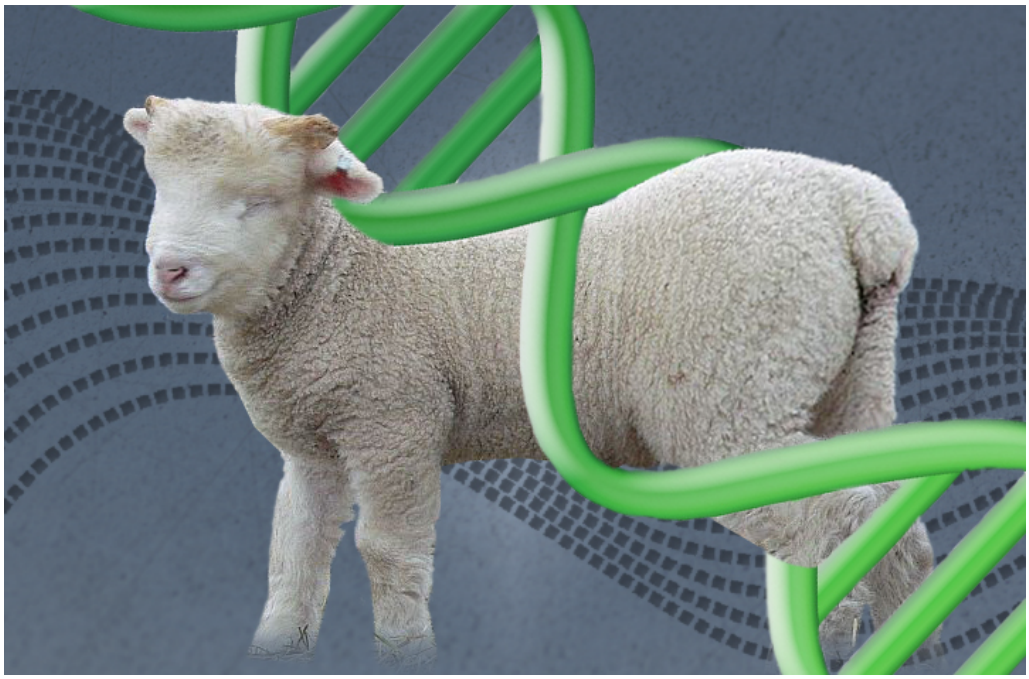
Académie universitaire Wallonie-Europe
Université de Liège
Faculté des sciences
Unité de Génomique Animal

Contribution à la compréhension du phénomène de surdominance polaire au locus callipyge du mouton



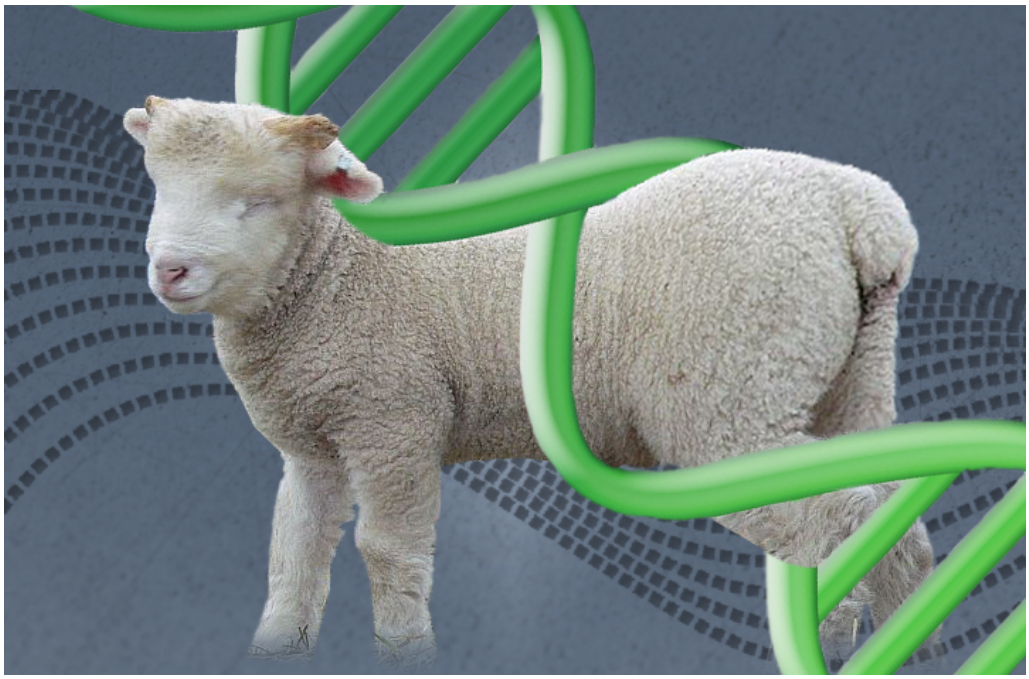
Par

CAIMENT FLORIAN

Thèse présentée en vue de l'obtention du doctorat en sciences
Collège de doctorat de Biochimie, Biologie moléculaire et cellulaire, Bioinformatique et modélisation en décembre 20010

Académie universitaire Wallonie-Europe
Université de Liège
Faculté des sciences
Unité de Génomique Animal

Contribution à la compréhension du phénomène de surdominance polaire au locus callipyge du mouton



Par

CAIMENT FLORIAN

Thèse présentée en vue de l'obtention du doctorat en sciences
Collège de doctorat de Biochimie, Biologie moléculaire et cellulaire, Bioinformatique et modélisation en décembre 2010

A mon père

Remerciements

En plus de six ans de thèse, la liste des personnes à remercier serait suffisamment longue pour nécessiter la parution d'un second tome. Je ne citerais donc que les personnes qui ont eu un rôle majeur dans la clôture de ce projet, tout en gardant en mémoire toutes les autres.

Merci tout d'abord à ma mère (Annita) et mon père (Norbert), qui se sont toujours sacrifiés pour financer mes études. Merci également à mes grands parents maternels (Ginette et Martial) pour leur aide quand le paiement des factures ne me laissait plus grand-chose à manger. J'aurais aimé que vous soyez encore tous là aujourd'hui.

Merci à Michel Georges de m'avoir donné la chance de faire un doctorat dans un cadre scientifique (et financier) idéal. Merci à Carole Charlier, chef du projet callipyge, d'avoir su me mettre sur les rails au début de ma thèse et de m'avoir soutenu les jours de tourmente.

Merci à Erica Davis et Haruko Takeda de m'avoir enseigné les techniques élémentaires de laboratoire. Je ne pouvais avoir de meilleurs professeurs en la matière.

Merci à ceux qui sont devenus de vrai(e)s ami(e)s au fil des années : Corinne, Nico, Mallory, Arno et Sarah. Travailler à vos côtés a toujours été un plaisir. Vous allez beaucoup me manquer.

Merci à ma compagne Pamela de m'avoir prêté ses talents pour la couverture, d'avoir cru en moi et de partager ma vie chaque jour.

Et je me dois de terminer par celui sans qui ma thèse n'aurait été qu'une longue traversée du désert : Denis Baurain. Merci d'avoir pris le temps pour m'apprendre les bases du métier de chercheur et d'avoir considéré mon travail avec autant d'attention.

Résumé de la thèse en français

THÈME DE RECHERCHE :

Le phénotype callipyge est une hypertrophie musculaire généralisée post-natale décrite chez le mouton (introduit dans le Chapitre 1). Son mode de transmission non mendélien, qualifié de surdominance polaire, est unique : seuls les hétérozygotes ayant reçu la mutation callipyge (*CLPG)* de leur père expriment le phénotype. La mutation *CLPG,* localisée dans un domaine soumis à l'empreinte parentale, est une mutation ponctuelle (SNP) détruisant un élément qui contrôle, en *cis,* le taux d'expression musculaire des gènes voisins. L'hypertrophie musculaire, sans doute causée par l'expression ectopique de la protéine DLK1 chez les animaux $+^{mat}/C^{pat}$, n'est pas observée chez les animaux C^{mat}/C^{pat} dont l'allèle maternel apparaît *trans*-inactiver la traduction du messager *DLK1*. Cette thèse a pour objectif d'approfondir notre compréhension de la surdominance polaire, aussi bien au travers de l'étude des mécanismes impliqués dans l'effet de contrôle à longue distance en *cis* de la mutation que de la caractérisation des inhibitions en *trans* induites par les transcrits maternels sur les messagers paternels.

RÉSULTATS :

L'étude de l'effet en *cis* (Chapitre 2) a démontré que la mutation *CLPG* se différenciait de l'allèle sauvage par au moins trois marques épigénétiques distinctes : (i) l'hypométhylation de l'ADN à proximité du SNP^{CLPG}, (ii) la création d'un site d'hypersensibilité à la DNase, et (iii) l'activation de la transcription bidirectionnelle de la région autour de la mutation. En démontrant que la mutation inactivait bel et bien un élément de contrôle à longue distance, ces données nous ont permis d'élaborer un modèle plus précis de la surdominance polaire.

L'étude de la *trans*-inhibition des transcrits protéiques à expression paternelle par les transcrits non codants maternels (Chapitres 3 et 4) se fondait sur l'hypothèse d'une implication des nombreux microARNs (miRNAs) du domaine *DLK1-GTL2*. Nous avons ainsi pu montrer que les cinq miRNAs abrités par *anti-PEG11* étaient capables, par leur parfaite complémentarité de séquence, de cliver le messager *PEG11* (Chapitre 3). Bien que le rôle de

la protéine PEG11 dans le phénotype callipyge soit inconnu, cette étude a néanmoins démontré l'existence d'une *trans*-inhibition entre les deux allèles du locus, tout en identifiant le premier cas de dégradation miRNA/cible impliquant des gènes soumis à l'empreinte chez les mammifères.

L'étude de la *trans*-inhibition appliquée au messager *DLK1* (Chapitre 4) a quant à elle nécessité la génération par séquençage haut débit d'un catalogue exhaustif des miRNAs du muscle squelettique ovin. Ce travail nous a permis de démontrer que les miRNAs du domaine *DLK1-GTL2* étaient bien exprimés maternellement et eux aussi soumis à l'effet en *cis* de la mutation. Si aucun miRNA capable d'inhiber le messager *DLK1* n'a été identifié sans ambiguïté, nos analyses d'affinité ont toutefois révélé un effet significatif de miR-376c sur sa 3'UTR, ainsi que de l'ensemble des miRNAs du locus DLK1-GTL2 sur sa région codante. Notons que nous avons également considéré comme potentiels candidats à la *trans*-inhibition les snoRNAs du locus, de même que les miRNAs du locus soumis à l'édition ARN.

CONCLUSIONS ET PERSPECTIVES :

À l'issue de cette thèse, nous avons donc contribué à la compréhension du phénomène de surdominance polaire au locus *DLK1-GTL2* du mouton callipyge. Si de nombreuses questions restent en suspend, elles devraient néanmoins trouver réponse grâce aux nouveaux outils d'analyse en cours de développement. Ainsi, deux lignées de souris transgéniques développées dans notre laboratoire, respectivement porteuses de la mutation *CLPG* ou surexprimant *PEG11* dans le muscle, éclaireront les mécanismes moléculaires de l'effet en *cis* de la mutation ainsi que le rôle de PEG11 dans l'établissement du phénotype callipyge. Par ailleurs, pour confirmer *in vivo* l'effet inhibiteur sur *DLK1* des miRNAs identifiés dans notre étude, les progrès apportés par le séquençage haut débit (notamment le HITS-CLIP) et la transfection en modèle cellulaire se révéleront sans doute d'une grande aide.

Résumé de la thèse en anglais

RESEARCH OVERVIEW:

The callipyge phenotype is a generalized muscular hypertrophy described in sheep (see Chapter 1). It features a non-mendelian mode of inheritance termed polar overdominance: only heterozygous animals having inherited the mutation from their father display the phenotype. The callipyge mutation (*CLPG*), which falls in an imprinted domain, is a single-nucleotide polymorphism (SNP) that disrupts a putative long-range control element affecting, in *cis*, the expression level of neighboring imprinted genes in skeletal muscle. The callipyge phenotype is thought to be caused by ectopic expression of DLK1 protein in $+^{mat}/C^{pat}$ animals. In contrast, C^{mat}/C^{pat} animals exhibit a wild-type phenotype that is certainly due to a *trans*-inhibition from the maternal allele on *DLK1* translation. The objective of our thesis was to improve our knowledge of polar overdominance, both by studying mechanisms of long-range *cis* regulation and by characterizing the *trans* effect of maternal noncoding transcripts on ~~the~~ paternally-expressed messenger RNAs.

RESULTS:

Our study of the *cis* effect (Chapter 2) showed that the *CLPG* mutation differs from the wild-type allele by at least three distinct epigenetic marks: (i) DNA hypomethylation in the vicinity of the SNP^{CLPG}, (ii) creation of a DNase-hypersensitive site, and (iii) activation of a bidirectional transcription start site centered on the mutation. Altogether, our data provided strong evidence for the SNP^{CLPG} inactivating a long-range control element and allowed us to refine our model of polar overdominance.

Our working hypothesis for the *trans*-inhibition of paternally-expressed genes by maternal noncoding transcripts (Chapters 3 and 4) involved microRNAs (miRNAs) from the *DLK1-GTL2* domain. In this respect, we showed that five miRNAs from *anti-PEG11* were able to cleave *PEG11* transcript, owing to perfect sequence complementarity (Chapter 3). This study was the first demonstration of miRNA-mediated RNAi involving imprinted genes in mammals. Furthermore, it allowed us to confirm the existence of a *trans*-inhibition between

both alleles in the domain, albeit the role of PEG11 protein in the callipyge phenotype is still unknown.

To study the *trans*-inhibition of *DLK1* messenger (Chapter 4), we used high-throughput sequencing to build an exhaustive catalogue of skeletal muscle-specific miRNAs in sheep. Our analyses showed that miRNAs from the *DLK1-GTL2* domain are maternally expressed and affected by the *cis* effect of the *CLPG* mutation. Even if we could not find miRNAs unambiguously able to repress *DLK1* transcript, affinity analyses revealed a significant effect on its 3' UTR for miR-376c, as well as on its coding region for all miRNAs considered as team. Of note, we also investigated snoRNAs and miRNAs subjected to RNA editing in the *DLK1-GTL2* locus.

CONCLUSIONS AND PERSPECTIVES:

During the course of this research, we contributed to improve the understanding of polar overdominance in the sheep *DLK1-GTL2* locus. Although many questions remain, most will eventually be answered thanks to upcoming analysis tools. Hence, our lab has already generated two transgenic mouse lines, either carrying the *CLPG* mutation or over-expressing *PEG11* in skeletal muscle. These mice should respectively shed light on the molecular mechanisms underlying the *cis* effect of the *CLPG* mutation and on the role of PEG11 in the callipyge phenotype. Finally, to confirm inhibiting effects of miRNAs identified in our study, technological improvements granted by high-throughput sequencing (such as HITS-CLIP) and miRNAs transfection in cell model systems will both prove to be very useful.

Tables des Matières

Liste des Figures

CHAPITRE 1

CHAPITRE 2

CHAPITRE 3

CHAPITRE 4

FIGURE S5 Sequence context of the polymorphic miRNA-*GDF8* interaction in sheep.

