this highly reliable exon was assigned differed completely, except for the start site for transcription, which was also conserved between the two species. This conserved start site and exon are now known to be genuine and part of the *GTL2* gene (see below).

Finally, a >4 kb long open reading frame (ORF) shared on the – strand by both species was detected between positions 235 and 239.5 kb. The position of this remarkable ORF coincides with an equally long phylogenetic footprint as well as a shared cluster of CpG islands. This ORF is postulated to correspond to a novel gene, which will be referred to as *PEG11* (see below).

Gene Prediction from EST Matching

Next, we performed BLAST searches between the human repeat-masked genomic sequence and the human ESTs in the dbEST database (<http://www.ncbi.nlm.nih.gov/dbEST/>), as well as with the ovine repeat-masked genomic sequence and the bovine ESTs in the dbEST database. The corresponding results are shown in Figure 2.

By so doing, we identified three regions that were spanned in both species by multiple physically connected EST matches. By physically connected, we mean either two noncontiguous matches with the same EST (therefore, the intervening sequences more than likely corresponding to introns), or two noncontiguous matches to distinct ESTs originating from the same cDNA clone. The first of these regions was located between position 67 and 77 kb and corresponds to the *DLK1* gene also identified by GENSCAN. The second such region extends from position 176 to 215 kb and corresponds to the *GTL2* gene, for which the transcription start site and one exon were also predicted by GENSCAN. The third region extends from position 250 kb to the end of the alignment between the ovine and human sequence. To the best of our knowledge, this sequence corresponds to a novel gene that we will refer to as *MEG8* (see below).

In addition to these physically connected EST clusters, we identified a number of isolated ESTs or UniGene clusters hybridizing in silico across the analyzed region. Most noticeably, we found a series of ESTs, both human and bovine that would map to the region immediately 3' of the *DLK1* gene (positions 77–81 kb), which will jointly be referred to as *DLK1-associated transcripts* or *DATs*; a series of isolated ESTs mapping between positions 235 and 239 kb, that is, coinciding with the *PEG11* ORF predicted by GENSCAN and a number of ESTs mapping between *PEG11* and *MEG8*.

Gene Confirmation

DLK1

The gene identified on the + strand between positions 67 and 77 kb both by ab initio gene prediction and EST matching corresponds to the previously described *Delta-like (DLK1)* gene (also known as *pG2*, *FA-1*, *PREF1*, *SCP-1* or *Zog*). *DLK1* is a transmembrane protein that belongs to the EGF-like homeotic protein family. *DLK1* is thought to play an important role in differentiation (particularly adipogenesis, hematopoiesis, B cell lymphopoiesis, and neuroendocrine differentiation) and tumorigenesis of several cellular types (for review, see Laborda 2000). *DLK1* is characterized by five exons corresponding to the five highly reliable exons predicted by GENSCAN in this region. However, alternative splicing of the fifth exon yields at least six

distinct messengers in the mouse, referred to as A, B, C, C2, D, and D2 (Smas et al. 1994). The largest two splicing variants (A and B) encode proteins that undergo proteolytic cleavage in the juxtamembrane region to release a soluble 50-kD ectodomain (Smas et al. 1997). In the bovine, the equivalents of splicing variants A and C2 have been found, as well as an additional one referred to as E (Fahrenkrug et al. 1999). In the human, the equivalent of the A form has been described as well as a splicing variant closely resembling the murine and bovine C2 variants, which we will refer to as the C3 variant (Lee et al. 1995). Figure 3 summarizes the structure of the splicing variants found in man and ruminants.

Primers were designed in exons 1 (DLK1.UP1) and 5 (DLK1.DN1) of the ovine *DLK1* gene (see Fig. 3 and Table 2) and used to study its expression by RT-PCR in a range of tissues obtained from an 8-wk-old, conventional (i.e., noncallipyge) sheep (individual 9535): brain, heart, adipose tissue, kidney, liver, lung, lymph nodes, skeletal muscle, pancreas, intestine, spleen, and thymus. A strong amplification product of 1314 bp was obtained from skeletal muscle mRNA (Fig. 3). This amplicon was shown by cycle sequencing to correspond to the C2 splice variant. A smaller amplification product (565 bp) was obtained from the same tissue as well, sequenced and shown to contain exon 1, the 5' 44 bp of exon 2, and the C2 version of exon 5 truncated by

