The callipyge mutation enhances bidirectional long-range *DLK1-GTL2* intergenic transcription in cis

Haruko Takeda*, Florian Caiment*, Maria Smit†, Samuel Hiard‡, Xavier Tordoir*, Noelle Cockett†, Michel Georges*§, and Carole Charlier*

*Unit of Animal Genetics, Faculty of Veterinary Medicine and Centre of Biomedical Integrative Genoproteomics (CBIG), University of Liège (B43), 20 Boulevard de Colonster, 4000-Liège, Belgium; †Animal, Dairy, and Veterinary Sciences, College of Agriculture, Utah State University, Logan, UT 84322-4700; and ‡Research Unit in Systems and Modelling, Department of Electrical Engineering and Computer Science and CBIG, Faculty of Applied Sciences, University of Liège (B29), 4000-Liège, Belgium

Communicated by James E. Womack, Texas A&M University, College Station, TX, April 7, 2006 (received for review February 2, 2006)

The callipyge mutation (*CLPG*) is an A to G transition that affects a muscle-specific long-range control element located in the middle of the 90-kb *DLK1-GTL2* intergenic (IG) region. It causes ectopic expression of a 327-kb cluster of imprinted genes in skeletal muscle, resulting in the callipyge muscular hypertrophy and its non-Mendelian inheritance pattern known as polar overdominance. We herein demonstrate that the *CLPG* mutation alters the muscular epigenotype of the *DLK1-GTL2* IG region in cis, including hypomethylation, acquisition of novel DNase-I hypersentivite sites, and, most strikingly, strongly enhanced bidirectional, long-range IG transcription. The callipyge phenotype thus emerges as a unique model to study the functional significance of IG transcription, which recently has proven to be a widespread, yet elusive, feature of the mammalian genome.

DNA methylation | DNase-I hypersensitivity | intergenic region | noncoding RNA

The callipyge phenotype is an inherited muscular hypertrophy of sheep. It is characterized by polar overdominance, an unusual mode of inheritance in which only heterozygotes having received the *CLPG* mutation from their sire express the phenotype (1). The *CLPG* mutation is an A-to-G transition in a conserved dodecamer motif located in the 90-kb intergenic (IG) region separating the imprinted *DLK1* and *GTL2* genes on sheep chromosome 18 (refs. 2 and 3; Fig. 1). This motif was assumed to be part of a muscle-specific locus control region (LCR), because the *CLPG* mutation causes ectopic expression of a core cluster of neighboring genes in postnatal skeletal muscle, a tissue in which these genes are normally silenced (6, 7). Genes whose expression is affected by the *CLPG* mutation include (*i*) the paternally expressed protein encoding *DLK1* and *PEG11* genes, located, respectively, 64 kb proximally and 88 kb distally from the *CLPG* mutation, and (*ii*) the maternally expressed noncoding RNA genes *GTL2*, *antiPEG11*, *MEG8*, and *MIRG*, located between 33 and 262 kb distally from the *CLPG* mutation, as well as their multiple C/D small nucleolar RNA and microRNA (miRNA) guests (8, 9). With the exception of *PEG11*, all these genes are transcribed toward the telomere. The effect of the *CLPG* mutation is cis-restricted and subordinate to imprinting control because it does not perturb the monoallelic expression of the target genes (6).

It was recently shown that the callipyge phenotype can be caused by ectopic expression of DLK1 protein in skeletal muscle as observed in $+/C^{Pat}$ individuals (10). The lack of phenotypic expression in C/C animals is postulated to be due to translational inhibition of padumnal *DLK1* transcripts by noncoding madumnal transcripts (11). A direct role for miRNAs in this trans effect is suggested by the demonstration of RNA interference-mediated degradation of padumnal *PEG11* transcripts by miRNAs processed from madumnal *antiPEG11* transcripts (12).

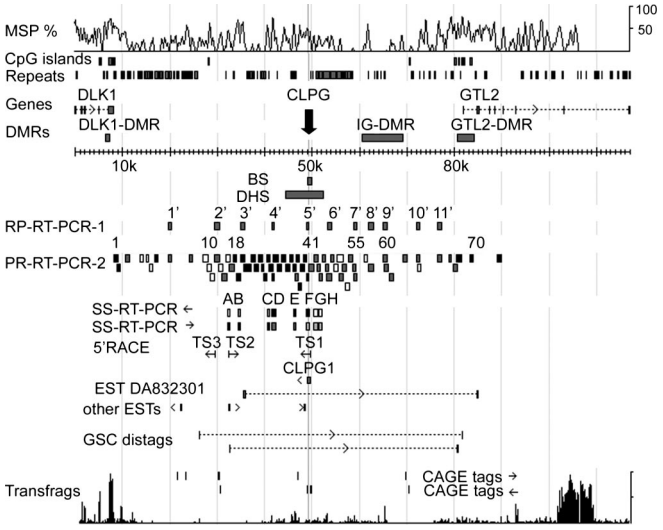

Fig. 1. Schematic representation of the ovine *DLK1-GTL2* IG region. MSP %, multispecies similarity profile; DMRs, differentially methylated regions; BS, bisulfite-sequenced segment (Fig. 2); DHS, segment explored for DNase-I hypersensitive sites (Fig. 3); RP-RT-PCR and SS-RT-PCR, location of amplicons used respectively in random primed (Fig. 4 *B* and *C*) or strand-specific RT-PCR experiments (Fig. 4 *A* and *D*); ←, targeting *D*←*G* transcripts; →, targeting *D*→*G* transcripts. For RP-RT-PCR-2 and SS-RT-PCR, which were performed in a $+/CLPG^{Pat}$ fetus, amplicons that gave strong RT-PCR products are labeled in black, those yielding weak RT-PCR products are in gray, and those that did not yield any RT-PCR product are in white. 5′ RACE, transcription start sites (TS) identified by 5′ RACE. GSC ditags and CAGE tags, ''gene signature cloning'' ditags and ''cap analysis gene expression'' tags identified in ref. 4. Transfrags, a local transfrag profile obtained by microarray analysis (5).