Liste des Tables

Liste des abréviations

3C	Chromosome Conformation Capture
A	adenosine
ADAM17	ADAM metallopeptidase domain 17
ADAR	Adenosine Deaminase Acting on RNA
ADN	acide désoxyribonucléique
AGO	argonaute
anti-PEG11	antisense paternally expressed gene 11
ARN	acide ribonucléique
ARNm	acide ribonucléique messager
ARNr	acide ribonucléique ribosomique
ARNt	acide ribonucléique de transfert
ATP	adenosine triphosphate
BAC	Bacterail artificial Chromosome
BEGAIN	brain-enriched guanylate kinase-associated protein
BTA	Bos taurus
BTRCP1	beta-transducin repeat containing
C	cytosine
°C	degree Celcius
CAF1	Chromatin Assembly Factor 1 subunit
CATS	Comparative Anchor Tagged Sequence
CFA	*Canis familiaris*
CLPG	callipyge
cM	centiMorgan
CRD-BP	c-myc mRNA coding Region Determinant-Binding Protein
DATS	dlk-1 associated trancripts
DHS	DNase-I Hypersensitive Sites
DIO3	iodothyronine deiodinase, type 3
DIO3as	DIO3 antisens

DLK1	delta-like 1
DLL1	delta 1
DLL2	delta 2
DLL3	delta 3
DMR	Région Différentiellement Méthylée
Dnmt1	DNA (cytosine-5-)-methyltransferase 1
Dnmt3a	DNA (cytosine-5-)-methyltransferase 3 alpha
Dnmt3b	DNA (cytosine-5-)-methyltransferase 3 beta
dNTP	déoxynucléotide triphosphate
d.p.c.	days post coitum
DSL	Delta Serrate Lage2
EBV	Epstein-Barr virus
EDTA	acide éthylènedinitrolotétraacétique
EGF	Epidermal Growth Factor
EMSA	electrophoretic mobility shift assay
ES	embryonic steam (souche embryonnaire)
EST	expressed sequence tags
FA-1	Fetal-Antigen-1
g	Gramme
G	guanine
GA-II	genome analyser 2 (séquenceur haut débit)
GATA1	GATA binding protein 1
GDF8	myostatine
GMBT	Gene Mark Bos Taurus
GO	gene ontology
GTL2	gene trap locus 2
HITS-CLIP	HIgh-Throughput Sequencing of RNAs from in vivo Cross-Linking and Immuno-Precipitation
I	inosine
IG	intergenic trancript

IG-DMR	intergenic differentially methylated region
Igf2	insulin-like growth factor 2
Igf2r	insulin-like growth factor 2 receptor
IMP-II	Igf2 mRNA-binding Protein 2
HOX	homeobox
HSA	homo sapiens
HTS	High-throughput sequencing
kb	kilobase(s)
KO	knockout
LCR	locus control region
LD	longissimus dorsi
LINE	long interspersed nuclear element
Lod	Logarith of Odd
LRCE	long range control element
LTR	long terminal repeat
M	molaire
MASA	marker-assisted segregation analysis
MatUPD12	disomie uniparentale maternelle du chromosome 12
Mb	mégabase
MAT	maternal
MEG3	maternally expressed gene 3
MEG8	maternally expressed gene 8
MIRG	microRNA containing gene
miRISC	microRNA RNA induced silencing complex
miRNA	micro ARN
min (ou ’)	minute(s)
ml	millilitre
mM	millimolaire
MMU	Mus musculus
MSTN	myostatine

MYOD1	myogenic differentiation 1
Nanog	Nanog homeobox
ncRNA	non coding RNA
ng	nanogramme
OAR	Ovis aries
OAR2	*Ovis aries* chromosome 2
OM	*ovum mutant*
ON	over night
ORF	open reading frame
PAPB	Poly(A)-Binding Proteins
PAGE	polyacrylamide gel electrophoresis
PAT	paternal
PatUPD12	disomie uniparentale paternelle du chromosome 12
P-bodies	processing bodies
PCR	Polymerase Chain Reaction
PEG1	paternally expressed gene 1
PEG3	paternally expressed gene 3
PEG11	paternally expressed gene 11
pg	picogramme
piRNA	piwi RNA
pre-miRNA	precursor microRNA
pri-miRNA	primary microRNA transcript
Pref-1	préadipocyte factor-1
QRT-PCR	quantitative RT-PCR (aloso known as Real time RT-PCR)
QTL	quantitative trait loci
QTN	quantitative trait nucleotides
RACE	rapid amplification of cDNA ends
Rian	RNA imprinted and accumulated in the nucleus
RISC	RNA induced silencing complex
RLM 5' RACE	RNA ligase mediated 5' rapid amplification of cDNA ends

RNA	ribonucleic acid
RNO	Rattus norvegicus
Rpm	rotation par minute
RT	**r**everse transcriptase
RTL1	retrotransposon-like 1
Scp-1	stromal cell protein-1
SDS	sel sodique de dodécyl sulfate
SILAC	Stable Isotope Labelling with Amino acids in Cell culture
siRNAs	small interfering ribonucleic acid
snoRNA	small nucleolar RNA
SNP	Single Nucleotide Polymorphism
STS	Sequence Tag Site
T	thymine
T3	triiodothyronine
T4	Thyroxine
TR	tandem repeat
TRA	tandem repeat A
TRB	tandem repeat B
TRC	tandem repeat C
Tris	Tris(hydroxyméthyl)-aminométhane
TS	transcription site
UTR	untranslated region
WARS	tryptophanyl-tRNA synthetase
WC	Waston-Crick
YY1	YY1 transcription factor
ZOG	zona glomerulosa-specific factor
µg	microgramme
µl	microlitre

CHAPITRE I

Introduction :

Il était une fois Callipyge

I. Le Phénotype callipyge

A. Apparition du phénotype

En 1983, dans une ferme d'Oklahoma, venait au monde Solid Gold, un agneau de race Dorset d'apparence tout à fait semblable à ses congénères. Cependant, il allait rapidement manifester un développement musculaire supérieur à la moyenne, essentiellement prononcé au niveau de son arrière-train, et associé à une diminution de masse graisseuse. Ce phénotype, que Solid Gold allait transmettre à une partie de sa descendance, attira l'attention des producteurs de viande ovine et des biologistes. La localisation particulière de l'hypertrophie musculaire poussa Michel Georges à baptiser ce phénotype « Callipyge » (du grec καλι (kallos) : beauté et πιγε (pûge)) : fesse.

B. Description du phénotype

Le phénotype callipyge se caractérise chez le mouton par une hypertrophie musculaire généralisée survenant de 4 à 8 semaines après la naissance. Cette manifestation postnatale, contrairement à d'autres phénotypes d'hypertrophie musculaire (notamment la vache blanc bleue belge), a l'avantage d'éviter les complications à la mise bas. L'augmentation musculaire est en moyenne de 27.6 % et suit un gradient rostro-caudal. Ainsi, si la masse de certains muscles postérieurs des moutons callipyges peut présenter une augmentation pouvant atteindre jusqu'à 63 %, l'hypertrophie des muscles des membres thoraciques n'excède pas 15 % (Jackson et al. 1997). Ce gradient rostro-caudal est caractéristique du phénotype callipyge et indépendant du contexte génétique (Freking et al. 1998b)

Carpenter et al. (1996) ont étudié la morphologie des fibres musculaires des animaux callipyges et ont découvert que l'hypertrophie est surtout due à une double augmentation du diamètre et du nombre des fibres glycolytiques rapides (type IIb) (cf. Figure 1).

L'hypertrophie musculaire n'est pas la seule caractéristique remarquable du phénotype callipyge. En effet, malgré cette hypertrophie, la masse totale d'un animal

callipyge à l'abattoir est similaire à celle d'un mouton de type sauvage. Outre la diminution des tissus graisseux, de nombreux organes du mouton callipyge présentent une réduction significative ($p<0.1$), tels que les poumons, le foie et les reins (Koohmaraie et al. 1995).

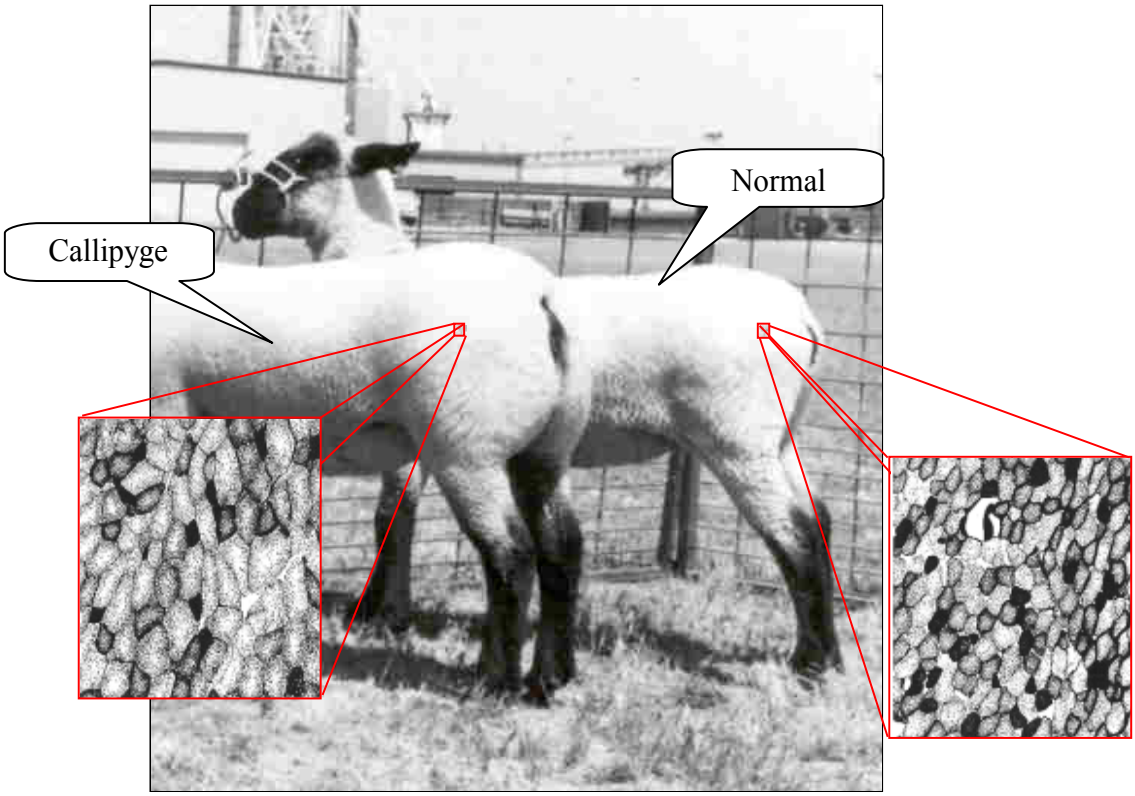


Figure 1 : Le phénotype callipyge. Hypertrophie musculaire d'un mouton callipyge (droite) comparé à un mouton de phénotype sauvage (gauche). Les coupes histologiques correspondantes illustrent l'augmentation de la proportion et du diamètre des fibres à contraction rapide chez l'individu callipyge. Les fibres oxydatives lentes (type I) sont noires, les fibres oxydatives rapides (type IIA) sont gris foncé et les fibres glycolytiques rapides (type IIb) sont en gris clair.

C. Déception économique

Le mouton callipyge semblait doté de toutes les caractéristiques pour révolutionner la production de viande ovine : plus de fibres musculaires pour la même quantité de nourriture, moins de gras (donc moins de cholestérol), un coût d'abattage réduit... Certains économistes américains prévoyaient déjà le gain potentiel que pouvait apporter le mouton callipyge à l'industrie agricole (Busboom et al. 1999).

Hélas, la viande du mouton callipyge se révéla rapidement incapable de satisfaire aux exigences des consommateurs. Son inconvénient majeur réside dans sa faible tendreté, la chair étant beaucoup plus coriace que celle d'un mouton de type sauvage. À cela, il faut en outre ajouter une faible jutosité et une saveur moindre.

Taylor et Koohmaraie (1998) ont démontré que cette diminution de tendreté était directement liée au taux de protéines myofibrillaires dans le muscle post-mortem. Chez un mouton de type sauvage, les protéines myofibrillaires sont progressivement dégradées après la mort, ce procédé étant majoritairement assuré par les calpaïnes, une famille de protéases calcium-dépendantes, inhibées par les calpastatines. Or, il est connu, grâce à Koohmaraie (1995), que les muscles des moutons callipyges présentent un taux post-mortem de calpastatines supérieur de 82.6 % comparé au phénotype normal (confirmé par Delgado et al. 2001). Il en résulte donc une importante inhibition des calpaïnes et une faible tendreté de la viande, et ce, en dépit d'un taux de protéines myofibrillaires normal au moment de l'abattage.

De nombreux auteurs tentèrent alors de proposer des solutions à ces problèmes. Ainsi, l'équipe de Clare et al. (1997) montra qu'une injection post-mortem de chlorure de calcium permettait de retrouver une tendreté normale. D'autres méthodes, comme une congélation précédant la rigidité cadavérique (pour contrer l'effet des calpastatines) ou encore des traitements électriques, donnèrent des résultats satisfaisants (Koohmaraie et al. 1998). Cependant, même si certaines équipes cherchent encore des alternatives, notamment différentes méthodes de cuisson (Shackelford et al. 2004 ; Everts et al. 2010), le mouton callipyge n'est toujours pas sérieusement envisagé pour la production de viande ovine. Si Solid Gold n'était pas une aubaine pour les producteurs agricoles, il allait toutefois s'avérer un formidable modèle pour les généticiens, grâce au mode de transmission unique du phénotype callipyge : la surdominance polaire.

II. La génétique et callipyge

A. Une transmission non mendélienne

La descendance directe de Solid Gold et les premiers croisements de ses fils laissèrent tout d'abord penser que le phénotype callipyge impliquait un seul gène autosomal non récessif soumis à un mode de transmission mendélien (cf Figure 2.1). En effet, deux béliers callipyges furent utilisés pour générer 200 agneaux dont 97 (48.5 %) présentaient l'hypertrophie musculaire après quelques semaines, indépendamment du sexe. Le symbole *CLPG* fut alors proposé pour caractériser l'allèle mutant responsable du phénotype callipyge (Cockett et al. 1994). A cette époque, rien ne semblait indiquer que la génétique de la mutation callipyge deviendrait une aubaine pour les scientifiques.

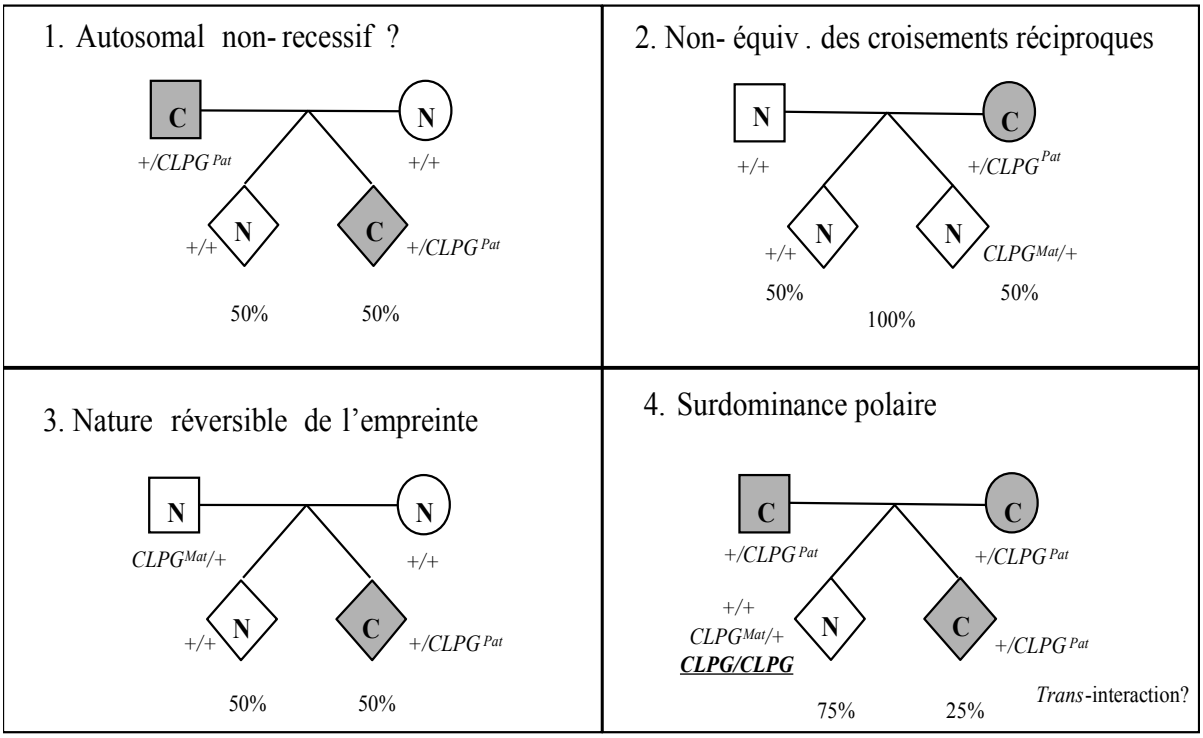


Figure 2 : La surdominance polaire associée au phénotype callipyge. Les quatre types de croisements ayant abouti à la définition du phénomène de surdominance polaire sont illustrés. Les animaux présentant le phénotype callipyge (C) sont représentés en gris. La mutation callipyge est dénotée $CLPG^{Pat}$ ou $CLPG^{Mat}$ en fonction de son origine parentale paternelle ou maternelle.

Une première observation vint rapidement réfuter le modèle monogénique mendélien. En effet, le croisement de brebis présentant l'hypertrophie musculaire avec des béliers sauvages aurait dû donner un taux de transmission de 100 % du phénotype dans le cas d'un gène autosomal non récessif si la mutation est à l'état homozygote (ou de 50 % à l'état hétérozygote). Or, tous les descendants de ce croisement se révélèrent de phénotype normal (cf Figure 2.2). Cette « polarité » paternelle de la transmission indiquait que la

mutation *CLPG* était en fait soumise à l'influence de l'origine parentale (cf Encadré 1 sur l'empreinte parentale). Conformément à ce nouveau modèle, un bélier ayant reçu la mutation *CLPG* de sa mère (et donc de phénotype normal) donnait à nouveau des animaux callipyges dans sa descendance (cf Figure 2.3). La mutation *CLPG* pouvait donc bien être réactivée après le passage dans la lignée germinale mâle (réversibilité caractéristique de l'empreinte parentale).