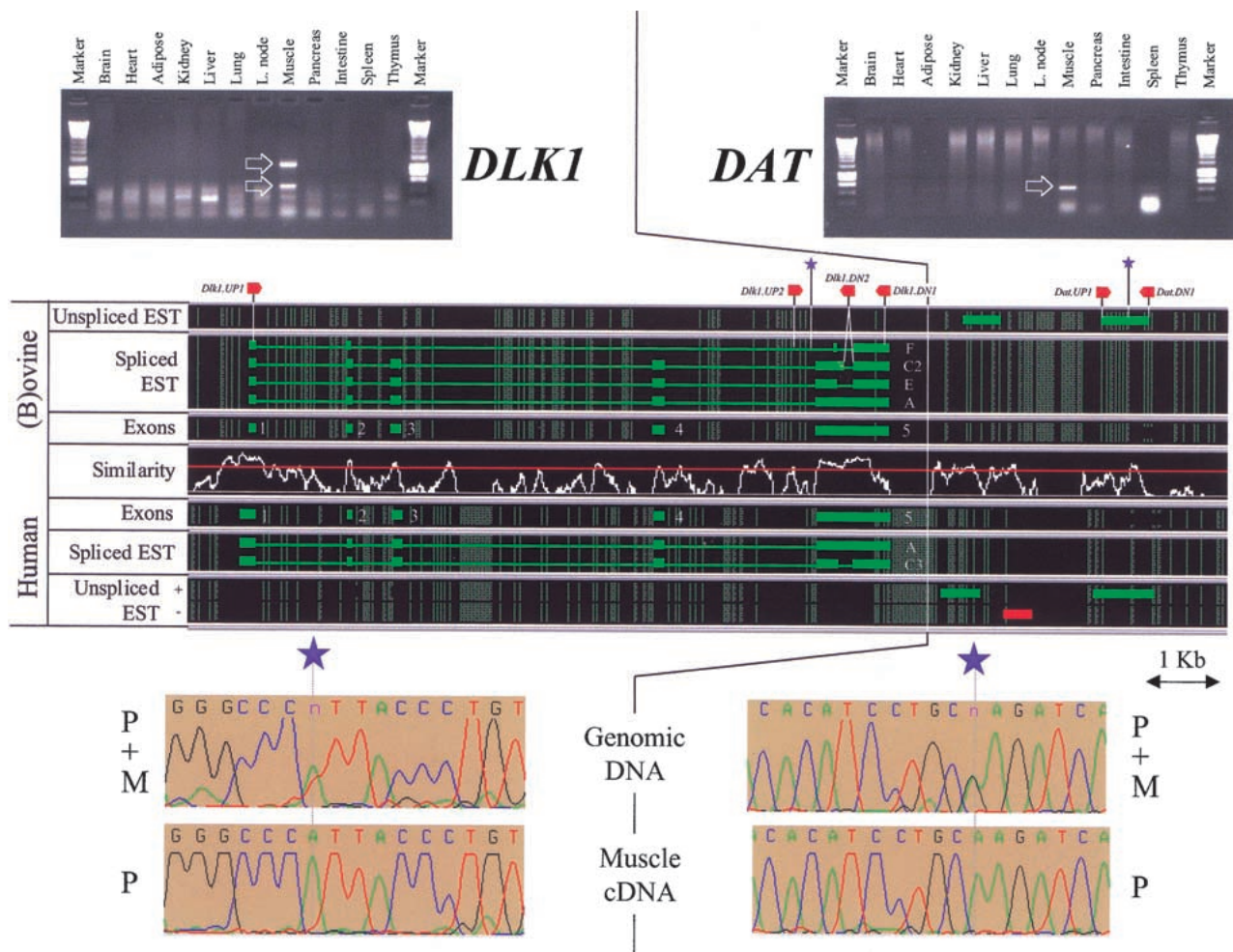


Figure 3 Organization, expression profile, and imprinting of *DLK1* and *DAT*. Spliced and unspliced ESTs whose transcriptional orientation could be inferred [(green) transcribed from the + strand; (red) transcribed from the - strand], were aligned with respect to the moving average similarity profile obtained for the human-ovine framework alignment (see Fig. 2). (Thick lines) EST matches; (thin lines) predicted introns. (Red arrows) The approximate positions of the primers used for genomic PCR, RT-PCR, and cycle sequencing; (blue asterisks) the approximate position of the SNPs used to evaluate the imprinting status. Expression profiles deduced from the analysis—by agarose gel electrophoresis—of RT-PCR products obtained with RNA from a range of tissues are shown, with specific amplification products marked by white arrows. Imprinting is demonstrated by comparison of the sequence traces obtained by cycle-sequencing equivalent amplification products obtained by genomic PCR versus RT-PCR. The parental origin of the alleles are labeled P (paternal) and M (maternal).

Table 2. Utilized Primer Sequences

Name	DNA sequence (5'–3')
DLK1.UP1	CTTTCGCGTCCGCAACCAGAAGC
DLK1.UP2	ACTCCCTCACCTCGAGCGTTTGA
DLK1.DN1	GAACAGACCGCACAGAGACAGG
DLK1.DN2	GGATGGTGAAGCAGATGGCCTGGG
DAT.UP1	TCCTTACCAGCACTCAATG
DAT.DN1	TAGCCCTTGGTTTTGGTTTAT
GTL2.UP1	AAACAAGGCAACCACCGGCAGTGA
GTL2.DN1	GATCACTATTGCAAAGCGTCCCAG
PEG11.UP1	ACAGCTCAACAGTGGAGGCTATG
PEG11.DN1	ATCTCGTGGCAGAGCAGATGAAC
MEG8.UP1	CCGTAGAGTCCCTCGGGATGGATCTAAC
MEG8.UP2	CCCTGATTGGTACATGTGAACAGT
MEG8.DN1	CCAGGGTAATTCCATGTGAAGATGCTC
MEG8.DN2	ATCAGCCAGTCCCTCTGAAGCTGC

232 bp at its 5' end (Fig. 3). The corresponding transcript has the potential to be translated into a protein of 134 residues including a peptide signal, two half EGF-like homeotic domains, as well as the *DLK1* transmembrane and cytoplasmic domains. It will be referred to in Figure 3 as splice variant F. Note, however, that the corresponding exon 2 splice acceptor and exon 5 splice donor sites do not satisfy the usual consensus splice sites. Therefore, the biological relevance of this F form is uncertain.

The observation that *DLK1* transcripts were detected exclusively in skeletal muscle contradicts to some extent with previous results documented for *DLK1* expression in postnatal tissue and for which no expression was found in muscle, but was found in pancreas (for review, see Laborda et al. 2000).

DLK1-Associated Transcripts (DATs)

Interestingly, a series of ESTs were shown to cluster within 4 kb of the 3' end of *DLK1*, both in human and sheep. A Unigene cluster (Hs 116631) comprising 12 human ESTs maps between positions 79.5 and 80.5 kb, that is, at 2.5 kb from the 3' end of *DLK1* and coinciding with a well-conserved evolutionary footprint. These ESTs jointly cover 803 bp, have the same transcriptional orientation as *DLK1* and are characterized by a conserved polyadenylation signal at the 3' end, but have no obvious ORF. In addition, two human ESTs map to an evolutionary footprint between positions 77.5 and 78.8 kb, one (accession no. BF793204) transcribed from the same strand as *DLK1*, the other (accession no. AI807698) supposedly from the complementary strand. In sheep, a bovine 5' EST (accession no. BF040125) transcribed from the + strand maps to the same evolutionary footprint between positions 77.8 and 78.3 kb.

Primers were designed on the basis of the ovine sequence between positions 79.6 and 80.4 kb (Fig. 3 and Table 2) and used to detect potential transcripts

originating from this region by use of oligo(dT)-primed cDNA generated from the previously described battery of tissues. As for *DLK1* a strong specific amplification product of the expected size was obtained from skeletal muscle mRNA only and shown to be specific by cycle sequencing. As no amplification product was obtained in the absence of a reverse transcriptase step, this product is considered to be a genuine RT-PCR product and not to reflect contamination of the RNA preparation with genomic DNA.