How the *CLPG* mutation operates such profound, tissue-specific influence on the expression of genes, which can be as far as 262 kb away, remains unknown. Intriguingly, Freking *et al.* (2) detected an RNA species of unknown function (*CLPG1*) encompassing the mutation and transcribed toward *DLK1*. Using 5′ RACE, they identified a putative transcription start site at 478 bp from the *CLPG* site.

To gain additional insight into the mechanisms underlying the cis effect of the *CLPG* mutation, we studied its effect on three epigenetic features that are known to be correlated with the activation state of other LCRs: DNA methylation, DNase-I hypersensitivity, and IG transcription.

Conflict of interest statement: No conflicts declared.

Abbreviations: DHS, DNase-I hypersensitive site; IG, intergenic; LCR, locus control region; miRNA, microRNA; RT, reverse transcriptase.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AF354168 and DQ378061).

§To whom correspondence should be addressed. E-mail: michel.georges@ulg.ac.be.

Results

The CLPG Mutation Imposes a Distinct Hypomethylation Mark in Cis. To test whether the *CLPG* mutation might affect the methylation status of surrounding DNA, we performed bisulfite sequence analysis of a 777-bp segment encompassing the mutation and the putative *CLPG1* transcription start site (TS1 in Fig. 1). As is the case for most of the *DLK1-GTL2* IG region (8), this DNA fragment has a high G+C content (60.7%) but a lower than expected number of CpG dinucleotides (28 observed vs. 77 expected). It is characterized by five highly conserved elements, one of which spans the *CLPG* mutation. We performed the analysis on skeletal muscle DNA of 8-week-old animals, because the *CLPG* mutation is assumed to act in this tissue and the phenotype is expressed at that age. We studied two animals of each of the four *CLPG* genotypes. We studied the DNA strand that allows distinction of the *CLPG* and + allele. The PCR products were cloned, and the sequence of at least 38 independent clones were determined for each animal (65 on average) (Fig. 2*A*; see also Fig. 5, which is published as supporting information on the PNAS web site). The conversion rate for non-CpG C residues averaged 99.8%, demonstrating the efficacy of the bisulfite treatment.

We first examined the effect of *CLPG* genotype on the proportion of methylated CpG sites per molecule (Fig. 2*B*). The major conclusions from this analysis are as follows:

1. *CLPG* chromosomes clearly distinguish themselves from + chromosomes by virtue of a population of molecules with <20% CpG methylation.
2. This hypomethylated population accounts for >70% of *CLPG* molecules in C/C and $C^{Mat}/+$ animals but only for 45% in $+/C^{Pat}$ animals.
3. + molecules exhibit a broad, uniform distribution of methylation, irrespective of *CLPG* genotype.

We also examined the effect of *CLPG* genotype on the percentage methylation of individual CpG sites across molecules (Fig. 2*C*), leading to the following conclusions:

1. In +/+ animals, the methylation rate exhibits a wave-like pattern with amplitude of 20–30% methylation and wavelength of 100–150 bp. The wave oscillates around a mean that maximizes (80%) at position +314 (counting from the *CLPG* mutation).
2. In C/C animals, the methylation rate is flat throughout the molecule averaging at ≈20%, except for two adjacent highly methylated CpG sites (sites 18 and 19).
3. In heterozygotes ($C^{Mat}/+$ and $+/C^{Pat}$), the maternal chromosome in essence is identical to its counterpart in the corresponding homozygotes, i.e., C^{Mat} very much resembles C/C, whereas $+^{Mat}$ very much resembles +/+. The paternal chromosomes, on the other hand, differ from their counterparts in the homozygotes, leaning toward the status of the maternal homologue, i.e., the $+^{Pat}$ allele is less methylated than +/+ and C^{Pat} is more methylated than C/C.

The *CLPG* Allele Exhibits Specific DNase-I Hypersensitive Sites (DHSs) and Increased DNase-I Sensitivity. One of the hallmarks of LCR is the occurrence of DHSs in their immediate vicinity (13). Assuming that the *CLPG* mutation perturbs a LCR, we wanted to test for the presence and effect of the *CLPG* mutation on DHS in its neighborhood. To that effect, we purified nuclei from skeletal muscle of 8-week-old sheep of the four *CLPG* genotypes. The nuclei were treated with increasing concentrations of DNase-I. DNA was extracted and subjected to Southern blot analysis, focusing on an 8.2-kb SspI restriction fragment encompassing the *CLPG* mutation. We detected at least three DHSs in +/+ animals located at +690 bp (DHS_+1), −525 bp (DHS_+2), and −810 bp (DHS_+3) from the *CLPG* SNP. Most