Le « simple » modèle d'empreinte parentale allait cependant se montrer incapable d'expliquer à lui seul tous les cas observés. Ainsi, alors que le croisement d'un bélier callipyge et d'une brebis hétérozygote porteuse de la mutation (mais de phénotype normal) aurait dû donner 50 % de descendants callipyges (les animaux ayant reçu la mutation *CLPG* de leur père, indépendamment de l'allèle transmis par la mère), seuls 25 % présentaient l'hypertrophie musculaire (cf. Figure 2.4). Curieusement, les animaux homozygotes pour la mutation ne manifestaient pas le phénotype callipyge, suggérant la possibilité d'une trans-interaction des deux allèles parentaux (voir plus loin sur l'effet en *trans* de la mutation).

B. Découverte de la région d'intérêt

1. Cartographie de la région d'intérêt

Les premières données génétiques furent publiées dès 1994 par Cockett et al. À cette époque, très peu d'informations étaient disponibles sur le génome ovin. Les auteurs parvinrent cependant à contourner ce problème en utilisant des marqueurs bovins. Deux populations ovines, issues des croisements de deux béliers présentant l'hypertrophie musculaire (pour un total de 172 animaux), furent utilisées. Par chance, sur les 4 marqueurs microsatellites utilisés, *GMBT16* avait un Lod score combiné sur les différentes populations de 8.6 à 20 % de recombinaison, ce qui le liait à la mutation *CLPG*. Or, des expériences d'hybridation *in situ* avaient déjà montré que ce marqueur se situait en région terminale du chromosome 21 bovin, elle-même orthologue à la région télomérique du chromosome 18 ovin (Georges et al. 1991). De plus, des marqueurs du chromosome 21 bovin indiquaient qu'en dépit d'une distance variable entre chaque marqueur, la synténie était globalement conservée entre B*os taurus* et *Ovis aries*. Comme des analyses de liaison situaient le locus

CLPG à 3 cM du marqueur *CSSM18* (soit environ 3 Mb ;région confirmée par Freeking et al., 1998a), on envisagea alors la construction d'un contig à partir de ce marqueur afin de séquencer le locus callipyge.

Ce projet fut achevé en 2000 par Segers et al. Afin de réaliser un contig ovin couvrant le locus *CLPG*, les auteurs utilisèrent 16 STS (sequence tagged sites) bovins pour cribler par PCR une librairie de BACs ovins. Les 21 BACs positifs permirent de générer de nouveaux STS, spécifiques au mouton cette fois. Il ne restait alors plus dans la région d'intérêt que deux trous qui purent être comblés par des PCR à longue distance. Ce contig de BACs ovins, d'une longueur d'environ 900 kb, constituait un formidable outil pour le clonage positionnel et la caractérisation de la mutation *CLPG*. Grâce à l'ajout de huit nouveaux marqueurs microsatellites dans la région, une cartographie fine par analyse de déséquilibre de liaison réalisée par Berghmans et al. (2001) permit de réduire la zone candidate à moins de 400 kb (entre les marqueurs *MULGE5* et *OY3*).

Les premières données semblaient indiquer que cette région était assez riche en gènes puisque la fréquence de clivage de l'enzyme de restriction NotchI, spécifique aux îlots CpG, était supérieure à la moyenne (une fois tous les 126 kb contre 285 kb). Il fut également noté que cette région présentait un taux de recombinaison supérieur chez les mâles (6.7 cM contre 1.4 cM chez les femelles). Cette différence, caractéristique des régions soumises à l'empreinte, concordait parfaitement avec le modèle de surdominance polaire.

Par analogie avec la région orthologue humaine, située dans la zone télomérique du chromosome 14, Shay et al. (2001) développèrent des CATS (pour Comparative Anchor Tagged Sequences) pour amplifier quatre gènes bovins du locus impliqué dans callipyge : *YY1*, *GTL2*, *DLK1* et *WARS*. Fait intéressant, deux d'entre eux avaient été récemment identifiés comme soumis à l'empreinte : *GTL2* exprimé depuis l'allèle maternel, et *DLK1* depuis l'allèle paternel (Schmidt et al. 2000 ; Takada et al. 2000). Le rôle de ces gènes sera abordé plus tard.

2. Séquençage du locus callipyge

D'après la cartographie fine, la mutation *CLPG* était donc située dans une région inférieure à 400 kb contenant deux gènes connus et soumis à l'empreinte. Le séquençage de

la région totale devenait donc pertinent. Charlier et al. (2001 b) séquencèrent deux BACs complets, ainsi qu'un produit de PCR à longue distance reliant les deux, pour un total d'environ 250 kb. L'alignement de la séquence obtenue confirma l'orthologie avec la région télomérique du chromosome 14 humain, tandis que la présence des gènes *DLK1* et *GTL2* fut démontrée par prédiction informatique (grâce au logiciel GENSCAN) et par alignement des séquences génomiques avec celles des banques de données d'EST (Expressed Sequence Tags). Par ailleurs, quatre nouveaux gènes potentiels furent prédits : *PEG11*, son antisens *anti-PEG11*, *MEG8* et *DATS* (voir plus loin).

Charlier et al. (2001a) réalisèrent également les premières expériences d'expression génique liées à la mutation callipyge. Ainsi, ils démontrèrent par RT-PCR que tous ces transcrits s'exprimaient dans le tissu musculaire squelettique d'un mouton callipyge âgé de huit semaines. Après avoir identifié au moins un nucléotide polymorphique dans chaque transcrit, ils furent en outre capables de démontrer par séquençage des produits de RT-PCR que tous ces transcrits étaient soumis à l'empreinte dans le muscle squelettique du mouton, *DLK1*, *DATs* et *PEG11* présentant une expression paternelle, alors que *GTL2*, *AntiPEG11* et *MEG8* étaient transcrits maternellement.

C. La surdominance polaire

La mutation callipyge était donc soumise à un mode de transmission nommée la surdominance polaire, dans lequel seuls les individus hétérozygotes ayant reçu la mutation de leur père exprimaient la mutation (cf Figure 3). Bien qu'étant un phénomène exceptionnel, deux autres cas assez similaires avaient déjà été décrits.

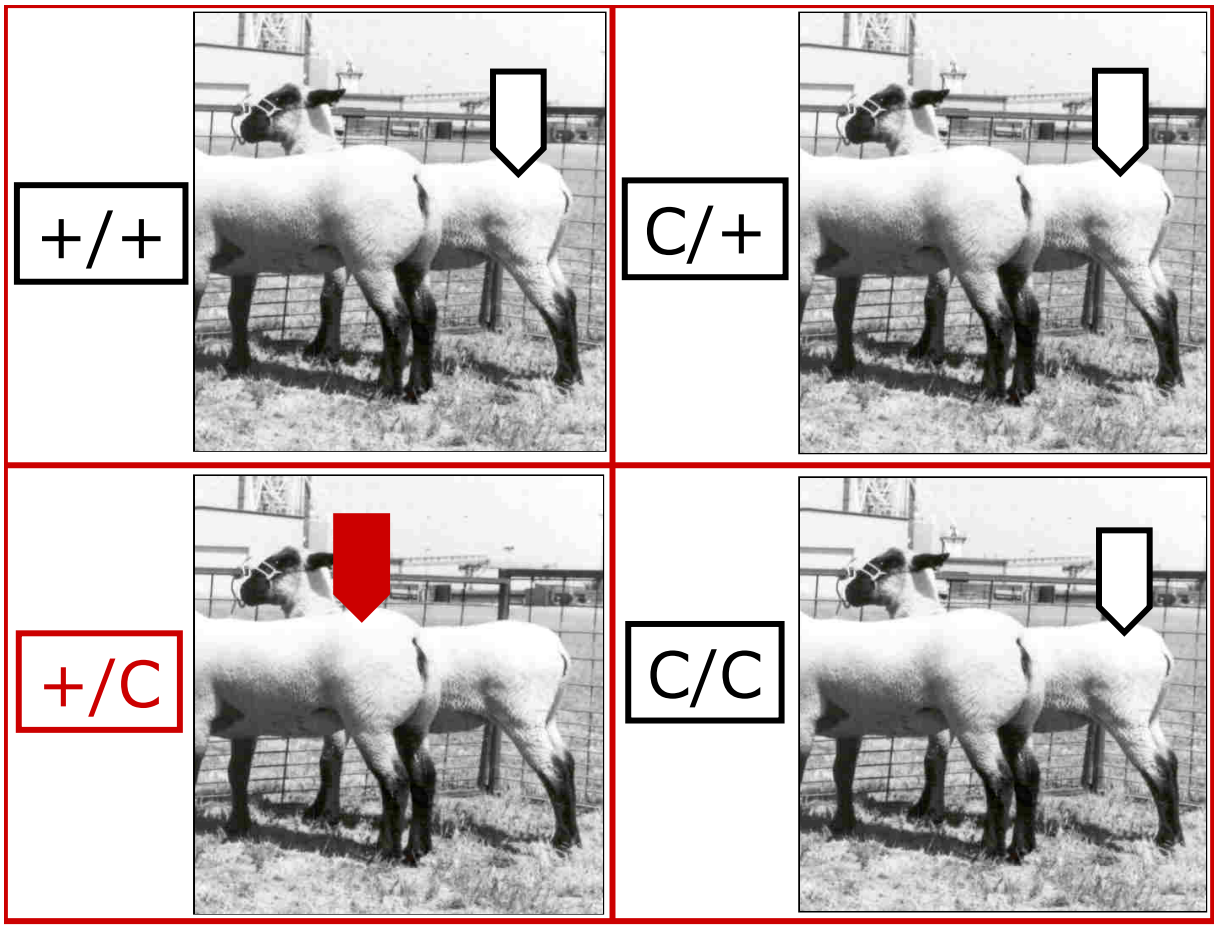


Figure 3 : Illustration de la surdominance polaire. L'allèle sauvage est indiqué par un +, l'allèle callipyge par un C. Par convention, on donne toujours en premier l'allèle maternel. Sur les quatre génotypes possibles, seul l'individu hétérozygote ayant reçu la mutation de son père présentera le phénotype callipyge (en rouge)

Le premier d'entre eux, appelé le syndrome DDK, fut rapporté dans les années soixante par un chercheur japonais du nom de Wakasugi, qui y consacre toujours ses travaux aujourd'hui (Wakasugi et al. 1967 ; Wakasugi 2007). Tout comme le phénotype callipyge, le syndrome DDK se caractérise par une polarité puisque le croisement de souris femelles de lignée DDK avec une autre lignée murine (non-DDK) est totalement stérile, alors que le croisement réciproque de mâles DDK avec des femelles d'une autre lignée produit des descendants parfaitement viables. Wakasugi explique la stérilité des femelles DDK par une mort embryonnaire résultant d'une incompatibilité entre un facteur cytoplasmique présent dans les oocytes DDK et d'un facteur présent dans les spermatozoïdes des autres lignées murines (Wakasugi 1974). Ces facteurs sont localisés dans un locus appelé *OM* (*ovum*

mutant) du chromosome 11 murin, mais malgré l'ancienneté de la découverte du syndrome DDK, leur véritable nature demeure mystérieuse. Renard et al. (1994) ont cependant déterminé que le facteur femelle est un ARN cytoplasmique, tandis que le facteur mâle a récemment été localisé avec plus de précision, par Bell et al. (2006), dans un cluster de gènes du chromosome 11 murin. Les mécanismes génétiques responsables de ce phénomène sont encore au stade des hypothèses, Wakasugi postulant dans une revue récente que l'instabilité de l'embryon serait due à un défaut de régénération du centrosome à chaque nouveau cycle cellulaire (Wakasugi 2007).

Le second fut également décrit dès le début des années soixante par Hiraizumi et Crow (1960). Nommé la dysgenèse hybride et affectant la drosophile, on y retrouve à nouveau la polarité caractéristique de la surdominance polaire. Ainsi, le croisement de femelles de lignée P avec des mâles de lignée M est parfaitement viable alors que le croisement réciproque est totalement létal. Cette létalité est due à la présence d'un élément transposable, appelé l'élément P, présent en de nombreuses copies dans la lignée P. Cette lignée a développé un inhibiteur de la transposase de l'élément P, qui se trouve notamment dans le cytoplasme des oocytes de la lignée P. Les femelles de la lignée M en étant dépourvues (tout comme les spermatozoïdes de la lignée M), les éléments P sont libres de transposer et de créer des dommages irréversibles au zygote (cf Figure 4). Depuis fin 2008, la nature de l'inhibiteur est connue : il s'agit de piwi ARNs (ou piRNAs, Aravin et al. 2007 pour revue), de petits ARNs synthétisés par les protéines PIWI, que l'on trouve uniquement dans les lignées de cellules germinales. Ces piRNAs bloquent la transposition par dégradation et clivage du messager de l'élément P, empêchant ainsi sa multiplication dans les cellules germinales (Brennecke et al. 2008).

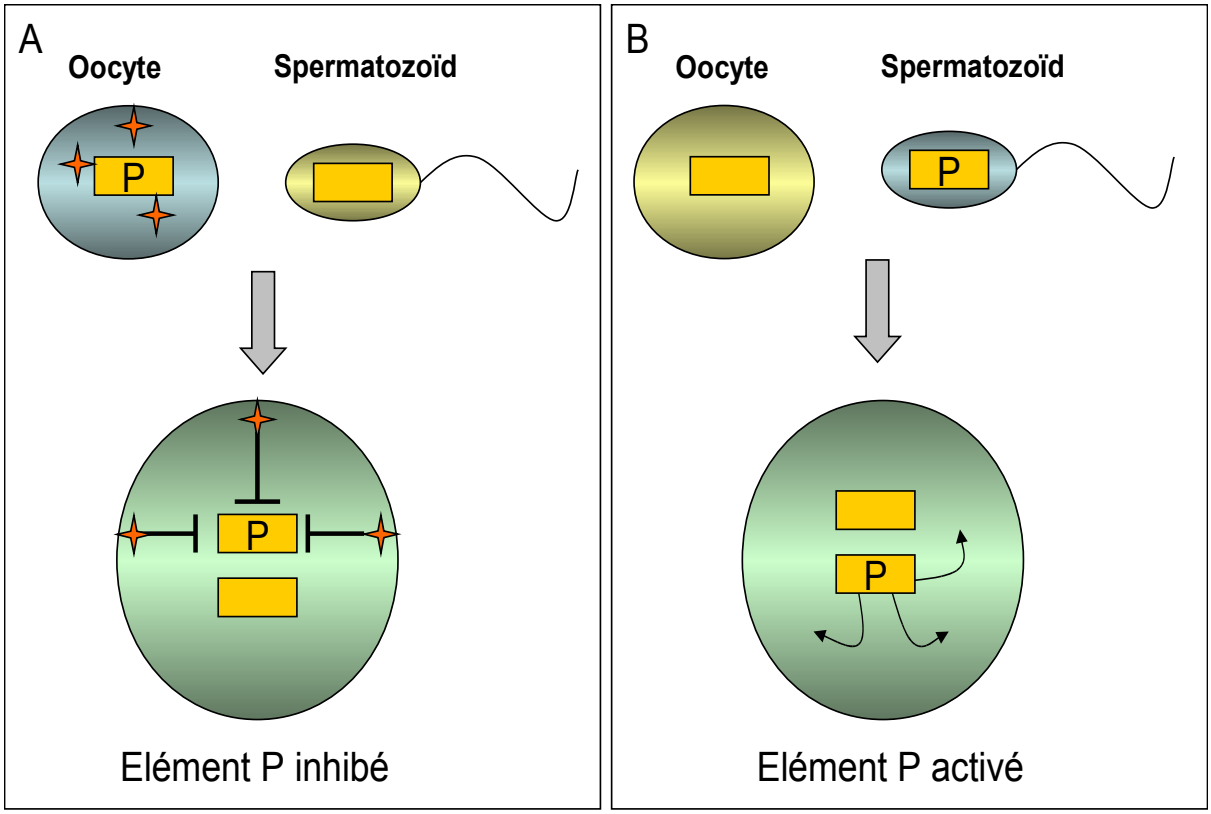


Figure 4 : Illustration de la dysgénèse hybride. **A**. Croisement d'une femelle de lignée P avec un mâle de lignée M. La transposition des élements P est inhibée par des facteurs présents dans le cytoplasme (etoiles rouges) des oocytes de la lignée P (en bleu). **B**. Croisement réciproque. Ces facteurs sont absents dans le cytoplasme de la lignée M (en jaune), le zygote résultant de ce croisement dégénérera rapidement par suite de la transposition des élements P.

Au moment de la description de la surdominance polaire chez le mouton callipyge, Cockett et al. (1996) n'avaient donc que ces deux exemples pour tenter d'apporter une explication aux phénomènes observés. La transposition et l'existence de facteurs cytoplasmiques semblaient difficilement pouvoir rendre compte de la spécificité musculaire de l'hypertrophie des moutons callipyge. Les auteurs envisagèrent alors l'implication de gènes soumis à l'empreinte parentale (cf Encadré 1).