Both 5' and 3' RACE experiments were performed with the same primers and skeletal muscle RNA. A 544-bp 3' RACE amplification product was obtained and sequenced. It confirmed the functionality of the conserved polyadenylation signal predicted from the analysis of the human ESTs. So far, no specific 5' RACE amplification products could be obtained. To verify whether the corresponding ESTs might in fact reflect the existence of novel, as-yet-undescribed *DLK1* transcripts, we performed RT-PCR experiments with a series of primers distributed across the *DLK1* coding sequence and the DAT.DN1 primer and skeletal muscle cDNA as template. So far, no specific amplification product could be obtained. The biological meaning of these DATs is being examined.

GTL2

The human and bovine EST clusters hybridizing in silico between positions 176 and 215 kb on the aligned human and ovine sequence (Fig. 2) correspond to the previously described *GTL2* or *MEG3* gene (Schuster-Gossler et al. 1996, 1998; Miyoshi et al. 2000). The structure of the gene as deduced from alignment of individual ESTs with the genomic sequence is illustrated in Figure 4.

The analysis of spliced messengers identified 12 exons in the human *GTL2* gene, and 10 exons in its (b)ovine homolog. All exons were bordered by the expected consensus splice acceptor and donor sites. It can be seen that the overall gene organization is very well conserved between the two species. The 10 exons identified in ruminants all have an obvious counterpart in the human gene. A detailed analysis of the intron-exon boundaries, however, reveals the sliding of the boundaries of exons 3, 7, 8, and 9 by up to 135 bp (Fig. 4). Comparison of the exon content of the differentially spliced ESTs reveals the existence of multiple alternatively spliced transcripts both in human and (b)ovine, thereby confirming earlier findings (Schuster-Gossler et al. 1998; Miyoshi et al. 2000). The absence of a conserved ORF, as well as the common occurrence of frame-shifting insertion deletions when comparing orthologous human and ovine exons, are in agreement with the previously formulated hypothesis according to which *GTL2* might function as an RNA (Schuster-Gossler et al. 1998; Miyoshi et al. 2000).

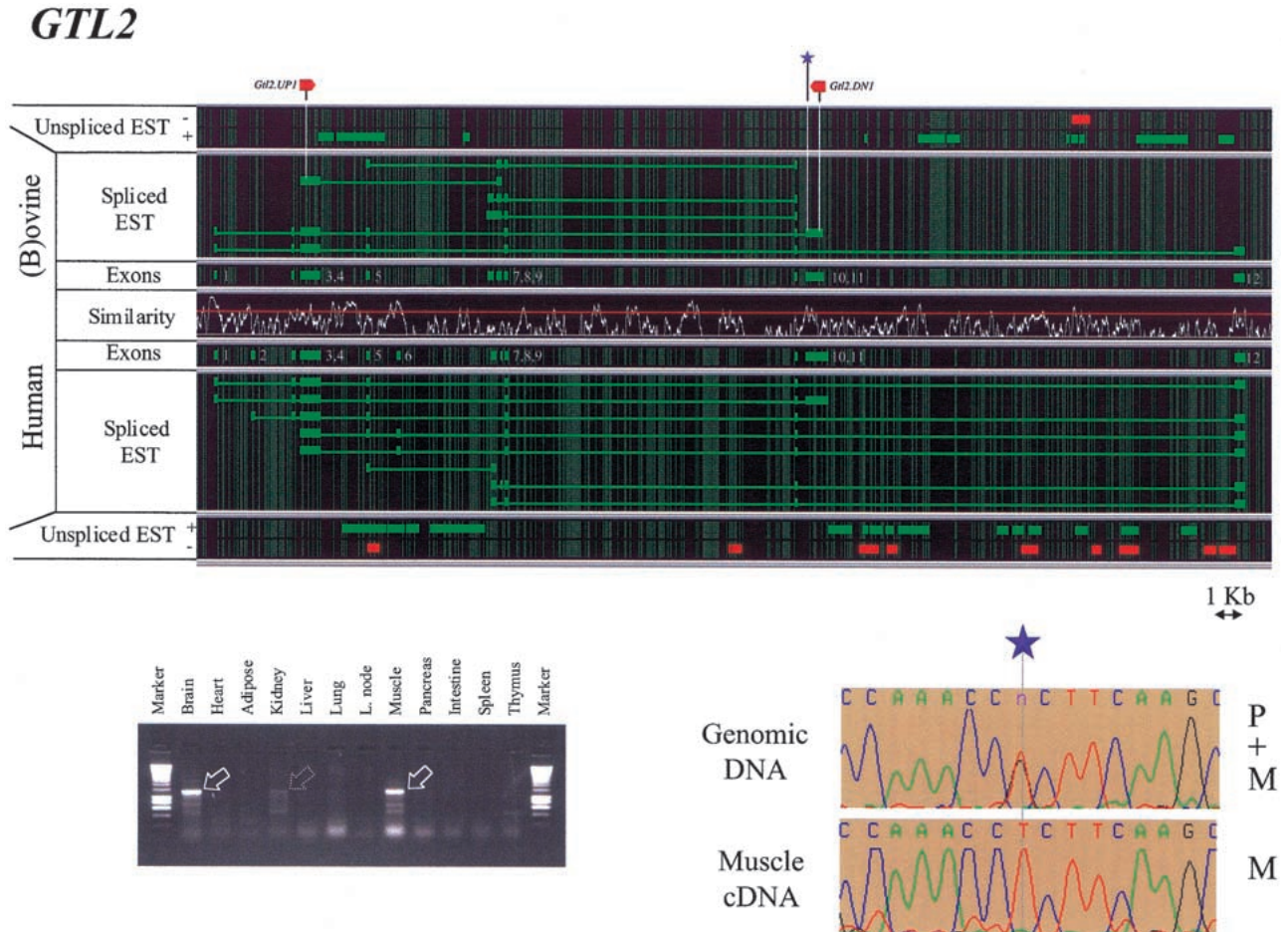


Figure 4 Organization, expression profile, and imprinting of *GTL2*. (For further details, see Fig. 3 legend.)

The mapping of unspliced human and bovine ESTs on their respective genomic sequences reveals the existence of additional transcripts containing both intronic sequences as well as unspliced exons. Both in human and bovine, these additional transcripts are concentrated between exons 4 and 7, as well as between exons 11 and 12. It is noteworthy that in addition to ESTs transcribed from the + strand, that is, having the same transcriptional orientation as all the spliced ESTs, a number of antisense transcripts originating from the – strand are observed as well, particularly in the region between exons 11 and 12.