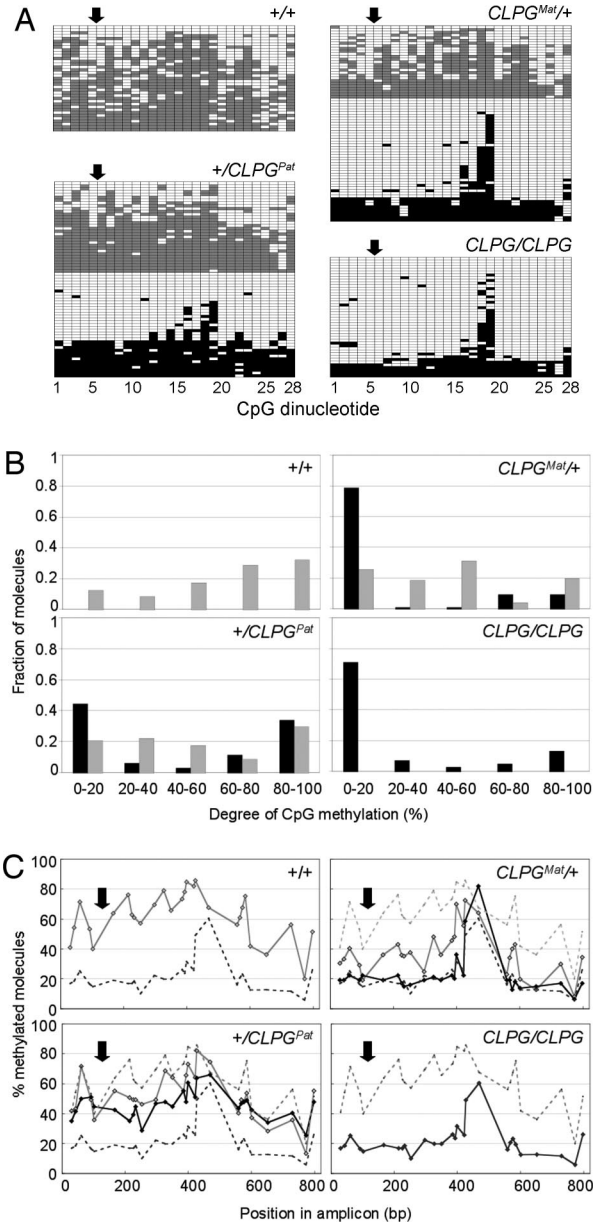

Fig. 2. DNA methylation analysis. (*A*) Representative bisulfite sequencing results for a 828-bp amplicon spanning the *CLPG* mutation for animals representing the four *CLPG* genotypes. Each line corresponds to a distinct molecule, each column to one of the 28 CpG dinucleotides in the amplicon. The corresponding coordinate is shaded in gray (+ allele) or black (*CLPG* allele) when methylated and white when unmethylated. The approximate position of the *CLPG* mutations is marked by the arrow. (*B*) Frequency distribution of the proportion of methylated CpG sites per molecule. The distribution for + alleles is shown as gray bars; *CLPG* alleles as black bars. (*C*) Percentage methylation for each of the 28 CpG sites across molecules. Results obtained for the + chromosomes are shown by the gray diamonds and *CLPG* chromosomes by the black diamonds. The curves obtained for the homozygous +/+ and C/C individuals are watermarked on all graphs. The position of the *CLPG* mutations is marked by the arrow.

interestingly, however, at least two additional DHSs were becoming apparent in C/C animals, located at +100 bp (DHS_C1) and –1,175 bp (DHS_C2) from the *CLPG* mutation. All DHSs, whether constitutive or *CLPG*-specific, correspond to regions of high conservation (Fig. 3*A*). Remarkably, the position of DHS_C1 coincides virtually exactly with the *CLPG* site and DHS_+1 coincides with TS1. The profiles obtained in heterozy-

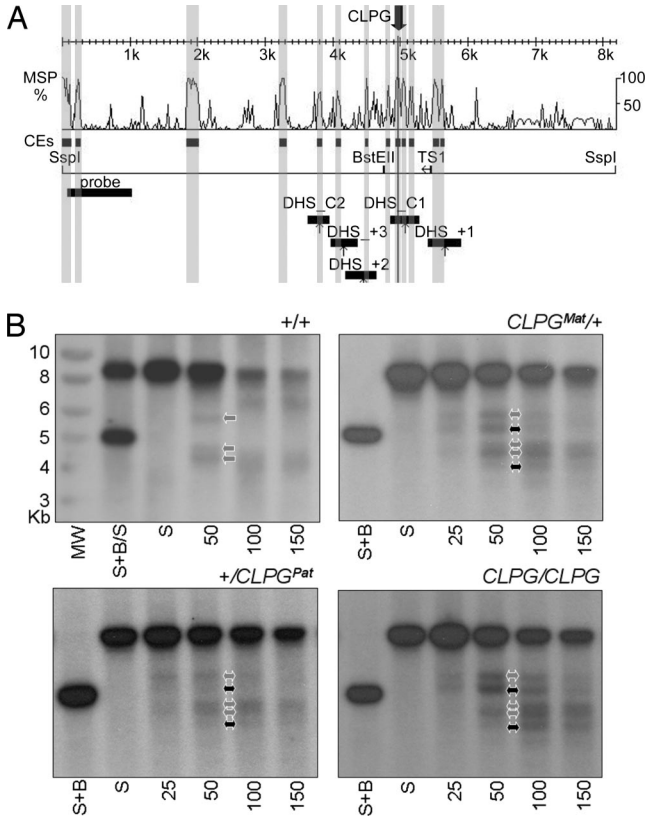

Fig. 3. DNase-I hypersensitive analysis. (*A*) Schematic representation of the 8.2-kb SspI fragment analyzed for the presence of DHS, showing (*i*) the position of the *CLPG* mutation (black arrow and vertical line), (*ii*) a multispecies (human, mouse, and ovine) similarity profile (MSP), (*iii*) ''*PhastCons*'' conserved elements (CEs) as obtained from http://genome.ucsc.edu, (*iv*) the position of the SspI, BstEII restriction sites, and the TS1 transcription start site identified in ref. 2, (*v*) the position of the probe used for Southern blot hybridization (black horizontal bar), and (*vi*) the position estimates of the constitutive (DHS_+x) and *CLPG*-specific (DHS_Cx) DHS. (*B*) Detection of DHS in nuclear DNA extracted from skeletal muscle of 8-week-old animals of the four possible *CLPG* genotypes. Purified nuclei were treated with increasing concentrations of DNase-I (25, 50, 100 and 150 units/ml), digested with SspI, and analyzed by Southern blot by using the probe shown in *A*. The same Southern blots included genomic DNA digested with SspI (S), SspI and BstEII (S+B), an equimolar mixture of both (S+B/S), and a molecular weight marker (MW). BstEII digests the 8.2-kb SspI fragment at 217 bp proximally from the *CLPG* mutation. Bands corresponding to the three DHS present on both the + and *CLPG* allele (DHS_+1, DHS_+2, and DHS_+3) are marked by gray arrows; bands corresponding to the two DHS that are specific for the *CLPG* allele (DHS_C1 and DHS_C2) are marked by black arrows.

gotes were compatible with a superposition of the + and *CLPG* patterns detected in the respective homozygotes. None of the DHSs were detectable in liver samples (data not shown). These results thus strongly suggest that the *CLPG* mutation uncovers allele- and tissue-specific DHS in cis.