Encadré 1 : L'empreinte parentale

Chez les organismes diploïdes, chaque individu reçoit deux copies de chaque gène : une de sa mère et l'autre de son père. Jusqu'au début des années 80, on pensait qu'à chaque locus, chacun des deux allèles était transcrit et fonctionnait de façon équivalente. Les premiers tests sur les embryons de souris démontrèrent que les embryons gynogénotes (possédant deux génomes femelles) étaient pourvus d'un placenta atrophié alors que les embryons androgénotes (deux génomes mâles) présentaient un développement embryonnaire très pauvre. Les deux copies du génome apportées par chaque parent n'étaient donc pas réciproquement fonctionnelles (Barton et al. 1984 ; McGrath and Solter 1984). La génération de souris à diploïdie uniparentale révéla l'existence de régions « soumise à l'empreinte », où un seul allèle était transcriptionnellement actif. Aujourd'hui, on estime que 600 gènes sont soumis à l'empreinte chez la souris (dont 64 % à expression maternelle), répartis sur une centaine de loci différents (Luedi et al. 2005). Une cartographie des domaines (ou « clusters ») dont la diploïdie uniparentale possède un phénotype est visible sur le site http://www.mgu.har.mrc.ac.uk/research/imprinting/largemap.html. Il est notable que la majorité des gènes de ces régions soit impliquée dans le développement embryonnaire ou placentaire.

De plus, l'empreinte génétique est spécifique aux mammifères placentaires. Dès lors, différentes théories ont été avancées pour expliquer l'avantage qui pourrait être conféré par l'expression monoallélique de certains gènes (a priori plutôt défavorable en cas de mutation). L'hypothèse la plus généralement acceptée est la « théorie du conflit parental » (Haig 2004; Haig and Wilczek 2006). Elle postule que l'établissement de l'empreinte est dû aux intérêts différents de chacun des parents. Ainsi, pour propager sa descendance, le père favorisera la transmission de gènes de croissance rapide, même aux dépens de la mère (la majorité des mammifères étant polygames, il pourra ensuite changer de femelle). La femelle, en revanche, cherchera à ménager ses ressources pour sa propre survie, et ce, afin de pouvoir assurer un plus grand nombre de portées. Les fonctions des gènes soumis à l'empreinte sont souvent en accord avec cette théorie : le père apporte des facteurs de croissance qui sont inhibés par des gènes à expression maternelle. Une autre hypothèse consiste à dire que l'empreinte est apparue dans le but de réduire l'expression des gènes étrangers. En effet, de nombreux gènes soumis à l'empreinte sont dérivés de rétrotransposons (d'origine virale) tandis que Chai et al. (2001) ont montré qu'un rétrotransposon inséré à proximité d'un gène soumis à l'empreinte acquerrait rapidement l'état d'empreinte.

Au niveau de la séquence primaire de l'ADN, rien ne distingue l'allèle paternel de l'allèle maternel. L'état d'empreinte repose donc sur des marqueurs épigénétiques qui sont effacés puis rétablis à chaque génération (hormis pour quelques domaines qui résistent à l'effacement gamétique). Pour la majorité des gènes, on sait aujourd'hui que le principal marqueur épigénétique de l'empreinte est la méthylation de l'ADN (particulièrement des cytosines des dinucléotides CpG) au niveau de zones différentiellement méthylées (régions DMR), elle-même accompagnée de diverses modifications des histones. L'établissement de l'empreinte se fait en deux étapes. Les marqueurs épigénétiques sont d'abord tous effacés au stage germinal primordial puis rétablis de novo par des méthyltransferases (*Dnmt3a* et *Dnmt3b*) en accord avec le sexe du futur gamète (cf Figure 5). Le maintien somatique de la méthylation est quant à lui assuré par *Dnmt1*.

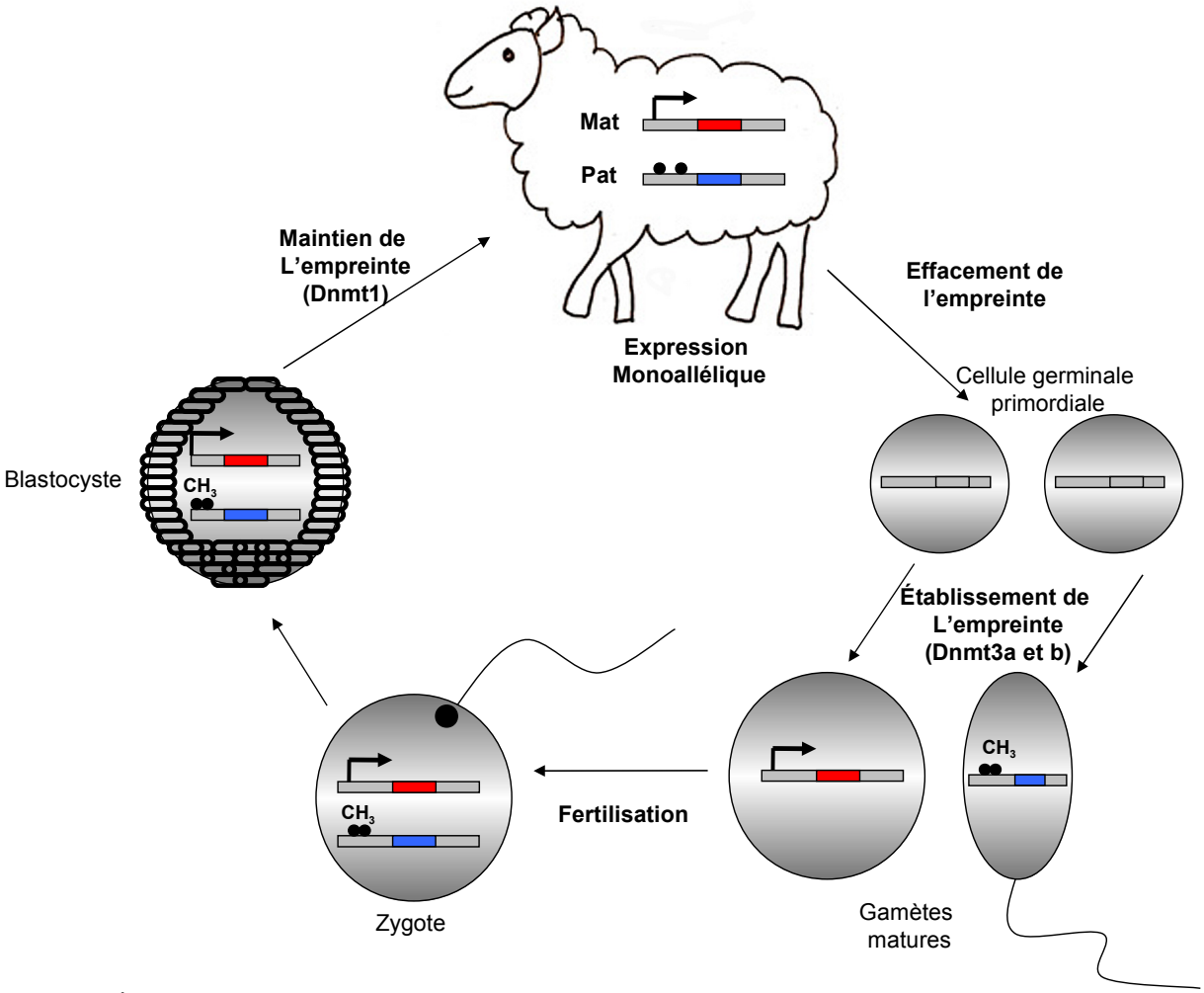


Figure 5 : Établissement et maintien de l'empreinte. Les marqueurs épigénétiques de l'empreinte sont effacés dans les cellules germinales primordiales puis rétablis de novo pendant leur maturation par Dnmta3 et Dnmt3b. Le zygote issu de la fertilisation présentera ainsi un état d'empreinte qui sera ensuite maintenu à chaque division somatique par Dnmt1.

III. Description du locus *CLPG*

A. Le locus DLK1-GTL2

Le locus impliqué dans la mutation callipyge présente de nombreuses particularités. Situé en région télomérique du chromosome 18 ovin, respectivement orthologue aux chromosome 14 humain et chromosome 12 murin, les transcrits qu'il exprime sont soumis à l'empreinte. On distingue quatre gènes à expression paternelle, encodant des protéines (*BEGAIN*, *DLK1*, *PEG11* et *DIO3*), et des transcrits non codants à expression maternelle (*GTL2*, *anti-PEG1*, *MEG8* et *MIRG*). De nombreux petits ARNs, comme des snoRNAs ou des

microARNs (miRNAs) sont également présents. Ce locus joue un rôle fondamental dans le développement embryonnaire. En effet, une souris présentant une disomie uniparentale paternelle du chromosome 12 (PatUPD12) ne survit pas à la gestation et montre une hypertrophie musculaire prénatale, un défaut de maturation des muscles et du cartilage ainsi qu'une placentomégalie caractéristique. De même, les souris MatUPD12 (maternal uniparental disomy 12) ne sont pas viables et présentent un retard global de croissance, à la fois embryonnaire et placentaire (Georgiades et al. 2000 ; Georgiades et al. 2001).

L'analyse par séquençage bisulfite du locus a permis de mettre en évidence trois régions différentiellement méthylées : deux spécifiquement liées à un gène (respectivement *DLK1* et *GTL2*), et une troisième, nommée *IG-DMR*, située dans une zone intergénique (cf Figure 6). Ces trois régions présentent une hyperméthylation sur l'allèle paternel et une hypométhylation sur l'allèle maternel, mais seule l'IG-DMR hérite de ces marques épigénétiques des cellules germinales (Paulsen and Ferguson-Smith 2001; Takada et al. 2002). Lin et al. (2003) ont montré que la délétion de l'*IG-DMR* chez la souris a pour effet, en cas de transmission maternelle, de totalement annuler tout phénomène d'empreinte génétique dans le locus. Tous les gènes du locus se comportent alors comme le chromosome paternel, ce qui se traduit par un phénotype comparable aux souris PatUPD12 : aucune expression des transcrits non codants et un double dosage des gènes codants. Inversement, la transmission paternelle de la délétion n'affecte pas le statut d'empreinte, prouvant par là que chaque chromosome contrôle son état d'empreinte différemment. Récemment, Kagami et al. (2010) ont utilisé deux patients nouveau-nés présentant chacun une microdélétion dans les régions *DMR* (respectivement dans la région *IG-DMR* et dans la région *DMR* de *GTL2*) afin de préciser leurs fonctions. Les auteurs ont ainsi mis en évidence que la région *DMR-GTL2*, hypométhylée dans l'embryon, joue un rôle essentiel dans l'établissement de l'état d'empreinte des gènes du locus embryonnaire alors que l'*IG-DMR* régule les gènes du placenta. L'*IG-DMR* a néanmoins un rôle de régulateur en amont de l'état de méthylation de la *DMR-GTL2* dans l'embryon. A ce jour, les mécanismes par lesquels ces *DMRs* établissent l'état d'empreinte restent toujours inconnus.

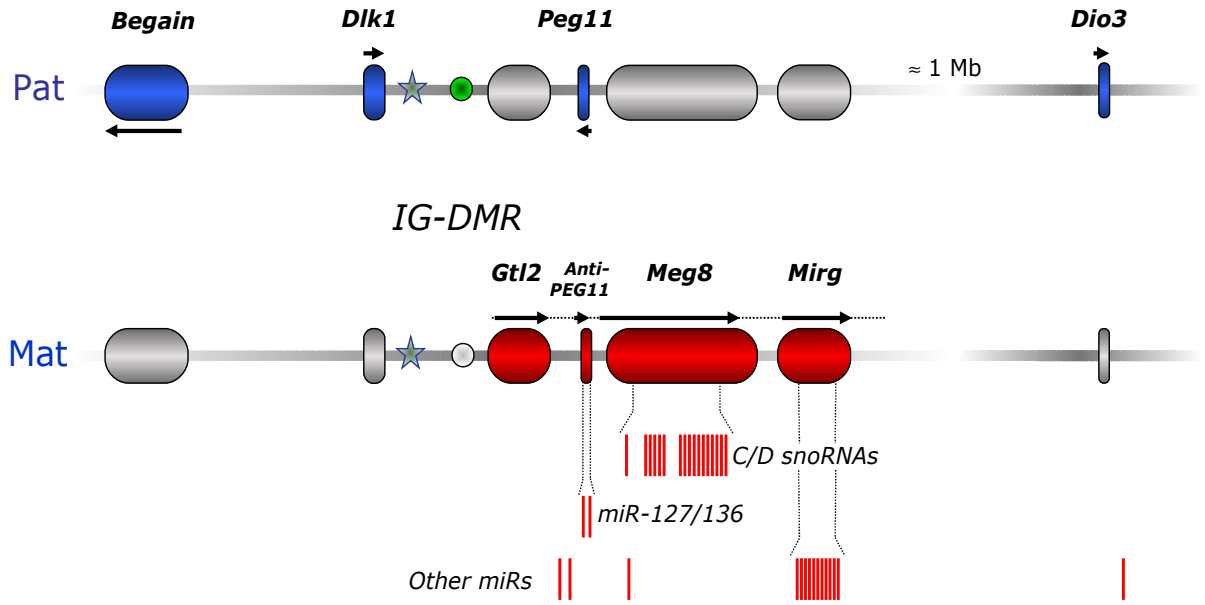


Figure 6 : Locus *DLK1-GTL2*. Quatre gènes, exprimés paternellement (en bleu) et 4 transcrits non codants maternellement exprimés (en rouge). Une région intergénique différentiellement méthylée (*IG-DMR*, rond vert) assure l'empreinte maternelle, tandis que la mutation *CLPG* (étoile) est responsable du phénotype d'hypertrophie musculaire. Certains gènes exprimés maternellement contiennent de petits ARNs non codants (microARNs et snoRNAs).

Certains transcrits du locus voient leurs taux d'expression affectés par la mutation callipyge (voir l'effet en cis de la mutation CLPG) et sont ainsi susceptibles d'avoir une influence sur le phénotype callipyge (cf Figure 6).

1. *DLK1*

Historiquement, la protéine DLK1 (Delta-Like 1) fut identifiée par le clonage d'un cDNA (nommé *Dlk1*) à partir de cellules 3T3-L1, fibroblastes murins capables de se différencier en adipocytes (Laborda et al. 1993). Avant l'officialisation des nomenclatures, *Dlk1* portait différents noms, chacun indiquant la fonction spécifique attribuée lors de sa découverte. Ainsi, *Dlk1* était connu chez la souris sous le nom de *Pref-1* (préadipocyte factor-1) ou *Scp-1* (stromal cell protein-1). *Dlk1* se nommait *FA-1* (pour Fetal-Antigen-1) chez l'humain, auquel on attribuait un rôle dans la différenciation des cellules pancréatiques. Très exprimé dans le cerveau, *Dlk1* y avait été nommé *PG2*, auquel on attribuait également un rôle de facteur de différenciation. Enfin, Okamoto et al. (1997) avaient identifié *Dlk1* dans le cerveau de rats sous le nom de *Zog* (zona glomerulosa-specific factor).

Toutes ces études conféraient à *Dlk1* un rôle dans la différenciation cellulaire. Le gène s'exprime dans un très grand nombre de tissus pendant le développement mais est éteint dans la grande majorité des tissus adultes. On lui impute aujourd'hui le maintien de l'état indifférencié (Floridon et al. 2000). Ainsi, une souris transgénique KO pour *Dlk1* souffre d'une obésité probablement due à la différenciation non contrôlée des fibroblastes en adipocytes (Moon et al. 2002). Cette souris présente également des retards importants de croissance et des malformations squelettiques.

Cette protéine est considérée comme proche de la famille Notch-Delta, dans laquelle on distingue les ligands DELTA (*Serrate*, *Dll1*, *Dll2*, *Dll3*, *Jagged1* et *Jagged2*) capables de se lier aux récepteurs NOTCH (au nombre de quatre chez les mammifères). Les ligands DELTA sont des protéines transmembranaires jouant un rôle essentiel dans la communication intercellulaire par l'intermédiaire de mécanismes de régulation contrôlant la différenciation et le passage de l'état embryonnaire à l'adulte. Lorsqu'ils entrent en contact avec un récepteur NOTCH, lui aussi membranaire, le ligand est clivé et sa partie intramembranaire migre vers le noyau où il agira sur l'expression de certains gènes en s'associant avec des facteurs de transcription. Ces voies métaboliques fondamentales sont fréquemment dérégulées dans les cellules tumorales. Si *Dlk1* présente des caractéristiques semblables à la famille Notch-Delta, comme la présence de 6 motifs EGF répétés (Epidermal Growth Factor) susceptibles de se lier aux NOTCHs dans un système double hybride (Baladron et al. 2005), il lui manque toutefois le domaine N–terminal DSL (Delta Serrate Lage2) indispensable à l'activité des ligands connus, tandis que son domaine intramembranaire est beaucoup plus petit (cf Figure 7). *Dlk1* n'est, par ailleurs, pas toujours transmembranaire puisque certains transcrits alternatifs peuvent être clivés par une protéase (ADAM17) rendant la protéine soluble et sécrétable hors de la cellule (Laborda 2000; Wang and Sul 2006). *Dlk1* pourrait donc bien être un antagoniste des récepteurs NOTCHs, capable de s'y lier sans activer aucun mécanisme de différenciation.