Primers were designed in exons 4 and 11 of the ovine *GTL2* gene (see Table 2) and used in RT-PCR reactions with the previously described battery of RNAs obtained from sheep 9535. A strong amplification product of 1248 bp was obtained from brain, skeletal muscle, and, to a lesser extent, kidney RNA. The corresponding brain and skeletal muscle amplification products were shown by sequencing to be identical and to contain exons 4, 5, 9, 10, and 11. Additional, mainly smaller and weaker, amplification products were detected in the same three tissues.

PEGII and antiPEGII

GENSCAN predicts a gene comprising a single exon, transcribed from the – strand between positions 235 and 239 kb in both human and sheep on the basis of the presence of a conserved ORF that has the potential to code for 1358- and 1333-amino-acid residue proteins in human and sheep, respectively (Figs. 2 and 5). The position of the predicted ORF coincides strikingly with a region of very high nucleic acid similarity between the two species including two conserved CpG islands. BLASTP and TBLASTN searches reveal highly significant similarities between the central portion of the predicted protein (amino acid residues ~200–980) and the gag and pol polyproteins of gypsy-like LTR-retrotransposons from a broad taxonomic range including *Fugu rubripes*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Oryza sativa*, *Arabidopsis thaliana* and *Schizosaccharomyces pombe*. Analysis of the sequences flanking the predicted ORF does not reveal obvious evidence for LTR sequences that might be expected for retroviral-like sequences. Note, however, that two clusters of simple sequence repeats shown in Figure 1 flank the corresponding ORF. The correspond-

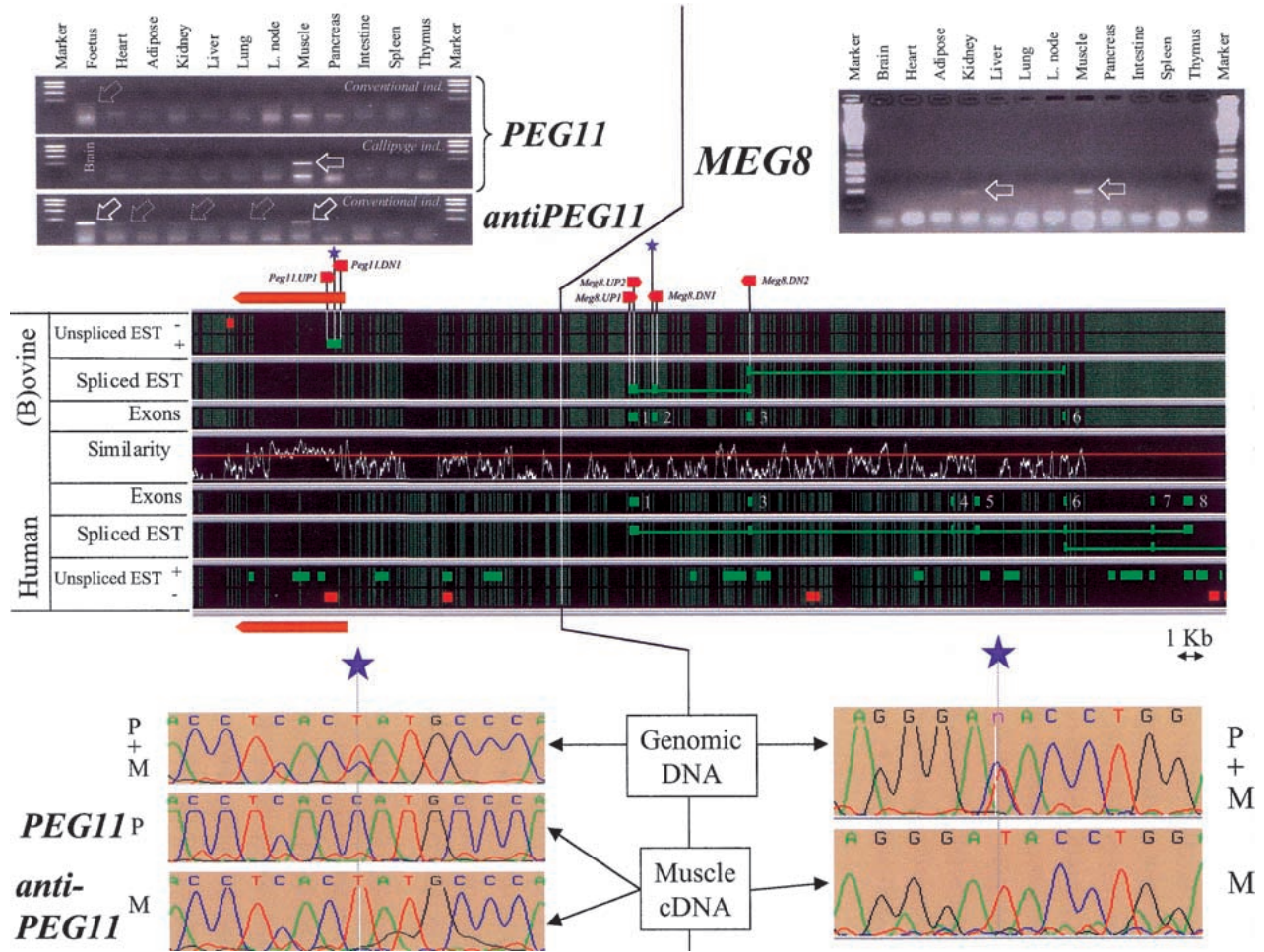


Figure 5 Organization, expression profile, and imprinting of *PEG11*, *antiPEG11*, and *MEG8*. For further details, see Fig. 3 legend. The large red arrow indicates the position of the *PEG11* ORF predicted by GENSCAN.

ing gene seems to be unique. Indeed, BLASTN searches performed with the human ORF against human genomic sequences (htgs database) do not reveal any highly similar hit other than the chromosome 14 locus itself, contrary to what would be expected if the corresponding gene was a member of a family of dispersed retrotransposon-like repetitive elements.

ESTs transcribed from the – strand and overlapping the predicted ORF were found in the human, bovine (Fig. 5), and mouse databases (data not shown), therefore supporting the fact that we might be dealing with a functional gene. The bovine EST corresponds to a 3' EST and is characterized by a poly(A)⁺ tail preceded by a canonical polyadenylation signal (conserved in human), therefore potentially defining a 3' end for the corresponding gene.