Transcriptionally active chromatin is known to exhibit increased, general sensitivity to DNase-I (14). Because the *CLPG* allele enhances transcriptional activity in cis in skeletal muscle, the *CLPG* allele is predicted to be more sensitive to DNase-I than the + allele in this tissue. To test this hypothesis, we used PCR-restriction fragment length polymorphism to measure the *CLPG*-to-+ allelic ratio in DNA extracted from skeletal muscle and liver nuclei of a $C^{Mat}/+$ animal, incubated for increasing lengths of time with DNase-I. The *CLPG*-to-+ allelic ratio was clearly reduced in DNase-I-treated skeletal muscle nuclei when compared with genomic DNA extracted by using standard procedures (Fig. 6, which is published as supporting information on the PNAS web site). Note that the effect was apparent even after very short exposure to DNase-I and only modestly enhanced with increased incubation time. It suggests that the observed effect could be due to endogenous nucleases. There was no evidence at all for a comparable effect in liver, demonstrating its tissue specificity and a likely genuine correlation with transcriptional activity.

The *CLPG* Mutation Enhances Bidirectional Long-Range *DLK1-GTL2* IG Transcription in Cis. To follow up on the *CLPG1* findings of Freking *et al.* (2), we repeated strand-specific RT-PCR experiments encompassing the *CLPG* mutation by using skeletal muscle RNA extracted from sheep of the four possible *CLPG* genotypes at two development stages: 2 weeks prenatal and 8 weeks postnatal.

Confirming Freking's findings, we detected transcripts oriented toward *DLK1* [hereafter referred to as $D(lk1)\leftarrow G(tl2)$ transcripts] in +/+ fetuses, albeit at low levels. The same low level $D\leftarrow G$ transcripts also were detectable in 8-week-old +/+ animals. In addition, we obtained RT-PCR products corresponding to antisense $D\rightarrow G$ transcripts from the prenatal +/+ samples at extremely low levels (Fig. 4*A*).

More remarkably, when compared with +/+ animals, we observed a strong increase in the yield of $D\leftarrow G$ RT-PCR product from $C^{Mat}/+$, $+/C^{Pat}$, and C/C pre- and postnatal samples. A similar effect also was noticed for the $D\rightarrow G$ products, albeit more modest. Sequencing the corresponding $C^{Mat}/+$ and $+/C^{Pat}$ amplicons indicated that both $D\leftarrow G$ and $D\rightarrow G$ transcripts were preferentially transcribed from the *CLPG* allele (Fig. 4*A*). These results demonstrate that, in addition to its previously reported effect on the expression of distant imprinted genes, the *CLPG* mutation enhances bidirectional expression of *DLK1-GTL2* IG transcripts in skeletal muscle, irrespective of its parental origin.

To study the extent of this previously undescribed cis effect, we designed 11 amplicons spanning the *DLK1-GTL2* IG region (1′-11′ in Fig. 1). They were amplified from genomic DNA and random primed skeletal muscle cDNA from animals of the four *CLPG* genotypes and the same two developmental stages. All PCR products were sequenced, and SNPs for which the individuals were heterozygous used to determine the allelic origin of the corresponding transcripts. The obtained results can be summarized as follows (Fig. 4*B*):

1. In +/+ fetuses, low-level discontinuous transcription is detected throughout the *DLK1-GTL2* IG region. More specifically, we obtained RT-PCR products for amplicons 3′, 4′, and 5′, which jointly span ≈15 kb from the *CLPG* mutation toward *DLK1*, as well as with the terminally located amplicons 1′, 10′, and 11′. The transcripts were preferentially of paternal origin on the *DLK1* side, of maternal origin on the *GTL2* side, and biallelic in the center. Single-stranded RT-PCR experiments performed on amplicon 5′ (data not shown) indicated that both the $D\leftarrow G$ and $D\rightarrow G$ transcripts are biallelically expressed in +/+ fetuses.
2. In 8-week-old +/+ animals, IG transcription is further reduced, restricted to the central part, and monoallelic.
3. At 2 weeks before birth, $C^{Mat}/+$, $+/C^{Pat}$, and C/C animals show a strong enhancement of transcript levels throughout the *DLK1-GTL2* IG region. In $C^{Mat}/+$ animals, the IG transcripts are exclusively produced from the maternal *CLPG* allele. In $+/C^{Pat}$ animals, the transcripts are virtually exclusively produced from the paternal *CLPG* allele, except for the two amplicons nearest *GTL2* that show biallelic expression. The effect of the *CLPG* mutation seems most pronounced for the segment spanned by amplicons 3′-5′, which also were yielding higher amounts of PCR product in +/+ fetuses. Note that transcription proceeds throughout the IG-DMR in these