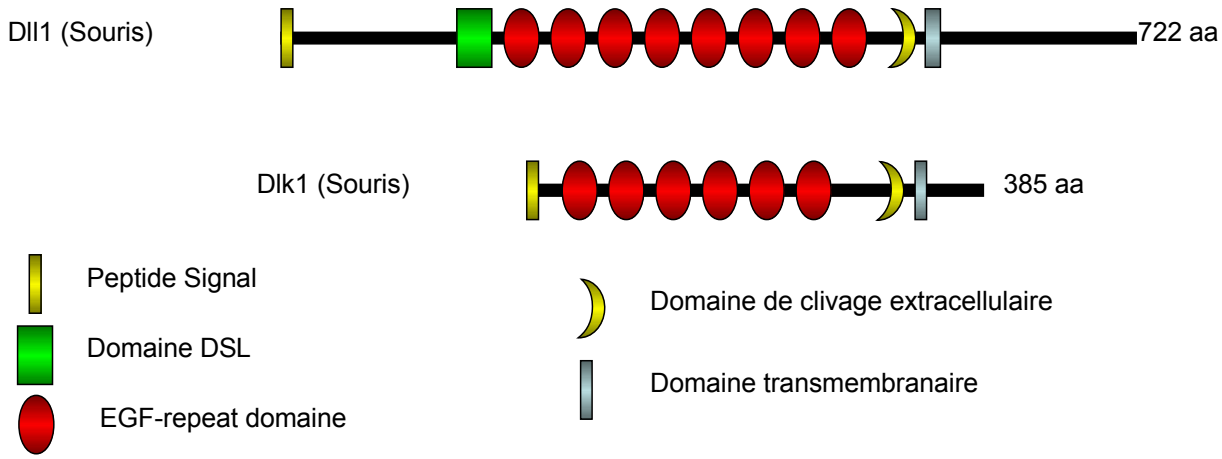


Figure 7 : Domaines de la protéine DLK1. La protéine DLK1 présente nombreux domaines en commun avec DLL1, une protéine de la famille Delta-Notch. L'absence du domaine DSL, indispensable à l'action du ligand ddl1, ainsi qu'une région transmembranaire plus courte, distinguent cependant les deux protéines l'une de l'autre. (adapté de Bray et al., 2008).

Un autre critère différenciant *Dlk1* des membres de la famille Delta-Notch est son statut d'empreinte génétique. En effet, le gène s'exprime uniquement depuis l'allèle paternel, contrairement à tous les membres de cette famille, qui sont à expression biallèlique. Ainsi, Moon et al. (2002) ont rapporté que la souris *null-dlk1* hétérozygote ayant reçu la mutation maternellement ne présente aucun trouble phénotypique. DLK1 est donc incontestablement essentielle au développement embryonnaire et ces auteurs ont été jusqu'à suggérer que toutes les anomalies observées chez les souris MatUPD12 étaient dues à l'absence de la protéine. La fonction exacte de DLK1 est cependant toujours inconnue.

2. GTL2

Gtl2 (Gene Trap Locus 2), également connu sous le nom *Meg3* (Maternally Expressed Gene 3), fut découvert en 1994 (Schuster-Gossler et al. 1996) lors d'une expérience de capture de gènes chez la souris. Les auteurs utilisèrent un vecteur contenant le gène rapporteur LacZ, dénué de promoteur, afin d'étudier les gènes importants pour le développement embryonnaire. Ce vecteur mit en évidence le gène *Gtl2* et son profil d'expression. Ainsi, *Gtl2* est exprimé dans la majorité des tissus embryonnaires. Chez la souris adulte, on retrouve en outre *Gtl2* dans tous les muscles squelettiques striés, ainsi que dans les cellules du système nerveux.

La lignée de souris transgéniques porteuses de ce vecteur permit aussi aux auteurs de mettre en évidence le statut d'empreinte de *Gtl2*. Ainsi, au travers d'une étude d'expression dans un embryon parthénogénique, Schuster-Gossler et al. (1996) démontrèrent l'expression maternelle du gène. Fait surprenant, les auteurs décrivirent que la transmission maternelle du transgène *Gtl2lacZ* n'avait aucun effet sur le phénotype alors que sa transmission paternelle générait des souris naines. Or, le vecteur rapporteur réduit considérablement l'expression (sans l'interrompre totalement) et l'empreinte maternelle de *Gtl2* laissait plutôt présager des troubles lors de transmission maternelle. Une explication plausible à cette observation pourrait venir de l'étude de Croteau et al. (2003), qui a montré que l'empreinte de *Gtl2* n'était pas totalement stricte selon les lignées murines. Une autre possibilité impliquerait le lieu d'insertion peu commun du gène rapporteur, situé 1.6 kb en amont de *GTL2,* en plusieurs copies et transcrit en orientation opposée. Steshina et al. (2006) ont rapporté que l'insertion d'une cassette *Neo* au même niveau que le site d'intégration du vecteur *LacZ* conduit exactement aux mêmes conséquences phénotypiques et que cet élément étranger a pour effet d'annuler le statut d'empreinte de *Gtl2*. Ainsi, ils ont postulé que la présence d'ADN exogène proche de la région DMR pourrait interférer avec sa méthylation et bouleverser l'état d'empreinte à proximité.

Très peu de données sont disponibles sur le rôle physiologique de *Gtl2*. Diverses études ont mis en évidence de nombreux transcrits alternatifs de taille différente. La forme majoritaire selon Schuster-Gossler et al. (1996) serait un long ARN polyadénylé de 6.5 kb. L'absence d'un fort signal d'initiation de traduction suggère que *Gtl2* n'encode aucune protéine, et agit donc sous la forme d'un ARN non codant. Une souris transgénique récemment développée et possédant une délétion des exons 1 à 5 de *Gtl2* (Takahashi et al. 2009) démontre l'importance capitale de ce gène chez la souris. La transmission maternelle de cette délétion donne naissance à des souriceaux normaux, mais qui présentent rapidement des troubles de croissance, principalement localisés au niveau des poumons et du foie, qui conduisent irrémédiablement à une mort prématurée (100 % de létalité après 4 semaines) . Lorsque la région promotrice est inclus dans le KO (Zhou et al., 2010), la région IG-DMR, normalement hypométhylée sur l'allèle maternel, se trouve alors hyperméthylé, avec pour conséquence une inhibition des transcrits maternels au profit des gènes paternels.

La transcription active de *Gtl2* serait donc capitale à la régulation de l'empreinte dans le locus. Plus surprenant, la transmission paternelle crée également un phénotype : un retard de croissance généralisé mortel pour 3 individus sur 4. Les homozygotes mutés sont quant à eux tout à fait normaux et donnent des adultes viables et fertiles. Les auteurs ont étudié les modifications d'expression géniques associées à cette délétion et ont notamment découvert une réduction de 85 % de l'expression de *Dlk1* dans le KO-paternel, ainsi qu'une perte de l'état d'empreinte des gènes *Mirg* et *Rian*. Cependant, ces observations concernent la souris et des tissus non musculaires ; elles sont donc difficilement transposables au modèle callipyge.

Gtl2 seul contient un promoteur bien défini (de type « TATA box ») parmi les transcrits maternels du locus. Tous ces transcrits non codant ayant la même orientation, il est suggéré qu'ils proviennent d'un seul long ARN polycistronique processé après la transcription en plusieurs ARNs distincts (da Rocha et al., 2008).

3. *PEG11* et *anti-PEG11*

PEG11 (pour Paternally Expressed Gene 11) a été initialement prédit par Charlier et al. (2001b) lors du séquençage du contig ovin couvrant le locus de la mutation callipyge (voir plus haut). Dépourvu d'intron et possédant des éléments très proches des domaines POL et GAG, sa structure rappelle celles des rétrotransposons, d'où le nom *Rtl1* conféré à son orthologue murin (pour retrotransposon-like 1). Contrairement aux rétrotransposons classiques, *PEG11* est très conservé chez les mammifères, ce qui laisse présager d'un rôle important. Charlier et al. (2001a) ont détecté son expression dans les muscles squelettiques et, dans une moindre mesure, dans le cœur, les reins et les poumons. La mise en évidence de la protéine PEG11 n'a été entreprise que très récemment grâce à des techniques d'immunologie couplées à de la spectrométrie de masse (Byrne et al. 2010).

Une particularité de *PEG11* est l'existence d'un transcrit antisens, découvert et nommé *anti-PEG11* par Charlier et al. (2001b), dont le statut d'empreinte est réciproque. Ainsi, si *PEG11* est exprimé depuis l'allèle paternel, *anti-PEG11* est transcrit maternellement. Seitz et al. (2003) ont décrit la présence de deux miRNAs (voir partie IV, C, 1) issus du transcrit *anti-PEG11* : miR-127 et miR-136. Les auteurs ont prédit que ces miRNAs, de

séquence parfaitement complémentaire à *PEG11* devaient logiquement pouvoir réguler l'expression de la protéine. Les régions les plus conservées d'*anti-PEG11* sont les miRNAs, et pour deux d'entre eux, les régions équivalentes correspondent à un domaine structural ou fonctionnel de la protéine, démontrant ainsi une évolution parallèle des deux transcrits en orientation antisens l'un par rapport à l'autre (Byrne et al. 2010).

Afin d'élucider le rôle dans le développement embryonnaire de ces deux transcrits antisens, Sekita et al. (2008) ont récemment réalisé la délétion totale des domaines POL et GAG de *Rtl1* (l'orthologue murin de *PEG11*). La réciprocité d'empreinte des deux transcrits permet d'observer l'effet de chaque KO suivant l'origine de transmission. Ainsi, la transmission paternelle (pat-KO) de la délétion génère une souris dépourvue de la protéine PEG11. La transmission maternelle (mat-KO), amputée des miRNAs d'*anti-PEG11*, présente une surexpression du transcrit *PEG11*. Les souris Pat-KO n'ont aucun phénotype anormal jusqu'au stade 15 d.p.c. (days post coitum), mais les auteurs ont constaté 50 % de létalité entre 15 d.p.c. et la fin de la gestation. Tous les animaux survivant à la gestation meurent à la naissance et présentent un retard de croissance. De plus, l'étude histologique de leur placenta montre de nombreuses anomalies anatomiques. Les souris mat-KO subissent également un taux de mortalité élevé, associé à d'importantes modifications placentaires. Ainsi, l'absence ou la surexpression de *PEG11* créent des troubles majeurs et ce gène proche des rétrotransposons nécessite donc une régulation fine. Bien que sa véritable fonction reste encore un mystère, les auteurs postulent que *PEG11* joue un rôle majeur dans le développement et l'évolution du placenta chez les mammifères. Byrne et al. (2010) ont démontré que le profil d'expression de la protéine PEG11 est compatible avec l'effet de la mutation *CLPG* et sa surexpression pourrait donc, conjointement à DLK1, participer à l'émergence du phénotype d'hypertrophie musculaire.

4. *MEG8*

MEG8 (Maternally Expressed Gene 8) fut identifié pour la première fois chez la souris lors d'une analyse de visualisation différentielle par fluorescence des transcrits d'embryons androgéniques et parthénogéniques (Hatada et al. 2001). Ce nouveau transcrit, localisé à proximité de *Gtl2*, était exprimé uniquement maternellement, exclusivement dans le noyau

des cellules du cerveau adulte, et fut baptisé *Rian* (RNA imprinted and accumulated in the nucleus). Chez le mouton, il a été identifié et appelé *MEG8* par comparaison de séquences et prédiction informatique (Charlier et al. 2001 b). Il ne contient aucune phase ouverte de lecture, présente plusieurs exons et transcrits alternatifs et s'exprime dans les muscles squelettiques (ainsi que dans les reins).

Si sa fonction est toujours inconnue dans toutes les espèces mammifères, Cavaille et al. (2000 ; 2001) ont découvert que le transcrit *MEG8* contenait deux clusters de répétitions en tandem dont les motifs sont caractéristiques des snoRNAs (small nucleolar RNAs). Ces clusters, nommés 14q(I) et 14q(II), contiennent respectivement 9 et 31 snoRNAs, auxquels s'ajoute 14q(0), un unique snoRNA conservé entre humain et ovin. Localisés dans les introns de *MEG8*, ils possèdent le même profil d'expression que leur transcrit hôte. Les snoRNAs sont des petits ARNs non codants dont la fonction est la modification chimique d'autres ARNs, ciblés par complémentarité de séquence. Ils assurent notamment la 2'O-ribose-méthylation des ARN ribosomaux et la pseudouridylation des ARNs de transfert. Cependant, les snoRNAs du locus, au contraire de tous ceux déjà identifiés, ne montrent aucune complémentarité de séquences avec des ARNr ou des ARNt (ils sont dits snoRNAs orphelins). Ainsi, tout comme *MEG8*, leur fonction reste encore à découvrir.

5. *MIRG*

Mirg (pour microRNA-containing gene) fut décrit initialement chez la souris comme un cluster de séquences répétées d'environ 75 nucléotides (Cavaille et al. 2002). Rapidement, Seitz et al. (2003) en parlèrent comme un nouveau gène, soumis à l'empreinte et exprimé maternellement. Ce transcrit semble dénué de cadre ouvert de lecture et fonctionne donc certainement comme un ARN non codant. Ces auteurs y identifièrent deux miRNAs (cf encadré miRNA), mir-134 et mir-154, ouvrant ainsi les premières hypothèses de *trans*-interaction impliquant le locus *DLK1-GTL2* (voir effet en *trans*).

Une étude informatique, basée sur la détection *in silico* de molécules susceptibles d'adopter une structure en épingle à cheveux conservée entre l'homme et la souris, a prédit l'existence de 29 miRNAs potentiels localisés dans le transcrit *Mirg* (Seitz et al. 2004). Par expérience d'extension d'amorce fut confirmée l'existence de onze d'entre eux, exprimés

dans le cerveau de souris. Tous ces nouveaux miRNAs étaient détectables dans les embryons murins matUPD12, mais non dans les patUPD12, prouvant ainsi leur statut d'empreinte maternelle chez la souris. Le transcrit *Mirg*, ainsi qu'une grande partie des miRNAs qu'il contient, sont conservés chez le mouton (cf Chapitre 4).

6. *DIO3*

Dio3 (deiodinase iodothyronine de type III) code pour une sélénoprotéine qui catalyse la dégradation des hormones thyroïdiennes T3 (3,5,3'-triiodothyronine) et T4 (Thyroxine) et participe ainsi à leur régulation dans de nombreux tissus fœtaux et dans le placenta (Salvatore et al. 1995). Le gène fut localisé dans le locus *DLK1-GTL2* par Hernandez et al. (1998) chez la souris. *Dio3* s'exprime uniquement depuis l'allèle paternel, comme tous les gènes encodant une protéine qui ont été identifiés dans le locus (Hernandez et al. 2002). Comme beaucoup de gènes soumis à l'empreinte, *Dio3* possède un transcrit antisens non codant, *Dio3as,* dont la partie 5' recouvre la séquence promotrice de *Dio3*. Les deux transcrits présentent le même profile d'expression, suggérant donc un mécanisme commun de régulation.

Une souris transgénique déficiente pour *Dio3* a récemment démontré le rôle crucial de ce gène dans le développement embryonnaire (Hernandez et al. 2006). En effet, ces souris présentent des troubles majeurs de croissance, de viabilité et de fertilité. En revanche, le transcrit antisens n'est pas affecté par ce KO et sa fonction reste inconnue à ce jour. Smit et al. ont montré en 2005 que *Dio3* n'était pas soumis à l'effet de la mutation *CLPG* (voir effet en *cis*) et n'a donc pas d'effet direct sur le phénotype d'hypertrophie musculaire.

7. *BEGAIN*

Originellement décrit chez le rat par Deguchi et al. en 1998 comme un gène spécifique du cerveau, le gène *Begain* (pour Brain-enriched guanylate kinase-associated protein) a depuis lors été étudié en détail chez le mouton (Smit et al. 2005). Grâce à une analyse comparative de séquences génomiques et des banques de données d'EST, les auteurs purent non seulement confirmer que *Begain* était conservé chez les mammifères (et

donc présent chez le mouton), mais qu'il générait pas moins de quatre transcrits primaires issus de promoteurs alternatifs. Contrairement au rat, le gène *BEGAIN* ovin est exprimé dans de nombreux tissus. Ainsi, si le transcrit est détecté dans le cerveau, sa présence a été également constatée dans le cœur, le foie, les reins, les poumons, le muscle squelettique et la langue. Capable d'interagir avec le domaine guanylate kinase de la protéine PSD-95/SAP90, une protéine d'architecture membranaire des cellules post-synaptiques, la protéine BEGAIN pourrait avoir un rôle dans la transduction du signal cellulaire, dans l'établissement du cytosquelette et dans le mouvement intracellulaire de ses récepteurs (Deguchi et al. 1998 ; Sheng 2001).