Primers were developed on the basis of the ovine sequence (Table 2) and used to detect potential transcripts by RT-PCR. cDNA was synthesized by use of the *PEG11.UP1* primer, therefore producing strand-specific cDNA that would correspond to mRNA encoding the

predicted ORF. As can be seen from Figure 5, no RT-PCR product was obtained from the previously described range of tissues isolated from the conventional (i.e., noncallipyge) 9535 individual. A weak band of the expected size was, however, detectable from RNA extracted from a conventional 21-day-old fetus, suggesting that the *PEG11* gene might be expressed in early development. However, when we performed the same experiment using RNA obtained from an 6-wk-old individual expressing the callipyge phenotype (sheep 0573), we obtained a strong amplification product of the expected size in skeletal muscle. The specificity of this PCR product was demonstrated by sequencing. The absence of a detectable RT-PCR product from the corresponding RNA preparations when eliminating the reverse transcription step (as well as the demonstration of imprinting; see below) indicates that the amplification products are genuine RT-PCR products and not due to contaminating genomic DNA.

Taken together, these results strongly suggest that we were dealing with a genuine, highly conserved

gene, expressed in early development and in skeletal muscle of at least callipyge individuals and potentially coding for a protein showing homology with the gag and pol proteins of retrotransposons. This gene will be referred to as *PEG11* (see below).

Analysis of the human and bovine ESTs mapping to the *PEG11* gene also revealed several ESTs predicted to be transcribed from the + strand that might therefore correspond to *antiPEG11* transcripts. RT-PCR experiments were set up to detect such putative *antiPEG11* transcripts by use of the PEG11.DN1 primer for cDNA synthesis. Figure 5 shows that putative *antiPEG11* transcripts were indeed detectable primarily in fetal RNA and skeletal muscle RNA of 8-wk-old conventional sheep (9535) and to a lesser extent in heart, kidney, and lung. An identical pattern was obtained with tissues from callipyge individual 0573 (data not shown). As for the *PEG11* transcripts, no amplification products were obtained in the absence of reverse transcriptase step.

Unexpectedly, PCR amplification performed with the PEG11.UP1 and PEG11.DN1 primers and RNA treated with reverse transcriptase in the absence of any primer yields an amplification pattern weaker but very similar to that obtained with the PEG11.DN1 primer for reverse transcription (data not shown). This could be due to residual reverse transcriptase activity when initiating the PCR reaction, despite our attempts to inactivate it by incubation at 99°C for 5 min. Alternatively, it might indicate the presence of double-stranded *PEG11* RNAs with short *PEG11* transcripts priming reverse transcription of *antiPEG11* transcripts. Neither of these hypotheses, however, explains why the corresponding RT-PCR products are not detected when the PEG11.UP1 primer was used for reverse transcription. The significance of this observation is being examined.

MEG8

As shown previously, a cluster of physically connected EST matches was shown to map between position 250 kb and the end of the sequence alignment (Fig. 2) and hypothesized to correspond to a novel gene referred to as *MEG8* (see below). Analyses of the spliced ESTs defined nine exons of which four were present in (b)ovine (exons 1, 2, 3, and 6) and eight in the human (exons 1, 3, 4, 5, 6, 7, 8, and 9, the last one being located ~5 kb downstream of the aligned human and sheep sequence; data not shown and Fig. 5). All predicted exons were flanked by consensus splice donor and acceptor sequences. As for *GTL2*, the absence of a conserved ORF and the frequent occurrence of frameshifting mutations when comparing orthologous exons suggests that *MEG8* might act via an RNA rather than a protein product as well.

In addition to these spliced ESTs, and in a manner reminiscent of the *GTL2* gene, several unspliced ESTs were found to map to this region, encompassing both exonic and intronic regions and being transcribed from both the + and - strands (Fig. 5).

Primers were designed in the predicted ovine exons 1 (MEG8.UP1) and 2 (MEG8.DN1), and used to perform RT-PCR experiments with the same tissues of sheep 9535 described above. Clear amplification products were obtained mainly from skeletal muscle and, to a lesser extent, from kidney (Fig. 5).

Testing the Imprinting Status of the Confirmed Genes in Sheep

As mentioned previously, there is ample evidence that the corresponding chromosome region is subject to parental imprinting. Therefore, the imprinting status of the six confirmed genes/transcripts was tested in sheep by (1) searching for DNA sequence polymorphisms in the transcribed portions of the corresponding genes, (2) identifying sheep that would be heterozygous for the corresponding polymorphisms, and (3) determining the presence of the respective alleles at the RNA level.

DLK1

A series of primer pairs were designed in the ovine *DLK1* gene that allowed for the amplification of all exons with some flanking intronic sequences from genomic DNA. Amplification products obtained from genomic DNA of sheep 9535 were cycle sequenced to search for DNA sequence polymorphisms. No polymorphisms were detected in the exonic sequences of *DLK1*. However, 9535 proved to be heterozygous for an A to T substitution found in intron 4, 27 bp downstream of exon 5 (Fig. 3). Genotyping its parents and relatives for the corresponding SNP as well as flanking microsatellites allowed us to unambiguously determine the parental origin of the two alleles (data not shown).

This SNP was used to test the imprinting of the ovine *DLK1* gene by genotyping pre-mRNA splicing intermediates. This assay was accomplished by performing RT-PCR experiments from skeletal muscle cDNA (the tissue in which *DLK1* transcripts were found to be most abundant) with primer DLK1.UP2 located in intron 4 downstream of the SNP and primer DLK1.DN1 located in exon 5. The amplification product corresponding to the C2 transcript was gel purified and sequenced with primer DLK1.DN2. This primer can only hybridize to C2 transcripts and not to genomic DNA. This experiment clearly demonstrated that only the paternal allele corresponding to the A residue could be found in the skeletal muscle *DLK1*

pre-mRNA population (Fig. 3), and, therefore, imprinting with paternal expression of the *DLK1* gene in sheep (Fig. 6).

This finding was confirmed on a second 8-wk-old conventional sheep (9556) that proved heterozygous for the same SNP.