GENETICS

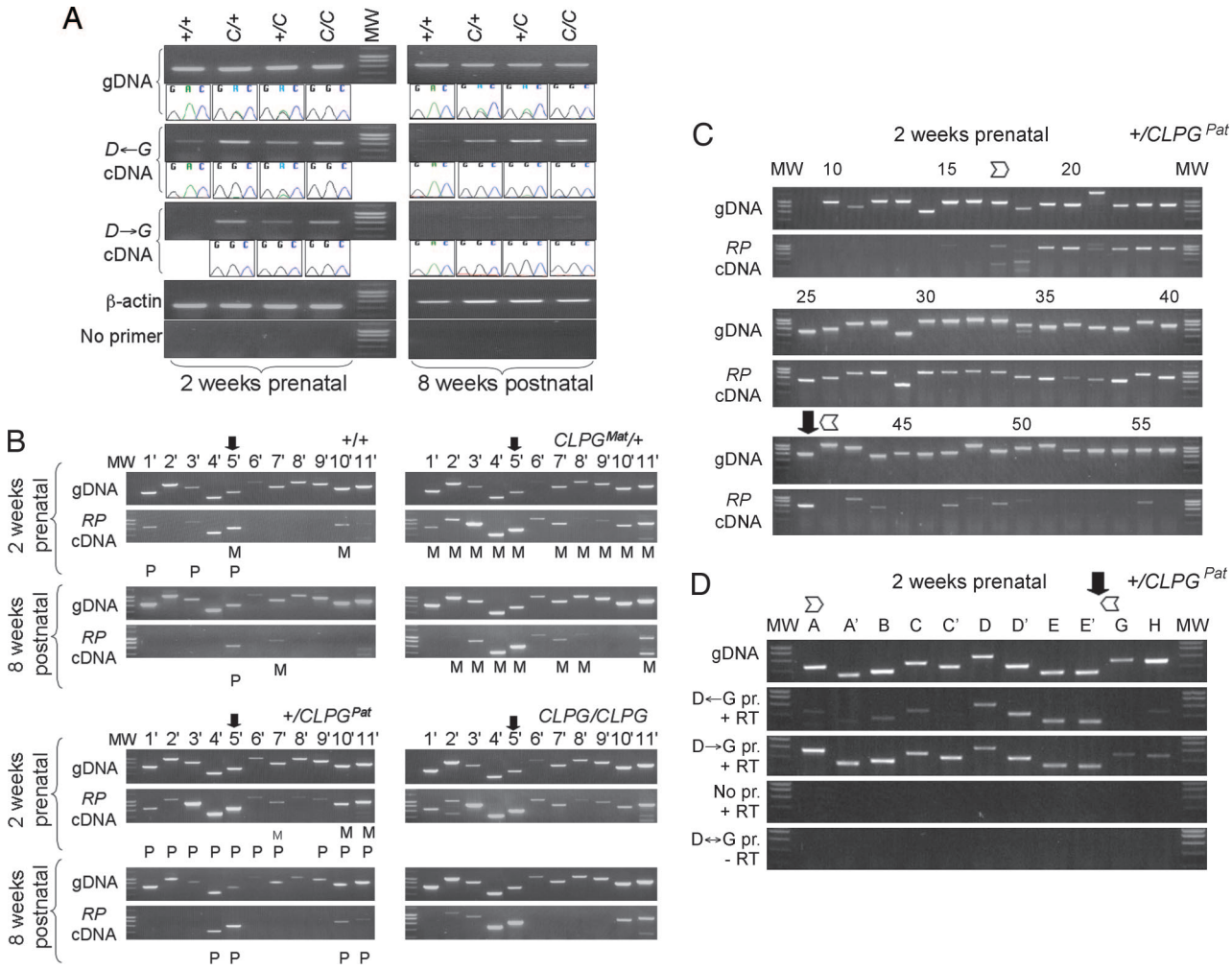

Fig. 4. Expression analysis of IG transcripts. The position of the *CLPG* mutation is marked by the black arrow. The white arrows correspond to the positions of the TS2 and TS1 transcription start sites, respectively. MW, molecular weight marker. (*A*) Results of strand-specific RT-PCR experiments by using a 593-bp amplicon spanning the *CLPG* site and *gluteus medius* RNA from animals of the four possible *CLPG* genotypes at 2 weeks before and 8 weeks after birth. The amplicon was amplified from the cognate genomic DNA extracted from skeletal muscle as positive control, *gluteus medius* cDNA synthesized by using either one (specific for *D←G* transcripts) or the other (specific for *D→G* transcripts) primer, and RT-treated cDNA in the absence of primers. A 428-bp β-actin amplicon was amplified to control for the quality of the RNA. The *CLPG* amplicons were directly sequenced; the portions of the electropherograms spanning the *CLPG* site are shown, revealing the preferential expression of the *CLPG* allele in $C^{Mat}/+$ and $+/C^{Pat}$ animals. (*B*) Results of RT-PCR experiments for 11 amplicons spanning the *DLK1-GTL2* IG region (1′–11′ in Fig. 1) by using random primed *gluteus medius* cDNA (*RP* cDNA) from animals of the four possible *CLPG* genotypes at 2 weeks before and 8 weeks after birth. Amplicon 5′ is marked by an arrow as it spans the *CLPG* site. The same amplicons were amplified from the cognate genomic DNA (gDNA) and randomly primed cDNA with or without RT. The latter were all negative and are not shown. The cDNA amplicons were directly sequenced, and SNP markers in the region were used to determine the parental origin of the transcripts when possible. Biallelically expressed amplicons are marked by both a maternal (M) and a paternal (P) of equal size. Preferential expression of one allele is reflected by the relative size of the corresponding symbols. Monoallelically expressed amplicons are marked by M or P if the allele is madumnal or padumnal, respectively. In the absence of informative polymorphisms, the amplicons are unlabeled. (*C*) Representative results of PCR experiments performed with 47 overlapping amplicons spanning a 32-kb IG segment (Fig. 1) by using genomic DNA (gDNA) and random primed gluteus medius cDNA (*RP* cDNA) from a $+/C^{Pat}$ fetus. Controls by using cDNA synthesized without RT were all negative and are not shown. (*D*) Results of strand-specific RT-PCR experiments performed with 11 amplicons, labeled A–H in Fig. 1 (i.e., A and A′ and B and B′, are distinct amplicons with virtually identical position) by using genomic DNA (gDNA), *gluteus medius* cDNA synthesized by using either one (specific for *D←G* transcripts) or the other (specific for *D→G* transcripts) primer, RT-treated cDNA in the absence of primers, and RT minus RNA with both primers.

animals, shown to operate as imprinting control element for the *DLK1-GTL2* domain (15).