BEGAIN est situé à proximité du locus *DLK1-GTL2*, mais son statut d'empreinte est longtemps resté inconnu. Smit et al. (2005) démontrèrent par étude de polymorphisme chez l'ovin que son expression était bien monoallélique et provenait de l'allèle paternel. Mais au contraire des autres gènes du locus, son statut d'empreinte n'est pas homogène et semble à la fois dépendre du promoteur et du tissu. Ainsi, l'une des formes du gène, *BEGAIN1*, est soumise à l'empreinte dans le cerveau, le cœur et le muscle squelettique, mais est exprimée de façon biallélique dans le foie. *BEGAIN2*, exprimé à partir d'un promoteur différent, n'est soumis à l'empreinte dans aucun des tissus testés. Chez la souris, *Begain* présente également une empreinte dépendante du promoteur (et non du tissu, le gène ne s'exprimant que dans le cerveau) (Tierling et al. 2009).

BEGAIN se démarque donc de l'ensemble du locus par son statut d'empreinte particulier. Tout comme le gène *DIO3*, la mutation *CLPG* n'affecte en rien son niveau d'expression (Smit et al. 2005) et ce gène ne semble donc pas participer au phénotype callipyge (voir plus loin l'effet en *cis* de la mutation).

B. La mutation CLPG est un SNP

Le clonage positionnel avait donc localisé la mutation responsable du phénotype callipyge dans un locus soumis à l'empreinte, contenant (i) des gènes codant pour des protéines essentielles au développement embryonnaire et placentaire à expression paternelle et (ii) de nombreux transcrits non codants maternellement exprimés et de

fonction souvent inconnue. Tous ces transcrits avaient été détectés dans le muscle squelettique strié et l'un d'eux était susceptible de porter une mutation induisant l'hypertrophie musculaire observée chez les moutons callipyges.

Pour localiser la mutation *CLPG*, deux groupes indépendants réalisèrent le séquençage total du locus chez le mouton porteur de l'allèle *CLPG* et d'un allèle sauvage (+) (Freking et al. 2002 ; Smit et al. 2003). Afin de minimiser les polymorphismes non impliqués dans le phénotype callipyge, l'allèle sauvage fut sélectionné sur base de sa proximité de séquence avec l'allèle *CLPG*, en s'assurant que tous les SNPs connus du locus soient homozygotes avec l'allèle *CLPG*. Sur les 184 kb séquencés à l'aide de 209 PCRs chevauchantes, un seul nucléotide différenciait les deux allèles. Nommé SNP^{CLPG}, ce polymorphisme est une substitution d'A (sauvage, ou +) vers G (*CLPG*) localisé à 32 775 pb en amont de *GTL2*. Le génotypage des cohortes ovines confirma que ce SNP était bien responsable de l'hypertrophie musculaire. Ainsi, les animaux présentant le phénotype callipyge ($+^{mat}/CLPG^{pat}$) étaient tous hétérozygotes (A/G) tandis que les trois autres génotypes (A/A ; G/A ; G/G) étaient de phénotype normal.

L'alignement de la séquence ovine sur celle de 12 espèces de mammifères mit en évidence que le SNP^{CLPG} se situait dans un motif de 12 pb parfaitement conservé (Smit et al. 2003). La conservation de ce dodecamère laisse à croire que le SNP^{CLPG} affecte un motif régulateur important du locus, tel qu'un site de fixation protéique. Freking et al. (2002) montrèrent par EMSA que cette région était capable de fixer le facteur de transcription musculaire MYOD1. Cependant, la mutation n'ayant aucun effet sur cette fixation, MYOD1 ne semblait pas directement impliqué dans le phénotype d'hypertrophie musculaire.

Fait intéressant, le génotypage d'ADN extrait de leucocytes de Solid Gold, le bélier fondateur du phénotype callipyge, révéla que ce dernier était mosaïque pour le SNP^{CLPG}. Ainsi, contrairement à un ratio de 1 : 1 attendu pour un animal présentant le phénotype callipyge, il fut estimé par hot-stop PCR que Solid Gold possédait environ 20 % d'allèle G (Smit et al. 2003). La transition d'A vers G affectant Solid Gold a donc probablement eu lieu au cours de son développement embryonnaire (et affecter en partie ses cellules de la lignée germinale puisqu'il a pu ensuite transmettre la mutation).

IV. Mécanisme moléculaire de la surdominance polaire

A. *Effet en cis de la mutation CLPG*

Avant même la découverte du SNP^{CLPG}, Charlier et al. (2001) avaient constaté que la mutation *CLPG* augmentait en *cis* l'expression de gènes du locus *DLK1-GTL2* sans affecter le statut de leur empreinte génétique. En effet, si durant le développement embryonnaire, tous les gènes du locus présentent un profil d'expression similaire dans les quatre génotypes callipyge, leur niveau de transcription postnatale est progressivement inhibé chez le mouton de type sauvage. La présence du SNP^{CLPG} a pour effet de contrer cette inhibition postnatale en *cis* et de conduire à l'expression ectopique des gènes exprimés depuis l'allèle muté. Ainsi, la mutation *CLPG* transmise par l'allèle maternel (C^{mat}) causera l'expression ectopique des transcrits non codants et la transmission paternelle (C^{pat}) modifiera la transcription de *DLK1* et *PEG11*, les transcrits *BEGAIN* et *DIO3* n'étant pas affectés par l'effet en *cis* de la mutation (voir Figure 8).

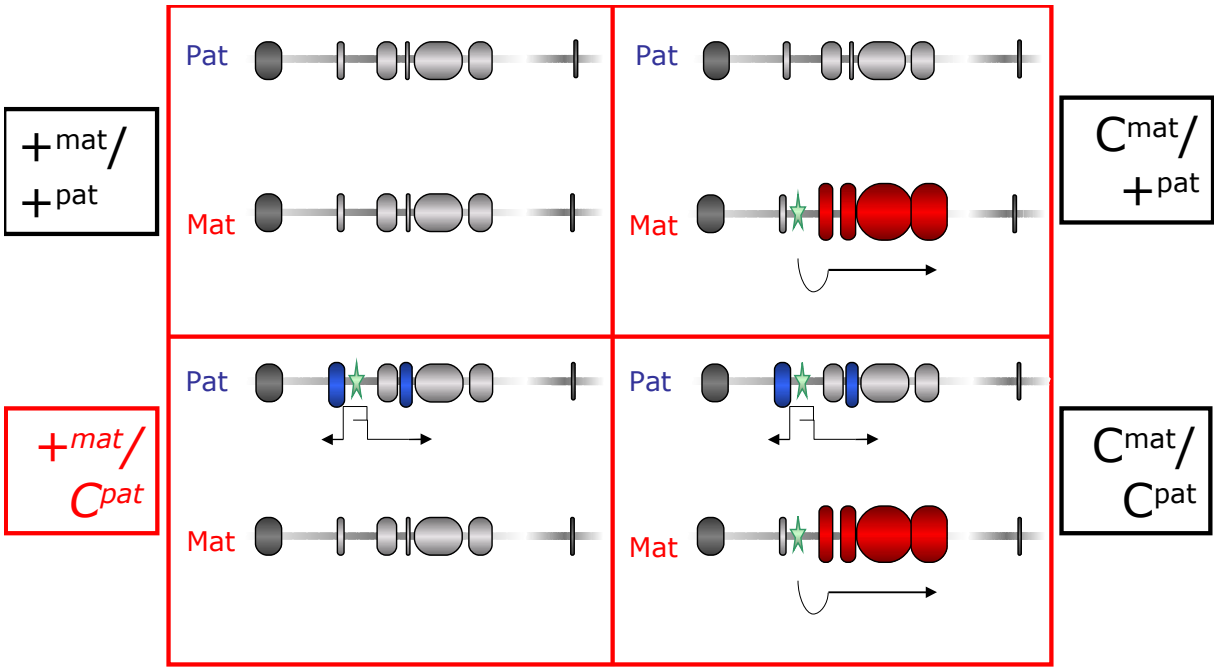


Figure 8 : Effet en *cis* de la mutation CLPG. La présence de la mutation *CLPG* (étoile verte) détruit un élément de contrôle longue distance et cause l'expression ectopique des gènes en *cis* sans affecter le statut de leur empreinte génétique. La mutation paternelle ne modifie l'expression que de deux gènes sur les quatre exprimés paternellement (*DLK1* et *PEG11*).

Ces données suggéraient que la mutation *CLPG* affecte un élément de contrôle longue distance (LRCE, pour Long Range Control Element) influençant l'expression des gènes du locus. Deux hypothèses furent alors envisagées : soit le SNP^{CLPG} détruisait la fixation d'un inhibiteur, soit il permettait la liaison d'un activateur (Georges et al. 2003). La génération de deux lignées de souris transgéniques, l'une avec la même transition A vers G que le mouton callipyge (nommée lignée AG) et l'autre avec la délétion du dodecamère conservé (lignée Δ12) permit de rejeter l'hypothèse de la création d'un motif activateur (Pirottin-Takeda, en préparation). En effet, ces deux lignées murines présentaient les mêmes profiles d'expression, démontrant par là que le LRCE n'est pas indispensable à l'effet en *cis* de la mutation et donc que le SNP^{CLPG} ne crée pas un site de fixation pour un activateur.

La mutation *CLPG* inactive donc un LRCE spécifique du muscle squelettique responsable, chez l'allèle sauvage, de l'inhibition postnatale des gènes voisins. Très peu d'informations sur les mécanismes moléculaires engendrant l'effet en *cis* de la mutation sont aujourd'hui disponibles. Freking et al. (2002) ont montré que le facteur musculaire MYOD1 pouvait se lier à la région, et que la région contenant le SNP^{CLPG} semblait transcrite (transcrit *CLPG1,* dont la fonction est inconnue). Par conséquent, la compréhension des mécanismes impliqués dans la mise en place de l'expression ectopique observée dans la surdominance polaire chez le mouton callipyge nécessitera soit la découverte de nouveaux facteurs capables de se fixer au dodecamère, soit l'étude des conséquences de la mutation sur le profil épigénétique du locus.

B. DLK1 et le phénotype callipyge

Le phénotype callipyge correspond à une expression ectopique des messagers *DLK1* et *PEG11* en l'absence de ~~la~~ surexpression des transcrits non codants maternels (cf Figure 8). Afin de clarifier les relations entre les profiles d'expression et le phénotype callipyge, Davis et al. (2004) étudièrent l'expression de la protéine DLK1 par immunohistochimie dans des coupes de muscles squelettiques (*Longissimus dorsi*) d'animaux des quatre génotypes callipyges. Différents stades de développement furent testés : deux semaines prénatales, deux semaines postnatales et huit semaines postnatales. Sur tous ces cas, seuls les animaux de phénotype callipyge âgés de huit semaines (et de génotype $+^{mat}/C^{pat}$) présentaient la

protéine DLK1. En outre, ces auteurs observèrent une corrélation positive entre le diamètre des fibres musculaires et le taux d'expression de la protéine.

L'expression ectopique dans les fibres musculaires de la protéine DLK1 semblait donc jouer un rôle direct dans l'établissement du phénotype callipyge. En vue de tester cette hypothèse, une souris transgénique exprimant *DLK1* sous l'influence d'un promoteur musculaire puissant fut générée (Davis et al. 2004). Les souris issues de cette transgénèse avaient un poids globalement inférieur aux souris de type sauvage. Cependant, tant la masse que le diamètre des fibres musculaires des souris mutantes se révélèrent significativement supérieurs aux contrôles (respectivement +9.3 % et 7.4 % chez les souris hétérozygotes mâles +/T comparées aux souris *+/+*).

L'expression ectopique de la protéine DLK1 est donc responsable, au moins en partie, du phénotype callipyge. Cette conclusion est étayée par (i) des travaux qui lient la voie d'activation des protéines de la famille Notch-Delta à la myogenèse (Hirsinger et al. 2001 ; Conboy et al. 2003), (ii) l'hypertrophie des myofibrilles des souris PatUPD12 (Georgiades et al. 2000) et (iii) l'implication de *DLK1* en tant qu'inhibiteur de la différentiation des adipocytes (le mouton callipyge présentant un taux de graisse inférieur à la normale).

En revanche, *PEG11*, le second transcrit paternel affecté par l'effet en *cis* de la mutation *CLPG* n'a jamais été associé à la myogenèse et la protéine correspondante a peu de fonctions connues à ce jour. Toutefois, la létalité causée par la délétion ou la surexpression du gène suggère un rôle essentiel. Dès lors, son implication dans le phénotype callipyge ne peut être écartée. Une lignée de souris transgénique exprimant *Peg11* dans les fibres musculaires est actuellement à l'étude (cf Discussion et Perspectives).

C. Effet en trans de la mutation

L'expression ectopique dans les muscles squelettiques du transcrit *DLK1* est donc responsable du phénotype callipyge chez les animaux hétérozygotes ayant reçu la mutation de leur père ($+^{mat}/C^{pat}$). Si le transcrit *DLK1* est tout autant surexprimé chez les animaux homozygotes mutants (C^{mat}/C^{pat}), pourquoi ne présentent-ils pas eux aussi le phénotype callipyge ? Tout porte à croire que l'expression ectopique des transcrits maternels non codants est capable de contrer les effets de la surexpression postnatale de *DLK1*.

Cette transinteraction entre les allèles paternels et maternels, également nommée effet en *trans* de la mutation *CLPG*, est observée à différents niveaux biologiques. L'analyse quantitative par northern blot (Charlier et al. 2001a) et par QRT-PCR (non publié) montrent que si les muscles squelettiques des animaux porteurs de la mutation paternelle surexpriment les transcrits *DLK1* et *PEG11*, les animaux C^{mat}/C^{pat} présentent un taux d'expression inférieur au $+^{mat}/C^{pat}$. La transmission maternelle de la mutation *CLPG* induit donc une légère inhibition des transcrits paternels. Inversement, les données d'expression indiquent que la mutation paternelle augmente la quantité de transcrits maternels (*GTL2* et *MEG8*) détectables chez les animaux C^{mat}/C^{pat}. On observe donc un double effet en *trans* de la mutation *CLPG*, résumé par la Figure 9.

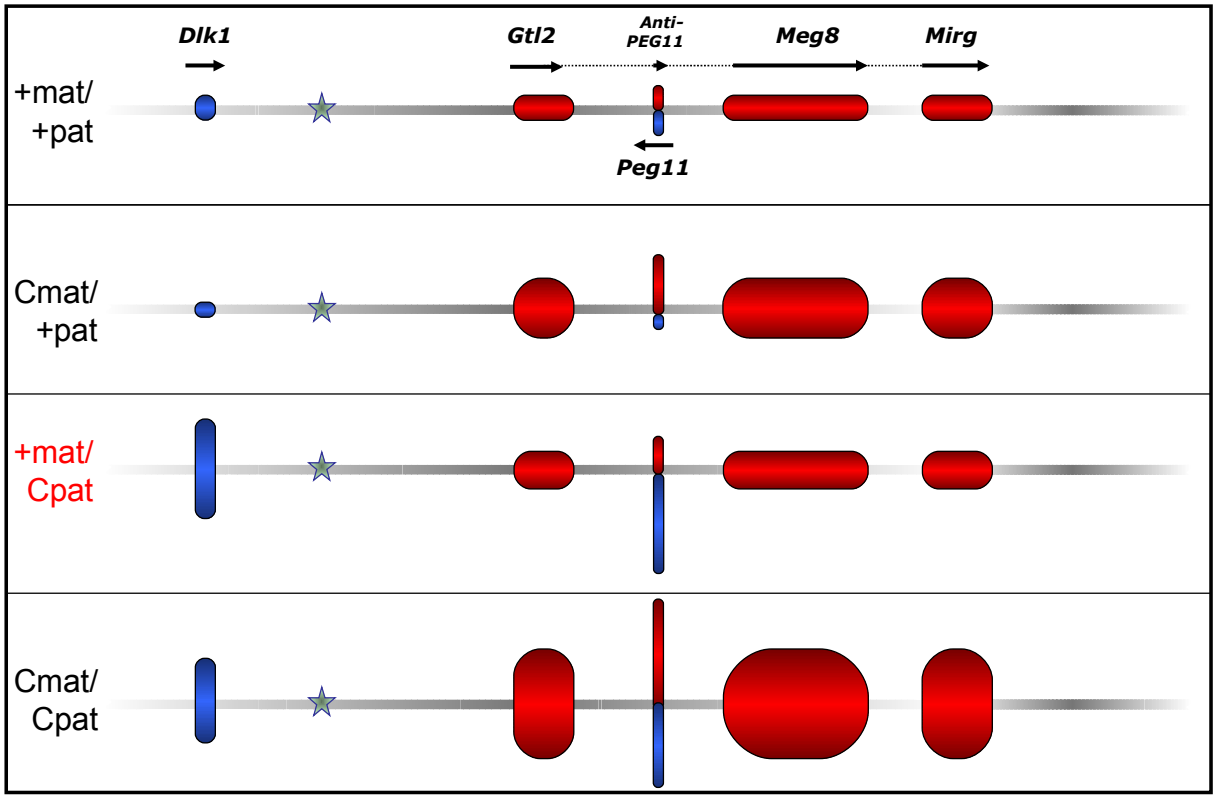


Figure 9 : Effet en *trans* de la mutation *CLPG* sur les quatre génotypes callipyges. Les transcrits maternels (en rouge) et paternels (en bleu) sont inversement affectés par les conséquences de la *trans*-interaction générée par la mutation *CLPG* (étoile). Ainsi, la présence de la mutation maternelle réduit le taux de transcrits paternels alors que la présence de la mutation paternelle renforce l'effet en *cis* de la mutation sur les transcrits maternels.