DAT

Individuals 9535 and 9556 were shown by cycle sequencing to be heterozygous for two polymorphisms and one polymorphism in the PCR product obtained from genomic DNA with primers DAT.UP1 and DAT.DN1, respectively: a G to A substitution and a deletion of an A residue. As for *DLK1*, genotyping relatives allowed for unambiguous identification of the paternal and maternal alleles. Sequencing the corresponding RT-PCR product obtained from skeletal muscle cDNA indicated that only the paternal allele was present in the mRNA population (Fig. 3), and, therefore, imprinting of the *DAT* transcript with paternal expression (Fig. 6).

GTL2

As for *DLK1*, primer pairs were designed on the basis of the ovine genomic sequence to allow for the amplification of several of the *GTL2* exons from genomic DNA. Individuals 9535 and 9556 were found to be heterozygous for a G to T transversion in exon 11. Sequencing the RT-PCR product obtained with primers *GTL2*-UP1 and *GTL2*-DN1 from skeletal muscle RNA clearly indicated that only the T allele, corresponding to the maternal allele, was present in the mRNA population, and, therefore, imprinting of the ovine *GTL2* gene with maternal expression in this tissue (Figs. 4 and 6).

PEG11 and antiPEG11

Sequencing the PCR product obtained with primers PEG11-UP1 and PEG11-DN1 from genomic DNA revealed a C to T transition for which 9535 proved heterozygous. The analysis of 12 callipyge individuals from our callipyge flock (e.g., see Shay et al. 2001) allowed us to identify two (0573 and 0525) that proved heterozygous for the same C to T transition and for which the parental origin of the respective alleles could be determined unambiguously.

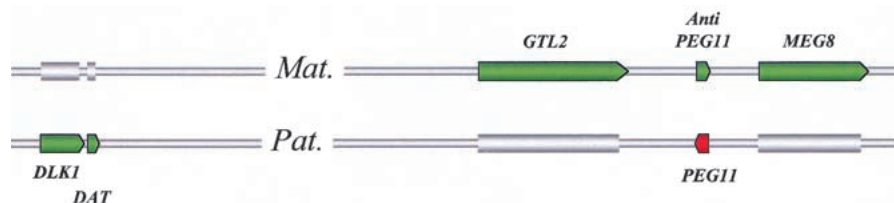


Figure 6 Summary of the imprinting status in sheep of the genes/transcripts identified in the callipyge imprinted domain. Expressed, parent-of-origin specific alleles (Pat. or Mat.) are marked by green and red arrows when transcribed from the + or – strand, respectively.

The RT-PCR products corresponding to the *PEG11* transcript and obtained from skeletal muscle RNA of callipyge individuals 0573 and 0525 were sequenced and shown to contain exclusively the paternal allele (Fig. 5), therefore, showing imprinting with paternal expression of the *PEG11* transcript (Fig. 6) and justifying its name (paternally expressed gene 11).

Sequencing the RT-PCRs product corresponding to the *antiPEG11* transcripts detected in skeletal muscle, heart, kidney, and lung from conventional individual 9535 indicated the exclusive presence of the maternal allele (Fig. 5), therefore, showing imprinting with maternal expression of the *antiPEG11* transcript (Fig. 6).

Note that sequencing the amplicons obtained from skeletal muscle, heart, kidney and lung, when the reverse transcription was performed in the absence of primers revealed the maternal allele only as for the *antiPEG11* transcripts (data not shown).

MEG8

By the same strategy used for the other genes, individual 9535 was shown to be heterozygous for three SNPs in the *MEG8* gene: a C to T transition in exon 1, a C to T transition in exon 2, and a C to A transversion in exon 2, whereas individual 9556 proved heterozygous for the same C to T transition in exon 2 and an additional T to C transition in exon 2. The RT-PCR products obtained from skeletal muscle RNA with the MEG8.UP2 and MEG8.DN2 primers (Table 2) were sequenced and shown to contain only the maternal allele (Fig. 5), therefore showing the imprinting status with maternal expression of the *MEG8* gene as well (Fig. 6), hence justifying its name (maternally expressed gene).

DISCUSSION

Here, we describe the ovine–human comparative sequence analysis of an ~250,000 bp genomic region encompassing a novel, evolutionarily conserved imprinted domain. In silico annotation led to the prediction of four genes: (1) *DLK1*, a previously described member of the EGF-like homeotic protein family, (2) *PEG11*, a novel unique gene that putatively codes for a protein showing homology with the gag and pol polyproteins of retroviral elements, (3) *GTL2*, a previously described noncoding RNA, and (4) *MEG8*, a novel noncoding RNA. In addition, we predicted the occurrence of two other transcripts, one associated with *DLK1* and referred to as *DAT*, the other with *PEG11* and referred to as *antiPEG11*.

The predictions were based on the outputs of the ab initio gene prediction program GENSCAN, results of in silico hybridization with EST databases, and

the conservation of the predicted features between human and ovine. The benefit of comparative sequence analysis in improving the reliability of gene predictions is well illustrated by examination of the gene predictions obtained initially by GENSCAN. GENSCAN predicted 57 exons in man (corresponding to 12 genes) and 74 exons in sheep (corresponding to 12 genes). We believe that only the seven exons that were predicted in both human and ovine are true functional exons, most of the remaining ones probably being erroneous. With the exception of *PEG11*, which was detected by GENSCAN primarily because of its unusually long intronless ORF, in this situation, the predictions based on EST matching were actually more informative than those of GENSCAN. Indeed, EST matching not only predicted *DLK1* as accurately as GENSCAN, but was essential for the identification of *GTL2*, *MEG8*, *DAT*, and *antiPEG11*. The inefficiency of GENSCAN in detecting the latter genes/transcripts is undoubtedly due to the fact that all of these seem to correspond to noncoding RNAs. Alignment of the human and ovine sequence also identified a number of highly conserved phylogenetic footprints that will guide future research toward defining functionally important elements in this region.

Analysis of the expression profile of the six predicted genes/transcripts by RT-PCR, not only confirmed their genuine nature but revealed that all of them have in common a preferential and pronounced expression in skeletal muscle. This expression suggests that all six transcripts are likely sharing common regulatory elements as previously suggested for *DLK1* and *GTL2* (Schmidt et al. 2000; Takada et al. 2000). The preferential expression in skeletal muscle is also particularly relevant in light of the muscular hypertrophy characterizing the callipyge phenotype.