4. The effect of the *CLPG* mutation on *DLK1-GTL2* IG transcription persists at 8 weeks of age, albeit attenuated. In $C^{Mat}/+$ and $+/C^{Pat}$ animals, expression is monoallelic and restricted to the mutant *CLPG* allele. Expression is concentrated in the central part in the vicinity of the *CLPG* SNP and distally from the IG-DMR.

To gain additional insight regarding the organization of the detected *DLK1-GTL2* IG transcripts, we performed further RT-PCR and RACE experiments by using skeletal muscle RNA from $+/C^{Pat}$ fetuses, i.e., a genotype and developmental stage showing pronounced IG activity. These experiments led to the following observations:

1. Random-primed RT-PCR products could readily be obtained for an uninterrupted chain of 24 overlapping amplicons that jointly span from ≈−15.8 kb to +138 bp from the *CLPG* site (18–41 in Fig. 4*C*). This chromosome segment is bounded on the *GTL2* side by TS1 and on the *DLK1* side by a strong *D→G* start site (TS2) detected in this work by 5′ RACE at position

−16,846 bp (Fig. 7, which is published as supporting information on the PNAS web site). Strand-specific RT-PCR experiments indicate that the transcription proceeds from both strands in this interval (Fig. 4*D*). The size of the RT-PCR products matched the genomic prediction for all amplicons in the interval, except amplicon 21, yielding a shorter fragment with splice-product compatible sequence.

2. IG transcripts were detected for 9 of 17 amplicons located between *DLK1* and position −16,846, and for 25 of 30 amplicons located between position +478 and *GTL2*, albeit at markedly lower levels than for the −16.846 to +478 interval (Fig. 4*C*; see also Fig. 8, which is published as supporting information on the PNAS web site). Strand-specific RT-PCR experiments performed on the *GTL2* side of the *CLPG* mutation (amplicons G and H) indicate that transcription proceeds primarily in the *D*→*G* direction in this region (Fig. 4*D*).

Discussion

We herein demonstrate that the mutant *CLPG* allele differentiates itself from the wild-type + allele by at least three epigenetic marks: DNA hypomethylation and the emergence of *CLPG*-specific DHS in the immediate vicinity of the mutation and enhanced bidirectional transcription throughout the *DLK1-GTL2* IG region.

Bisulfite sequencing revealed that *CLPG* alleles obtained from C/C or $C^{Mat}/+$ skeletal muscle were dominated by molecules with a distinct signature, being virtually completely unmethylated with the exception of two adjacent highly methylated CpG dinucleotides (Fig. 2). This pattern contrasted strikingly with the uniformly high level of methylation of + alleles as obtained from $+/+$ and $+/C^{Pat}$ samples. This hypomethylated population only accounts for ≈75% of the *CLPG* molecules in C/C and $C^{Mat}/+$ individuals. The remaining 25%, in essence, recapitulates the uniformly high methylation pattern typical of the + allele. A simple explanation of this bimodal behavior is tissue heterogeneity. The majority of molecules might originate from muscle tissue in which the mutation exerts its effect and the remainder from unaffected cell types.

It is noteworthy that the position of the two hypermethylated CpG sites in *CLPG* alleles coincides with the methylation peak of the + alleles. Moreover, sorting *CLPG* molecules by ascending methylation rate reveals a gradient emanating from these two adjacent sites. One interpretation is that this region acts as a nucleation site for methylation. Spreading of methylation from this nucleation site would be somehow hampered on the *CLPG* allele. In agreement with this conjecture, Murphy *et al.* (16) recently observed a postnatal acquisition of methylation in the vicinity of the *CLPG* site in skeletal muscle of $+/+$ but not of C/C animals.

Quite surprisingly, in heterozygous $C^{Mat}/+$ and $+/C^{Pat}$ animals, the methylation status of the maternal allele, whether C^{Mat} or $+^{Mat}$, recapitulates that of the corresponding allele in homozygotes. However, the paternal allele, whether $+^{Pat}$ or C^{Pat}, adopts an intermediate profile leaning toward the methylation status of its maternal homologue. If further confirmed, this observation might reveal the existence of a novel trans-sensing mechanism in the *DLK1-GTL2* domain in addition to the previously reported trans interaction between the products of reciprocally imprinted genes (10–12).

Because of its effect on the expression level of genes located within a large chromosomal domain, we hypothesized that the *CLPG* mutation might perturb a LCR element (6). So far, however, evidence supporting this hypothesis has been only indirect. The presence of DHS is typically considered pathognomonic for LCRs and other distant control elements (13). The identification of multiple tissue-specific DHS in the immediate vicinity of the *CLPG* mutation, and more specifically the demonstration of DHS that are unique for the *CLPG* allele, thus directly supports our hypothesis (Fig. 3).

We provide evidence that, when compared to the + allele, the *CLPG* allele exhibits an increase in general sensitivity to DNase-I in skeletal muscle, which is compatible with it adopting a more open, transcriptionally permissive chromatin configuration in this tissue (Fig. 6).