La *trans*-interaction ne concerne pas uniquement le niveau d'expression des messagers. En effet, Davis et al. (2004) ont démontré par immunohistochimie que la

protéine DLK1 est indétectable chez les animaux C^{mat}/C^{pat} (voir plus haut). Ainsi, malgré un niveau de transcrits élevé, la traduction de *DLK1* apparaît bloquée chez les homozygotes mutants. S'il semble évident que la *trans*-inhibition des messagers paternels est causée par l'un (ou plusieurs) des transcrits non codants exprimés maternellement, les mécanismes moléculaires restent à découvrir. L'hypothèse principale développée dans cette thèse consiste à considérer les miRNAs comme candidats idéals à la *trans*-inhibition (Chapitre 4).

1. Les microARNs : découverte et biosynthèse

Les microARNs (ou miRNAs) sont une famille d'ARNs non codants, de 21 à 24 nucléotides de longueur et dont la fonction est associée à la régulation post-transcriptionnelle et traductionnelle des ARN messagers. Leur découverte remonte à 1993, grâce à l'étude chez *C. elegans* d'un mutant incapable de franchir le stade larvaire (Lee et al. 1993 ; Wightman et al. 1993). Quinze ans plus tard, miRBase, base de données sur les miRNAs, dénombre 940 miRNAs chez l'humain (http://www.mirbase.org/)

Chez les plantes, la régulation par les miRNAs requiert une complémentarité de séquence parfaite entre le miRNA et son gène cible, ce qui conduit à la dégradation du messager. Chaque miRNA peut donc être associé aisément à son (ou ses) ARNm cible(s) par prédiction informatique. Chez les animaux, si les prédictions informatiques (utilisant notamment la conservation de la structure secondaire) et le séquençage haut débit des librairies de miRNAs générées à partir de nombreux tissus (Landgraf et al. 2007) ont permis l'identification de la majorité des miRNAs existant chez l'homme, la validation des messagers affectés est beaucoup plus difficile. En effet, les cas d'interactions parfaites sont très rares et diverses « règles » ont été déduites de l'observation *in vitro* de l'effet des miRNAs sur des gènes rapporteurs. Ainsi, les miRNAs seraient efficaces sur les 3' UTRs des messagers ciblés avec lesquels ils possèdent un appariement Watson-Crick parfait de 7 nucléotides (positions 2 à 8 du miRNA : la « graine »). Bien que la présence de cette graine formée avec la 3' UTR du messager semblait être une condition *sine qua non* de l'inhibition de la protéine, Tay et al. (2008) ont récemment rapporté qu'une graine imparfaite compensée par un appariement en 3' du miRNA, ciblant la partie codante d'un gène, pouvait aussi se solder

par une inhibition. On estime que chaque miRNA régulerait environ 200 messagers différents et que 20 à 30 % des gènes humains seraient ainsi affectés par un (ou plusieurs) miRNA(s).

Les miRNAs sont exprimés à partir d'un précurseur nommé pri-miRNA, transcrit majoritairement par l'ARN polymérase II. Ces pri-miRNAs, d'une taille minimale de 100 nucléotides, parfois polyadénylés, sont le plus souvent transcrits à partir d'un intron d'un gène ou sont concentrés en un « cluster » localisé dans un long transcrit intergénique (Ohler et al. 2004). Le pri-miRNA, toujours constitué d'une structure en épingle à cheveux, est clivé dans le noyau par une endonucléase de type III, DROSHA, qui ne conserve que la structure « tige-boucle » et définit l'une des extrémités du futur miRNA mature (Lee et al. 2003). Cette molécule intermédiaire, nommée pre-miRNA, est exportée dans le cytoplasme via les récepteurs Exportine-5/Ran-GTP (Lund et al. 2004) où elle est de nouveau clivée par une ribonucléase de type III, DICER. Il en résulte un duplex d'ARN double brin imparfait, dont l'un des brins (ou parfois les deux) donnera un miRNA mature en entrant en interaction avec le complexe RISC (RNA-Induced Silencing Complex). RISC, aussi appelé miRNP ou miRISC, contient notamment des protéines de la famille Argonautes (AGO 1 à 4 chez les mammifères), dont le rôle est essentiel à la fonction des miRNAs (cf Figure 10). Ce modèle initial de biosynthèse des miRNAs comporte des exceptions. En effet, les auteurs d'une étude récente ont découvert un cas de génération d'un miRNA (miR-451) chez la souris où le clivage en 3' n'est pas réalisé par la protéine Dicer mais directement par Argonaute 2 dans le RISC (Cheloufi et. al., 2010).

2. Les microARNs : mécanismes et fonctions

Les mécanismes impliqués dans l'inhibition de la traduction, souvent accompagnée d'une légère dégradation du messager, suscitent aujourd'hui de nombreuses discussions. Des arguments appuient un blocage au niveau de l'initiation ou de l'élongation de la traduction. D'autres auteurs décrivent la protéolyse du peptide naissant ou la déadénylation progressive du messager jusqu'à sa complète dégradation. Enfin, les miRNAs catalyseraient le transport de leur cible vers un compartiment virtuel de la cellule, nommé P-Body, où le

messager serait stocké ou dégradé (pour revue : Filipowicz et al. 2008). Des observations parfois contradictoires sur ces mécanismes laissent à penser que des phénomènes différents (mais aboutissant à une inhibition similaire de la traduction) pourraient être impliqués selon l'espèce, le tissu ou le miRNA concerné.

Des études commencent à catégoriser les miRNAs en fonction de l'action qu'ils exercent sur leurs cibles. Flynt et Lai (2008) distinguent trois classes d'interaction : (i) les miRNAs « interrupteurs » (Switch) qui répriment totalement leur cible, (ii) les miRNAs régulateurs (Tuning) à rôle plus faible, mais suffisamment important pour être conservés à travers les espèces et (iii) les miRNAs neutres (ni avantageux ni invalidants), dont l'évolution est par conséquent plus rapide. Certains miRNAs serviraient donc de déterminants clés pour la différenciation cellulaire alors que d'autres, affectant souvent un plus grand nombre de cibles, se contenteraient de renforcer l'identité tissulaire ou organique. Fait intéressant, si la séquence des miRNAs matures est souvent entièrement conservée à travers les espèces, de même que leurs fonctions physiologiques, les gènes ciblés par deux miRNAs orthologues sont généralement différents. Cette observation suggère donc une évolution rapide de ces interactions dans le règne animal.

L'impact réel des microARNs sur la régulation protéique est aujourd'hui sujet à controverse. Deux études indépendantes (Baek et al. 2008 ; Selbach et al. 2008) ont examiné par spectrométrie de masse l'effet de la transfection de miRNAs sur la traduction globale des protéines via une nouvelle méthode protéomique nommée SILAC (pour Stable Isotope Labelling with Amino acids in Cell culture) et consistant à utiliser des acides aminés radioactifs afin de différencier une même protéine avant et après traitement. Ces expériences ont conclu à un effet relativement faible des miRNAs sur la production de protéines (au maximum un facteur 4 en différence d'abondance), y compris pour les miRNAs possédant une graine parfaitement complémentaire aux 3' UTRs de leurs cibles, ce qui suggère que les miRNAs sont vraisemblablement de fins régulateurs de l'expression génique. Parallèlement à ces observations, Tay et al. (2008) ont observé une inhibition de plus de 50 % de l'expression protéique d'un gène *(Nanog*) ciblé par un miRNA formant un

appariement avec deux G : U dans la graine (en principe très néfaste pour l'efficacité des miRNAs) et ciblant la partie codante de ce gène. L'éventail des interactions possibles semble donc très difficile à estimer et une étude empirique au cas par cas se révèle souvent nécessaire.

Les miRNAs sont donc incontestablement l'une des découvertes essentielles de la dernière décennie, permettant notamment des traitements thérapeutiques d'une précision inégalée. Cependant, les mécanismes précis de leur mode d'inhibition (voire même d'activation dans les quelques rares cas rapportés) et l'étendue réelle de leur contribution in vivo restent encore à éclaircir. Il est aujourd'hui assez difficile de prédire un ensemble de règles fonctionnelles s'appliquant à l'ensemble des miRNAs, qui paraissent varier selon l'espèce, le type cellulaire ou simplement d'un miRNA à l'autre.

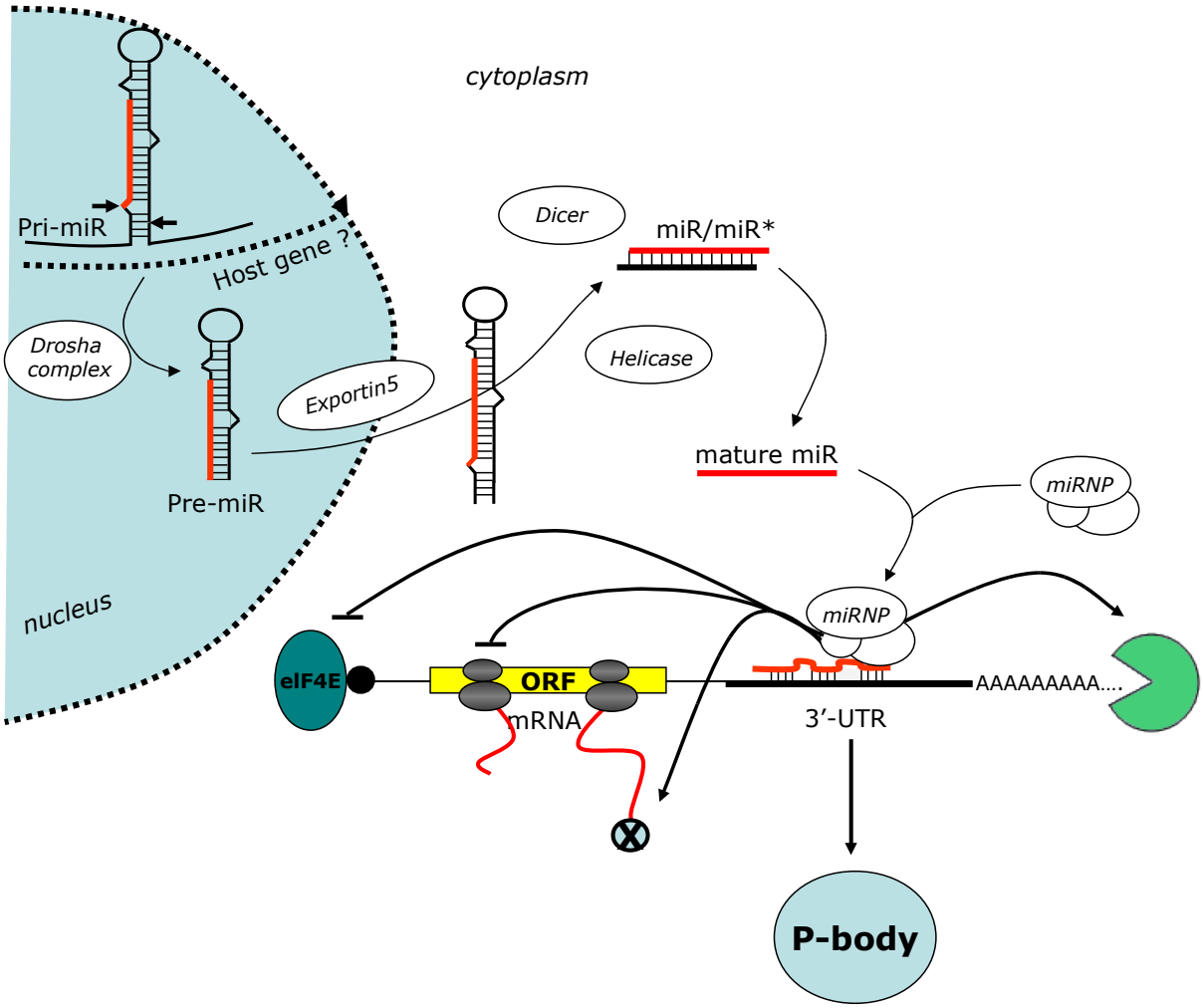


Figure 10 : Biosynthèse des miRNAs et principales voies de régulation de la traduction des messagers.

Encadré 2 : Édition des ARNs

On nomme « édition » tous les mécanismes post-transcriptionnels conduisant à modifier la séquence d'un ARN, que ce soit par insertion, délétion ou conversion nucléotidique. Après édition, la séquence du transcrit ne correspond donc plus entièrement à celle encodée par l'ADN génomique du gène correspondant. Le premier cas observé, impliquant la délétion de plusieurs nucléotides dans les transcrits mitochondriaux de trypanosomes, remonte à plus de 20 ans (Benne et al., 1986). L'édition d'ARN a cours dans tous les organismes et englobe un grand nombre de modifications différentes.

Chez les mammifères, l'édition majoritaire est la conversion d'une adénosine en inosine (A vers I). Cette modification peut survenir aussi bien dans les ARN messagers que dans les ARN non codants (notamment les ARNs de transfert). Reconnue comme une guanine par les mécanismes cellulaires, l'inosine introduite peut avoir diverses conséquences, comme la modification d'un acide aminé, la perturbation d'un site d'épissage ou la création d'un codon stop.

L'édition de l'adénosine en inosine est catalysée par les protéines ADARs (Adenosine Deaminase Acting on RNA). Il existe trois protéines ADAR conservées chez les vertébrés. ADAR1 et ADAR2 sont exprimées dans de nombreux tissus, contrairement à ADAR3, dont l'activité n'a jamais été démontrée et qui semble spécifique au cerveau (Bass and Weintraub 1988 ; Wagner et al. 1989). Une lignée de souris transgénique porteuse d'une délétion d'*Adar1* présente un phénotype létal, démontrant ainsi l'importance cruciale de l'édition sur le développement embryonnaire (Wang et al. 2000). Récemment, il a été découvert que l'édition A vers I est beaucoup plus courante chez l'homme que chez les autres primates (Paz-Yaacoov, 2010).

La sélectivité et spécificité de substrat des ces protéines sont assez mal connues. Afin d'être édité, un ARN doit présenter une structure double brin. Dans le cas d'un long (>100 pb) ARN double brin parfaitement complémentaire, de nombreuses adénosines seront globalement éditées, tandis que les structures plus courtes et imparfaites verront seulement certaines adénosines spécifiques éditées. Toutefois, les déterminants de structure permettant l'édition d'une adénosine en particulier restent inconnus à ce jour. De même, il est impossible de prédire si une structure donnée sera éditée ou non.

L'analyse systématique des banques de données d'EST a permis de déterminer que l'édition est un phénomène très rare dans les séquences codantes. La majorité des séquences éditées correspond en effet à des éléments répétés (notamment les séquences Alu chez les primates). C'est pourquoi certains auteurs postulent que l'édition pourrait être un mécanisme de défense contre la propagation de ces éléments (Kim et al. 2004 ; Levanon et al. 2004).

Récemment, les précurseurs des miRNAs (cf miRNAs) ont été ajoutés aux substrats potentiels des protéines ADAR. Dans certains cas, l'édition de précurseurs pourra interférer avec le processus de maturation du miRNA (modification du site de clivage de DROSHA ou DICER) et se terminer par la dégradation du précurseur (Yang et al. 2006). Si la modification touche la graine, le miRNA mature ne pourra plus réguler ses cibles légitimes, mais gagnera une affinité pour d'autres transcrits (Kawahara et al. 2007). Les protéines ADAR sont donc en compétition de substrat avec la machinerie d'interférence ARN et pourraient agir en inhibiteurs spécifiques de certains précurseurs.

V. Objectif de thèse

La surdominance polaire gouvernant la transmission du phénotype callipyge est le résultat d'une interaction complexe entre (i) un effet en *cis* de la mutation *CLPG* sur le niveau de transcription des gènes du locus *DLK1-GTL2* et (ii) une *trans*-interaction entre les allèles paternels et maternels. Si les manifestations observables de ces effets sont aujourd'hui bien décrites, les mécanismes moléculaires sous-jacents sont pour la plupart inconnus.