DLK1 and *GTL2* were recently shown to be imprinted in mouse and man (Miyoshi et al. 2000; Schmidt et al. 2000; Takada et al. 2000; Wylie et al. 2000). We demonstrate in this work that the ovine orthologs are likewise reciprocally imprinted: *DLK1* being expressed from the paternal allele, *GTL2* from the maternal allele. In addition, we show that the other studied genes/transcripts are all undergoing parental imprinting as well: *DAT* and *PEG11* are paternally expressed, whereas *antiPEG11* and *MEG8* are maternally expressed (Fig. 6).

It is striking that four of the defined genes/transcripts (*DAT*, *GTL2*, *antiPEG11*, and *MEG8*) apparently do not code for a protein product but rather might act as RNA effector molecules. This observation corroborates the recurrent findings of noncoding RNAs in imprinted domains (e.g., Sleutels et al. 2000). We cannot exclude the possibility that a number of these might in fact belong to the same transcriptional unit in a manner reminiscent of the *Air* transcript described for the *Igf2r* locus (Lyle et al. 2000). The fact that a

number of ESTs map between *GTL2* and *PEG11*, and even more between *PEG11* and *MEG8*, supports such a hypothesis.

An additional striking feature of both *GTL2* and *MEG8* is the occurrence of antisense ESTs that seem to cluster at defined positions within these genes. The fact that such ESTs are encountered at relatively high frequencies in the existing EST databases suggests that these are genuine antisense transcripts rather than annotation errors that would cause their erroneous orientation. The biological significance of these putative antisense transcripts, however, remains to be established.

The intronless ORF of the *PEG11* gene that has homologies with the gag and pol polyproteins of LTR-retrotransposons suggests, at first glance, that it belongs to a family of selfish transposonlike elements. However, the observation that the corresponding sequence is unique in the genome, its very high conservation among human, mouse, and (b)ovine reflecting a strong selective pressure, its expression in skeletal muscle, and the apparent absence of LTR sequences, suggest that *PEG11* is an essential mammalian gene. Its differential expression in conventional and callipyge individuals also suggest that it might be involved in the determinism of the callipyge phenotype.

This work is an essential step toward unraveling the molecular biology underlying the polar overdominance characterizing the callipyge phenotype. Examination of the effect of the *clpg* genotype on the expression of the six identified transcripts as well as sequencing of the entire domain to identify the CLPG mutation is in progress.

METHODS

RT-PCR Experiments and Cycle Sequencing

Total RNA was extracted from 1 g of tissue by use of RNA Insta-Pure (Eurogentech) according to the manufacturer's instructions. First strand cDNA was synthesized by use of the GeneAmp RNA PCR Kit (Perkin Elmer) starting from 1 µg of total RNA primed either with an oligo(dT) primer (*DLK1*, *GTL2*, *DAT*, *MEG8*) or the transcript-specific primers reported in Table 2 (*PEG11* and *antiPEG11*). The actual PCR amplifications were carried out by use of standard procedures, a DNA Thermal Cycler 480 (Perkin Elmer), and the gene-specific primers reported in Table 2. PCR products were separated on a 1% agarose gel and gel purified by use of the GeneClean Kit (Bio 101). Approximately 500 ng of purified PCR product was used as template for 25 cycles of cycle sequencing performed with the BigDye Terminator Cycle Sequencing Kit (PE Applied Biosystems). Products of the sequencing reaction were analyzed on an ABI377 automatic sequencer.

ACKNOWLEDGMENTS

This project was supported by grants from (1) Fonds de la Recherche Fondamentale Collective, Belgium (2.4525.96), (2) Crédit aux Chercheurs (1.5.134.00) from the Fonds National