The latter observation is in good agreement with the most remarkable observation of this study, namely the fact that the *CLPG* mutation strongly enhances biallelic, long-range transcription throughout the *DLK1-GTL2* IG region (Fig. 4). IG transcription has been demonstrated for a number of LCRs, but its role has remained elusive (17–20). More recently, genome-wide approaches have revealed that IG transcription is much more widespread than initially suspected (4, 5), but, in these studies as well, the functional significance of these findings was difficult to apprehend. The callipyge phenomenon might offer a unique opportunity to study the role of noncoding IG transcripts.

Murphy *et al.* (16) recently reported results focusing on *D*←*G* transcript in the immediate vicinity of the *CLPG* mutation. Their most important message, namely that the *CLPG* mutation cis enhances *CLPG1* transcription in skeletal muscle, agrees with our findings (Fig. 4*A* and ref. 21). Minor differences include the fact that, in $C^{Mat}/+$ and $+/C^{Pat}$, we find preferential expression from the *CLPG* allele not only after (8 weeks) but also before (2 weeks) birth. This difference could be due to the fact that their fetuses were at an earlier stage of development, before down-regulation of IG transcription from the + allele. Other discrepancies are the fact that Murphy *et al.* (16) do not report the detection of antisense *D*→*G* transcripts, nor of *D*←*G* transcripts in adult $+/+$ animals. It is likely due to a difference in sensitivity between the PCR assays used.

The major difference between the two studies is the fact that we herein demonstrate that the *cis* effect of the *CLPG* mutation on IG transcription is not limited to its immediate vicinity but extends throughout the entire 90-kb *DLK1-GTL2* IG region. Unraveling the precise organization of the corresponding transcripts will require additional work, but the following statements can be made. The *CLPG* mutation enhances *D*←*G* transcription from TS1 and *D*→*G* transcription from TS2, generating long complementary transcripts that have the potential to form double-stranded RNA molecules. It is worthwhile noting that this segment overlaps, in part, with a region of enhanced transcriptional activity detected by microarray ("transfrag") analysis (5). *CLPG* chromosomes also produce IG transcripts on both sides of this central 17-kb segment. It remains uncertain, however, whether these transcripts are physically connected with those originating from the central segment or whether they are the products of independent initiations, and, in that case, from which strand they originate. The simplest model assumes that they are just extensions of the TS1 and TS2 transcripts. However, the detection of an additional *D*←*G* initiation site at position –19,683 bp (TS3) (Fig. 7), reports of multiple CAGE tags, and GSC ditags corresponding potentially to alternative transcription initiation sites throughout the region (4) hints toward a more complex transcript network.

It remains an open question whether the detected transcripts are just innocent bystanders or play an active role in mediating the effects of the *CLPG* mutation on its target genes. Some observations, however, are intriguing and, in our opinion, suggest an active function. The first is the fact that the IG transcription induced by the *CLPG* mutation seems to bridge the gaps between the mutation and at least two of its major targets: *DLK1* and *GTL2*. The IG transcripts, thus, might physically connect the mutation and the genes that are affected by it, thus directly mediating the effect of the mutation. It is very intriguing that a spliced EST and GSC ditags are actually directly connecting IG with *GTL2* sequences, as if some *GTL2* transcripts

actually are initiated within the IG region (Fig. 1). The second is the observation that enhanced transcription in the *DLK1-GTL2* IG region is an early event when compared with phenotypic expression. We demonstrated in this work that IG transcripts are more abundant 2 weeks before birth than 8 weeks after birth. This finding contrasts with the observation that ectopic expression of DLK1 protein, and, hence, expression of the callipyge phenotype is only manifest several weeks after birth (10).

Our results allow us to propose the following model (Fig. 9, which is published as supporting information on the PNAS web site). The *CLPG* mutation would inactivate a silencer element that normally operates in fetal skeletal muscle to control the level of *DLK1-GTL2* IG transcription. Increased IG transcription would alter the chromatin epigenotype throughout the region. This alteration, for instance, could be achieved by promoting the incorporation of variant histone molecules (22), by preventing PcG mediated silencing (23), by preventing regional spreading of DNA methylation, or by going through an RNA interference-dependent mechanism (24). This permissive chromatin status would be maintained epigenetically in skeletal muscle throughout development, promoting high-level transcription of the genes known to be influenced by the *CLPG* mutation. In $+/C^{Pat}$ animals, competence to translate the *DLK1* mRNAs would be acquired only in muscles of the hindquarters later in development, possibly as a result of the down-regulation of specific miRNAs or any other translational control. In C/C animals, this competence never would be acquired because of the persistent expression of anti-*DLK1* miRNAs from the C^{Mat} allele.

The recent generation of transgenic mice with a deletion of the dodecamer motif, which are recapitulating the callipyge phenomenon (D. Pirottin, M.G., and C.C., unpublished data), should facilitate testing of the hypotheses that result from this work. Easy access to tissue at multiple developmental stages, combined with tiling arrays of the region, will allow a more extensive characterization of the regional epigenotype and structure of the IG transcripts. Targeted mutagenesis of the IG transcripts combined or not combined with the dodecamer deletion will directly test their functional relevance.

Methods

Bioinformatic analyses were conducted as described in ref. 25. Bisulfite sequencing was performed by using the CpGenome DNA modification kit (Chemicon International, Temecula, California). Detection of DHS and probing general DNase-I sensitivity was performed by following, respectively, Gregory *et al.* (26) and Gregory and Feil (27) with some modifications. Random primed and strand-specific RT-PCR experiments were performed by using cDNA synthesized, respectively, with the SuperScript-III First-Strand Synthesis System (Invitrogen) and EndoFree RT Kit (Ambion, Austin, TX), and RNA treated with the Turbo DNA-free kit (Ambion) to remove contaminating genomic DNA. 5′ RACE was performed by using the GeneRacer kit (Invitrogen). Detailed descriptions of the used procedures and primer sequences are provided in *Supporting Text* and Table 1, which are published as supporting information on the PNAS web site.