de la Recherche Scientifique, Belgium, (3) Crédit à la Recherche from the University of Liège, (4) the Utah Center of Excellence Program, (5) the USDA/NRI Competitive Grants Program (USDA/NRICGP grants 94-04358, 96-35205, and 98-03455), and (6) the Utah Agricultural Experiment Station, Utah State University, Logan, Utah 84322-4819. C.C. is Chargé de Recherche from the Fonds National de la Recherche Scientifique, Belgium. We thank Anne Ferguson-Smith and Martina Paulsen for fruitful discussions.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**: 403-410.
- Berghmans, S., Segers, K., Shay, T., Georges, M., Cockett, N. E. and Charlier, C. 2001. Breakpoint mapping positions the callipyge gene within a 285 kilobase chromosome segment containing the Gtl-2 gene. *Mamm. Genome* **12**: 183-185.
- Bernardi, G. 1995. The human genome: Organization and evolutionary history. *Annu. Rev. Genet.* **29**: 445-476.
- Burge, C. and Karlin, S. 1997. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**: 78-94.
- Cockett, N.E., Jackson, S.P., Shay, T.L., Nielsen, D., Green, R.D., and Georges, M. 1994. Chromosomal localisation of the callipyge gene in sheep (*Ovis aries*) using bovine DNA markers. *Proc. Natl. Acad. Sci.* **91**: 3019-3023.
- Cockett, N.E., Jackson, S., Shay, T.L., Farnir, F., Berghmans, S., Snowden, G., Nielsen, D., and Georges, M. 1996. Polar overdominance at the ovine callipyge locus. *Science* **273**: 236-238.
- Carpenter, C.E., Rice, O.D., Cockett, N.E., and Snowden, G.D. 1996. Histology and composition of muscles from normal and callipyge lambs. *J. Anim. Sci.* **74**: 388-393.
- Ewing, B. and Green, P. 1998. Base-calling of automated sequencer traces using PHRED I. Error probabilities. *Genome Res.* **8**: 186-194.
- Ewing, B., Hillier, L., Wendl, M.C., and Green, P. 1998. Base-calling of automated sequencer traces using PHRED I Accuracy assessment. *Genome Res.* **8**: 175-185.
- Fahrenkrug, S.C., Freking, B.A., and Smith, T.P.L. 1999. Genomic organization and genetic mapping of the bovine PEF-1 gene. *Biochem. Biophys. Res. Comm.* **264**: 662-667.
- Fahrenkrug, S.C., Freking, B.A., Rexroad III, C.E., Leymaster, K.A., and Smith, T.P.L. 2000. Comparative mapping of the ovine *clpg* locus. *Mamm. Genome* **11**: 871-876.
- Freking, B.A., Keele, J.W., Beattie, C.W., Kappes, S.M., Smith, T.P., Sonstegard, T.S., Nielsen, M.K., and Leymaster, K.A. 1998a. Evaluation of the ovine callipyge locus: I. Relative chromosomal position and gene action. *J. Anim. Sci.* **76**: 2062-2071.
- Freking, B.A., Keele, J.W., Nielsen, M.K., and Leymaster, K.A. 1998b. Evaluation of the ovine callipyge locus: II. Genotypic effects on growth, slaughter, and carcass traits. *J. Anim. Sci.* **76**: 2549-2559.
- Freking, B.A., Keele, J.W., Shackelford, S.D., Wheeler, T.L., Koohmaraie, M., Nielsen, M.K., and Leymaster, K.A. 1998c. Evaluation of the ovine callipyge locus: III. Genotypic effects on meat quality traits. *J. Anim. Sci.* **77**: 2336-2344.
- Gardiner-Garden, M. and Frommer, M. 1987. CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**: 261-282.
- Gordon, D., Abajian, C., and Green, P. 1998. CONSED: A graphical tool for sequence finishing. *Genome Res.* **8**: 195-202.
- Jackson, S.P., Green, R.D., and Miller, M.F. 1997a. Phenotypic characterization of Rambouillet sheep expressing the callipyge gene: I. Inheritance of the condition and production. *J. Anim. Sci.* **75**: 14-21.
- Jackson, S.P., Miller, M.F., and Green, R.D. 1997b. Phenotypic characterization of Rambouillet sheep expressing the callipyge gene: II. Carcass characteristics and retail yield. *J. Anim. Sci.* **75**: 125-132.
- Jackson, S.P., Miller, M.F., and Green, R.D. 1997c. Phenotypic characterization of Rambouillet sheep expressing the callipyge gene: III. Muscle weights and muscle weight distribution. *J. Anim. Sci.* **75**: 133-139.
- Laborda, J. 2000. The role of epidermal growth factor-like protein DLK in cell differentiation. *Histol. Histopathol.* **15**: 119-129.
- Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. 1992. CpG islands as gene markers in the human genome. *Genomics* **13**: 1095-1107.
- Lee, Y.L., Helman, L., Hoffman, T., and Laborda, J. 1995. DLK, pG2 and Pref1 mRNAs encode similar proteins belonging to the EGF-like superfamily. Identification of polymorphic variants of this RNA. *Biochim. Biophys. Acta* **1261**: 223-232.
- Lyle, R., Watanabe, D., te Vruchte, D., Lerchner, W., Smrzka, O.W., Wutz, A., Schageman, J., Hahner, L., Davies, C., and Barlow, D.P. 2000. The imprinted antisense RNA at the Igf2r locus overlaps but does not imprint Mas1. *Nat. Genet.* **25**: 19-21.
- Miyoshi, N., Wagatsuma, H., Wakana, S., Shiroishi, T., Nomura, M., Aisaka, K., Kohda, T., Surani, M., Kaneko-Ishino, T., and Ishino, F. 2000. Identification of an imprinted gene, Meg3/Gtl2 and its human homologue MEG3, first mapped on mouse distal chromosome 12 and human chromosome 14q. *Genes Cells* **5**: 211-220.
- Needleman, S.B. and Wunsch, C.D. 1970. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**: 443-453.
- Schmidt, J.V., Matteson, P.G., Jones, B.K., Guan, X.-J., Tilghman, S.M. 2000. The *Dlk1* and *Gtl2* genes are linked and reciprocally imprinted. *Genes & Dev.* **14**: 1997-2002.
- Schuster-Gossler, K., Simon, D., Guénet, J.-L., Zachgo, J. and Gossler, A. 1996. *GTL2lacZ*, an insertional mutation on mouse chromosome 12 with parental origin dependent phenotype. *Mamm. Genome* **7**: 20-24.
- Schuster-Gossler, K., Bilinski, P. Sado T., Ferguson-Smith A. and Gossler A. 1998. The mouse *gtl2* gene is differentially expressed during embryonic development, encodes multiple alternatively spliced transcripts, and may act as an RNA. *Dev. Dyn.* **212**: 214-228.
- Segers, K., Vaiman, D., Berghmans, S., Shay, T., Cockett, N., Georges, M., and Charlier, C. 2001. Construction and characterization of an ovine BAC contig spanning the callipyge locus. *Anim. Genet.* **31**: 352-359.
- Shay, T., Berghmans, S., Segers, K., Meyers, S., Womack, J., Beever, J., Georges, M., Charlier, C. and Cockett, N.E. 2001. Construction of a contig spanning a 4.6 centimorgan interval containing the CLPG locus. *Mamm. Genome* **12**: 141-149.
- Sleutels, F., Barlow, D.P., and Lyle, R. 2000. The uniqueness of the imprinting mechanism. *Curr. Opin. Genet. Dev.* **10**: 229-233.
- Smas, C.M., Green D., and Sul, H.S. 1994. Structural characterization and alternate splicing of the gene encoding the preadipocyte EGF-like protein Pref-1. *Biochemistry* **33**: 9257-9265.
- Smas, C.M., Chen, L., and Sul, H.S. 1997. Cleavage of membrane-associated pref-1 generates a soluble inhibitor of adipocyte differentiation. *Mol. Cell. Biol.* **17**: 977-988.
- Smit, A.F.A. 1999. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr. Opin. Genet. Dev.* **9**: 657-663.
- Smit, A.F.A. and Green, P. 2000. RepeatMasker server on World Wide Web. <http://repeatmasker.genome.washington.edu>.
- Takada, S., Tevendale, M., Baker, J., Georgiades, P., Campbell, E., Freeman, T., Johnson, M.H., Paulsen, M., and Ferguson-Smith, A.C. 2000. *Dlk* (*Delta-like*) and *Gtl2* are closely linked reciprocally imprinted genes on mouse chromosome 12 and are paternally methylated and co-expressed during development. *Curr. Biol.* **10**: 1135-1138.
- Wylie, A.A., Murphy, S.K., Orton, T.C., and Jirtle, R.L. 2000. Novel imprinted DLK1/GTL2 domain on human chromosome 14 contains motifs that mimic those implicated in IGF2/H19 regulation. *Genome Res.* **10**: 1711-1718.

Received November 27, 2000; accepted in revised form February 16, 2001.