This project was supported by Fund for Collective Fundamental Research Grant 2.4525.96; National Foundation for Scientific Research (FNRS) Crédit aux Chercheurs Grant 1.5.134.00; grants from the Crédit à la Recherche from the Université de Liège and the "GAME" Action de Recherche Concertée from the Communauté Française de Belgique; PAI P5/25 from the Belgian Ministry for Science, Technology, and Culture Grant R.SSTC.0135; grants from the European Union "Callimir" Specific Targeted Research Project and the Utah Center of Excellence Program; U.S. Department of Agriculture/National Research Initiative Competitive Grants Program Grants 94-04358, 96-35205, and 98-03455; and a grant from the Utah Agricultural Experiment Station, Utah State University. H.T. benefits from a European Union–Marie Curie Postdoctoral Fellowship. C.C. is chercheur qualifié from the FNRS.

1. Cockett, N. E., Jackson, S. P., Shay, T. L., Farnir, F., Berghmans, S., Snowder, G. D., Nielsen, D. M. & Georges, M. (1996) *Science* **273,** 236–238.
2. Freking, B. A., Murphy, S. K., Wylie, A. A., Rhodes, S. J., Keele, J. W., Leymaster, K. A., Jirtle, R. L. & Smith, T. P. (2002) *Genome Res.* **12,** 1496–1506.
3. Smit, M., Segers, K., Carrascosa, L. G., Shay, T., Baraldi, F., Gyapay, G., Snowder, G., Georges, M., Cockett, N. & Charlier, C. (2003) *Genetics* **163,** 453–456.
4. Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., *et al.* (2005) *Science* **309,** 1559–1563.
5. Cheng, J., Kapranov, P., Drenkow, J., Dike, S., Brubaker, S., Patel, S., Long, J., Stern, D., Tammana, H., Helt, G., *et al.* (2005) *Science* **308,** 1149–1154.
6. Charlier, C., Segers, K., Karim, L., Shay, T., Gyapay, G., Cockett, N. & Georges, M. (2001) *Nat. Genet.* **27,** 367–369.
7. Murphy, S. K., Freking, B. A., Smith, T. P., Leymaster, K., Nolan, C. M., Wylie, A. A., Evans, H. K. & Jirtle, R. L. (2005) *Mamm. Genome* **16,** 171–183.
8. Charlier, C., Segers, K., Wagenaar, D., Karim, L., Berghmans, S., Jaillon, O., Shay, T., Weissenbach, J., Cockett, N., Gyapay, G., *et al.* (2001) *Genome Res.* **11,** 850–862.
9. Seitz, H., Royo, H., Bortolin, M. L., Lin, S. P., Ferguson-Smith, A. C. & Cavaille, J. (2004) *Genome Res.* **14,** 1741–1748.
10. Davis, E., Jensen, C. H., Schroder, H. D., Farnir, F., Shay-Hadfield, T., Kliem, A., Cockett, N., Georges, M. & Charlier, C. (2004) *Curr. Biol.* **14,** 1858–1862.
11. Georges, M., Charlier, C. & Cockett, N. (2003) *Trends Genet.* **19,** 248–252.
12. Davis, E., Caiment, F., Tordoir, X., Cavaille, J., Ferguson-Smith, A., Cockett, N., Georges, M. & Charlier, C. (2005) *Curr. Biol.* **15,** 743–749.
13. Li, Q., Peterson, K. R., Fang, X. & Stamatoyannopoulos, G. (2002) *Blood* **100,** 3077–3086.
14. Stalder, J., Larsen, A., Engel, J. D., Dolan, M., Groudine, M. & Weintraub, H. (1980) *Cell* **20,** 451–460.
15. Lin, S. P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J. & Ferguson-Smith, A. C. (2003) *Nat. Genet.* **35,** 97–102.
16. Murphy, S. K., Nolan, C. M., Huang, Z., Kucera, K. S., Freking, B. A., Smith, T. P., Leymaster, K. A., Weidman, J. R. & Jirtle, R. L. (2006) *Genome Res.* **16,** 340–346
17. Gribnau, J., Diderich, K., Pruzina, S., Calzolari, R. & Fraser, P. (2000) *Mol. Cell.* **5,** 377–386.
18. Masternak, K., Peyraud, N., Krawczyk, M., Barras, E. & Reith, W. (2003) *Nat. Immunol.* **4,** 132–137.
19. Rogan, D. F., Cousins, D. J., Santangelo, S., Ioannou, P. A., Antoniou, M., Lee, T. H. & Staynov, D. Z. (2004) *Proc. Natl. Acad. Sci. USA* **101,** 2446–2451.
20. Dean, A. (2006) *Trends Genet.* **22,** 38–45.
21. Georges, M., Charlier, C., Smit, M., Davis, E., Shay, T., Tordoir, X., Takeda, H., Caiment, F. & Cockett, N. (2004) *Cold Spring Harb. Symp. Quant. Biol.* **69,** 477–483.
22. Henikoff, S. & Ahmad, K. (2005) *Annu. Rev. Cell Dev. Biol.* **21,** 133–153.
23. Schmitt, S., Prestel, M. & Paro, R. (2005) *Genes Dev.* **19,** 697–708.
24. Haussecker, D. & Proudfoot, N. J. (2005) *Mol. Cell. Biol.* **25,** 9724–9733.
25. Smit, M. A., Tordoir, X., Gyapay, G., Cockett, N. E., Georges, M. & Charlier, C. (2005) *Mamm. Genome* **16,** 801–814.
26. Gregory, R. I., Khosla, S. & Feil, R. (2001) *Methods Mol. Biol.* **181,** 269–284.
27. Gregory, R. I. & Feil, R. (1999) *Nucleic Acids Res.* **27,** e32.