L'objectif de mes travaux consistera donc en l'exploration des ces mécanismes afin de parvenir à une meilleure compréhension de l'établissement du phénomène de surdominance polaire chez le mouton callipyge. Dans un premier temps, j'aborderai les effets en *cis* du SNP^{CLPG} sur le profil épigénétique de la région (Chapitre 2, en collaboration avec Haruko Takeda). Puis je traiterai du rôle des microARNs dans l'élaboration de l'effet en *trans*, au niveau de *PEG11* (Chapitre 3) et de *DLK1* (Chapitre 4).

Mes travaux sur les miRNAs m'ont en outre conduit à collaborer à un projet de recherche sur une autre hypertrophie musculaire, affectant le mouton Texel, dont la mutation causale implique directement l'inhibition d'un gène par des miRNAs. Ces travaux touchant à des thèmes proches de ceux abordés chez le mouton callipyge, la publication correspondante est jointe en annexe de cette thèse.

CHAPITRE II

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

Takeda H, Caiment F, Smit M, Hiard S, Tordoir X, Cockett N, Georges M, Charlier C. 2006. PNAS 103(21): 8119-8124.

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

‡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

SUMMARY

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.

INTRODUCTION

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*, *anti-PEG11*, *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).

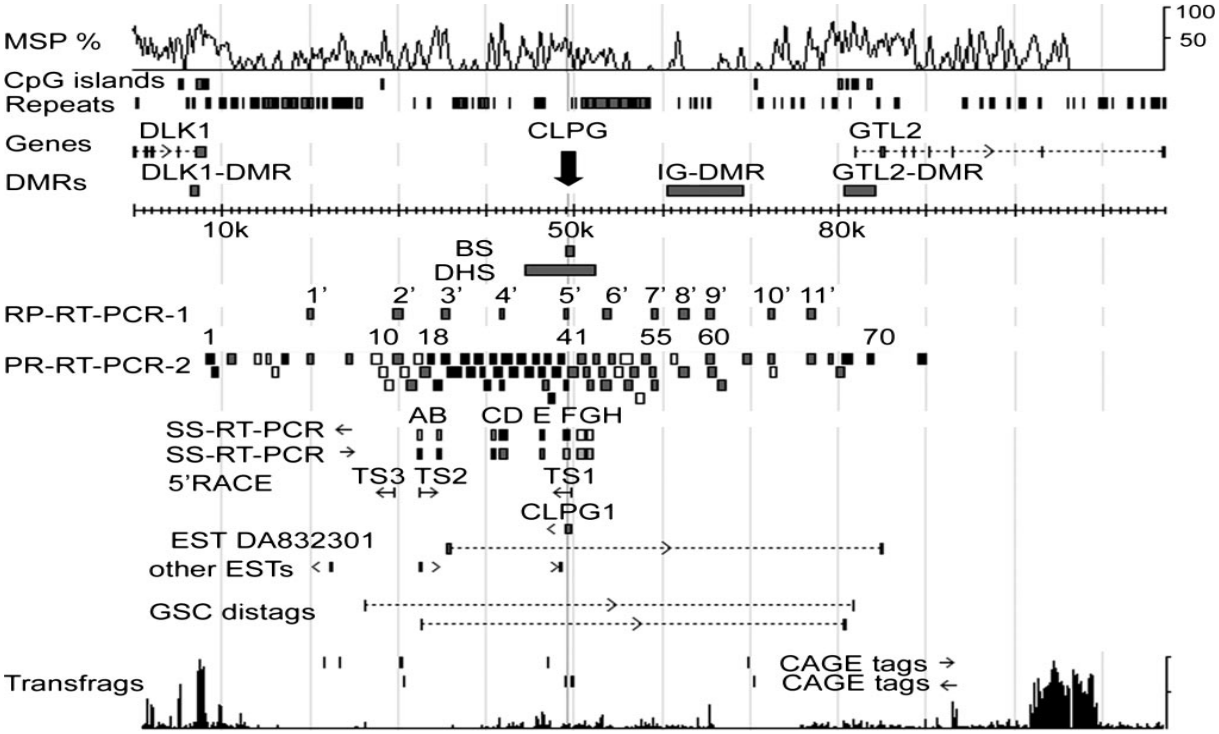


Figure 1 Schematic representation of the ovine *DLK1-GTL2* IG region. MSP %, multispecies similarity profile; DMRs, differentially methylated regions; BS, bisulfite-sequenced segment (Fig. 2); DHS, segment explored for DNase-I hypersensitive sites (Fig. 3); RP-RT-PCR and SS-RT-PCR, location of amplicons 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).

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).

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.

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). The conversion rate for non-CpG C residues averaged 99.8%, demonstrating the efficacy of the bisulfite treatment.

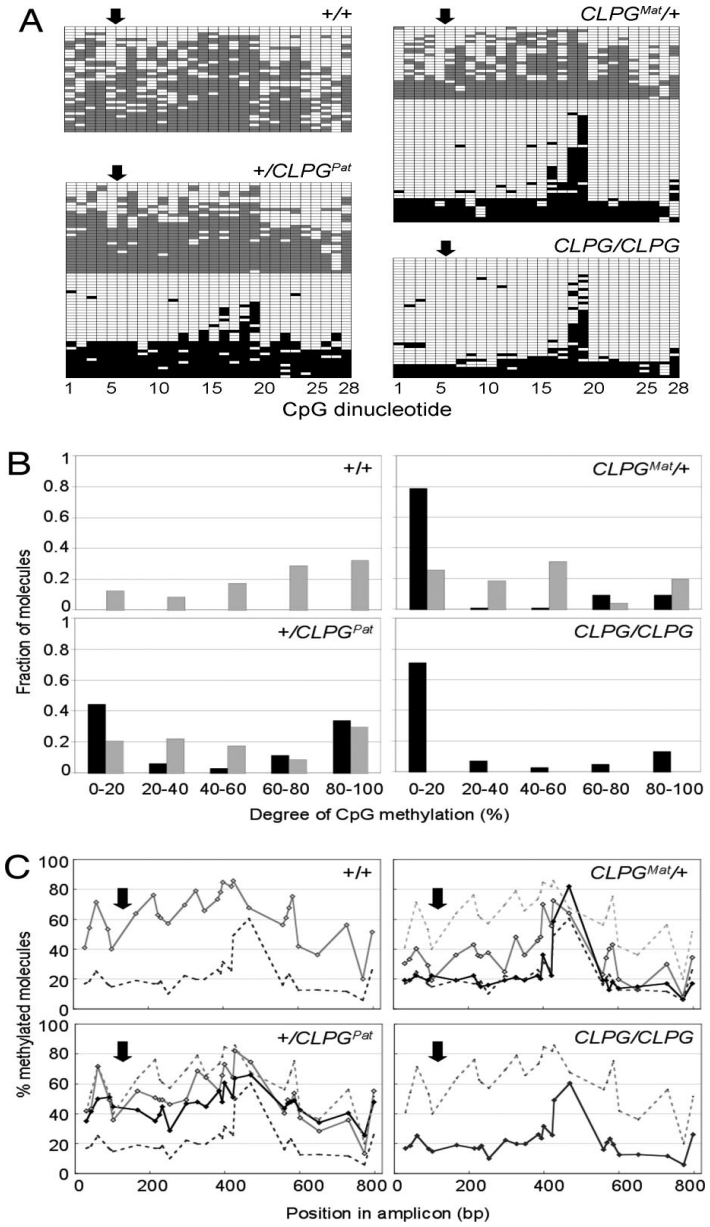


Figure. 2. DNA methylation analysis. (*A*) Representative bisulfite sequencing results for a 828-bp amplicon spanning the *CLPG* mutation for animals rep- resenting the four *CLPG* genotypes. Each line corresponds to a distinct mole- cule, each column to one of the 28 CpG dinucleotides in the amplicon. 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* muta- tions is marked by the arrow.

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 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 heterozygotes 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*.

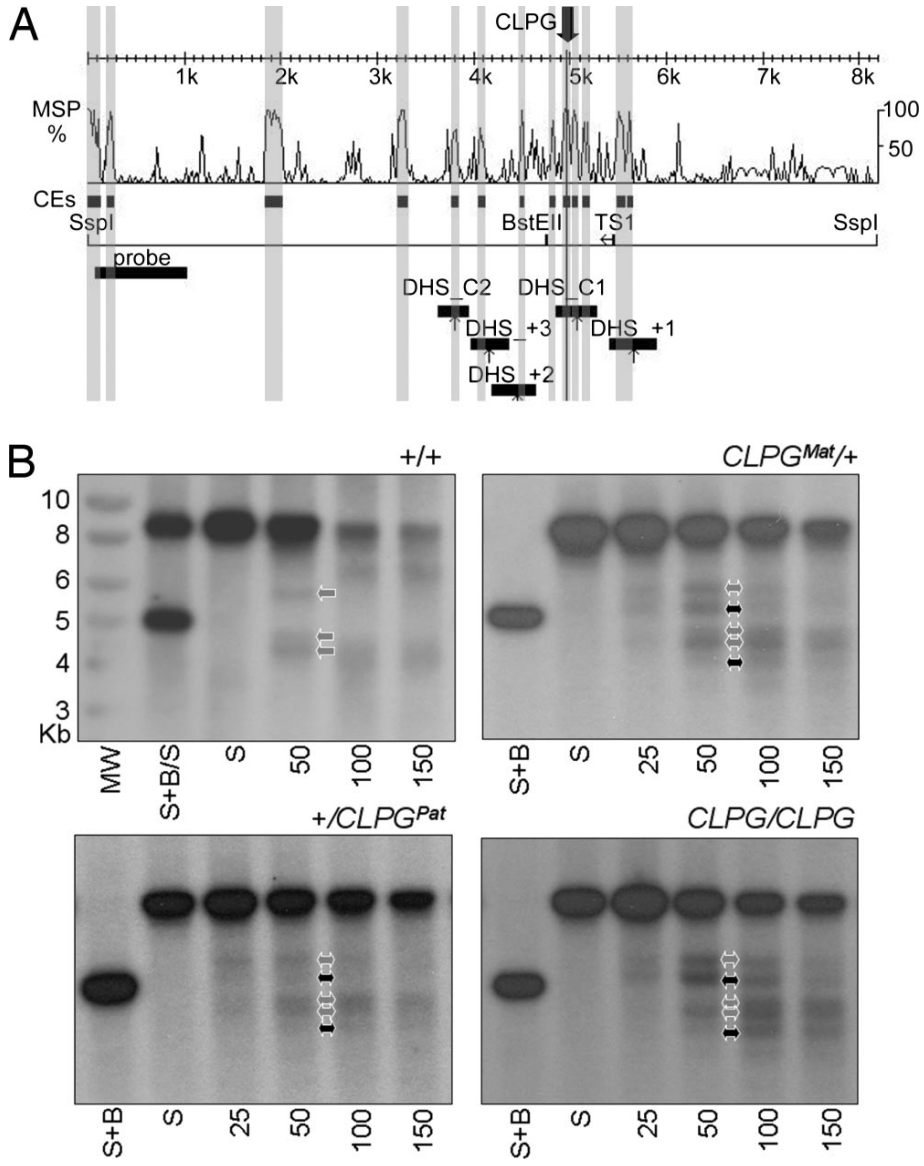


Figure. 3. DNase-I hypersensitive analysis. (*A*) Schematic representation of the 8.2-kb SspI fragment analyzed for the presence of DHS, showing (*i*) the position of the *CLPG* mutation (black arrow and vertical line), (*ii*) a multispecies (human, mouse, and ovine) similarity profile (MSP), (*iii*) "*PhastCons*" con- served 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.

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). 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*)←*G*(*tl2*) transcripts] in *+/+* fetuses, albeit at low levels. The same low level *D*←*G* transcripts also were detectable in 8-week-old *+/+* animals. In addition, we obtained RT-PCR products corresponding to antisense *D*→*G* transcripts from the prenatal *+/+* samples at extremely low levels (Fig. 4 *A*).

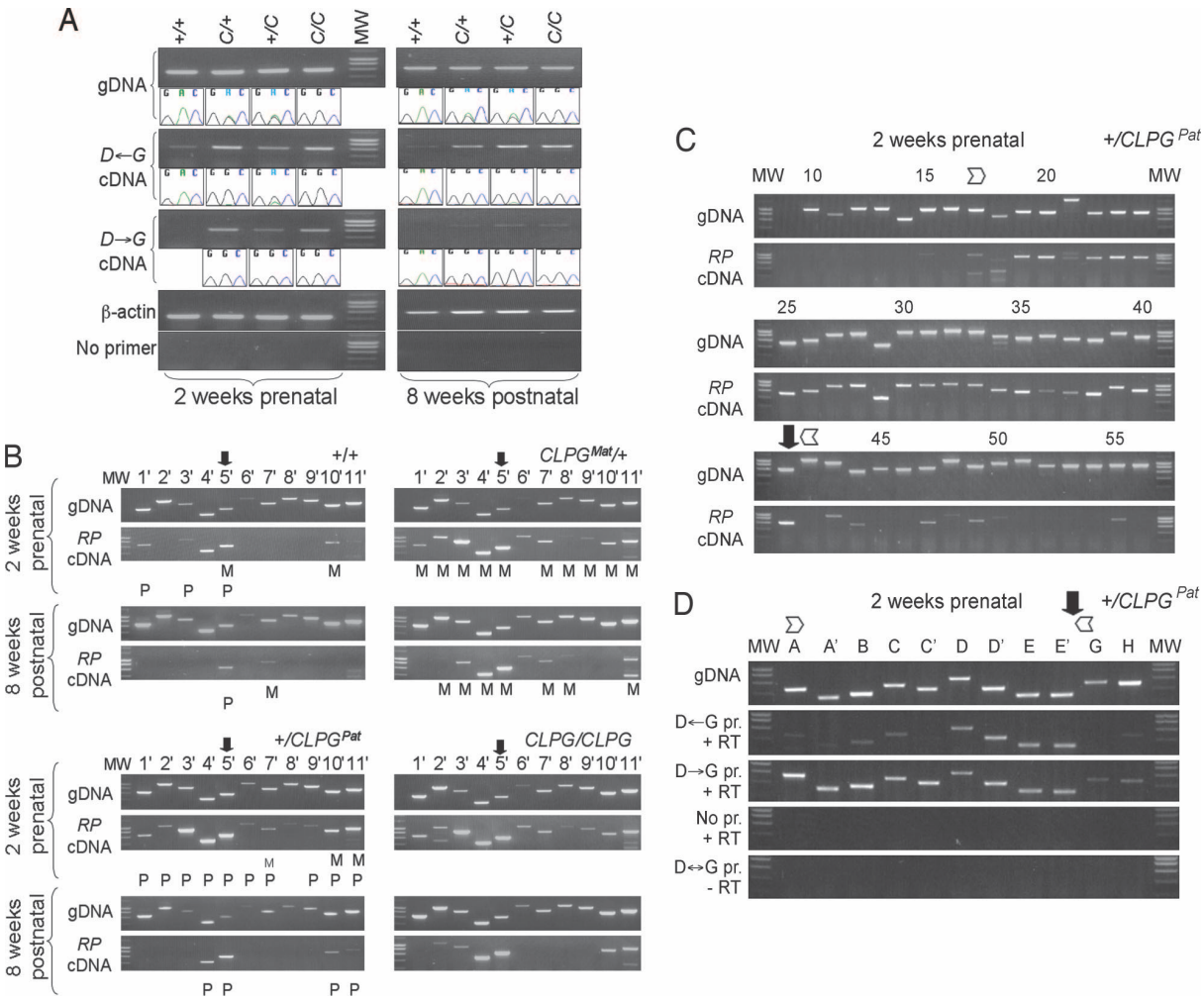


Figure. 4. Expression analysis of IG transcripts. The position of the *CLPG* mutation is marked by the black arrow. The white arrows correspond to the positions of the TS2 and TSl transcription start sites, respectively. MW, molecular weight marker. (*A*) Results of strand-specific RT-PCR experiments by using a 593-bp amplicon spanning the *CLPG* site and *g!uteus 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, *g!uteus medius* cDNA synthesized by using either one (specific for *D4G* transcripts) or the other (specific for *D3G* transcripts) primer, and RT-treated cDNA in the absence of primers. A 428-bp -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 **II** amplicons spanning the *DLK1-GTL2* IG region (**I'—II'** in Fig. **I**) by using random primed *g!uteus 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. **I**) 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 *D4G* transcripts) or the other (specific for *D3G* transcripts) primer, RT-treated cDNA in the absence of primers, and RT minus RNA with both primers.

More remarkably, when compared with *+/+* animals, we observed a strong increase in the yield of *D←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→G* products, albeit more modest. Sequencing the corresponding $C^{Mat}/+$ and $+/C^{Pat}$ amplicons indicated that both *D←G* and *D→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←G* and *D→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 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). 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). 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, genomewide 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). 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.

Acknowledgments

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.

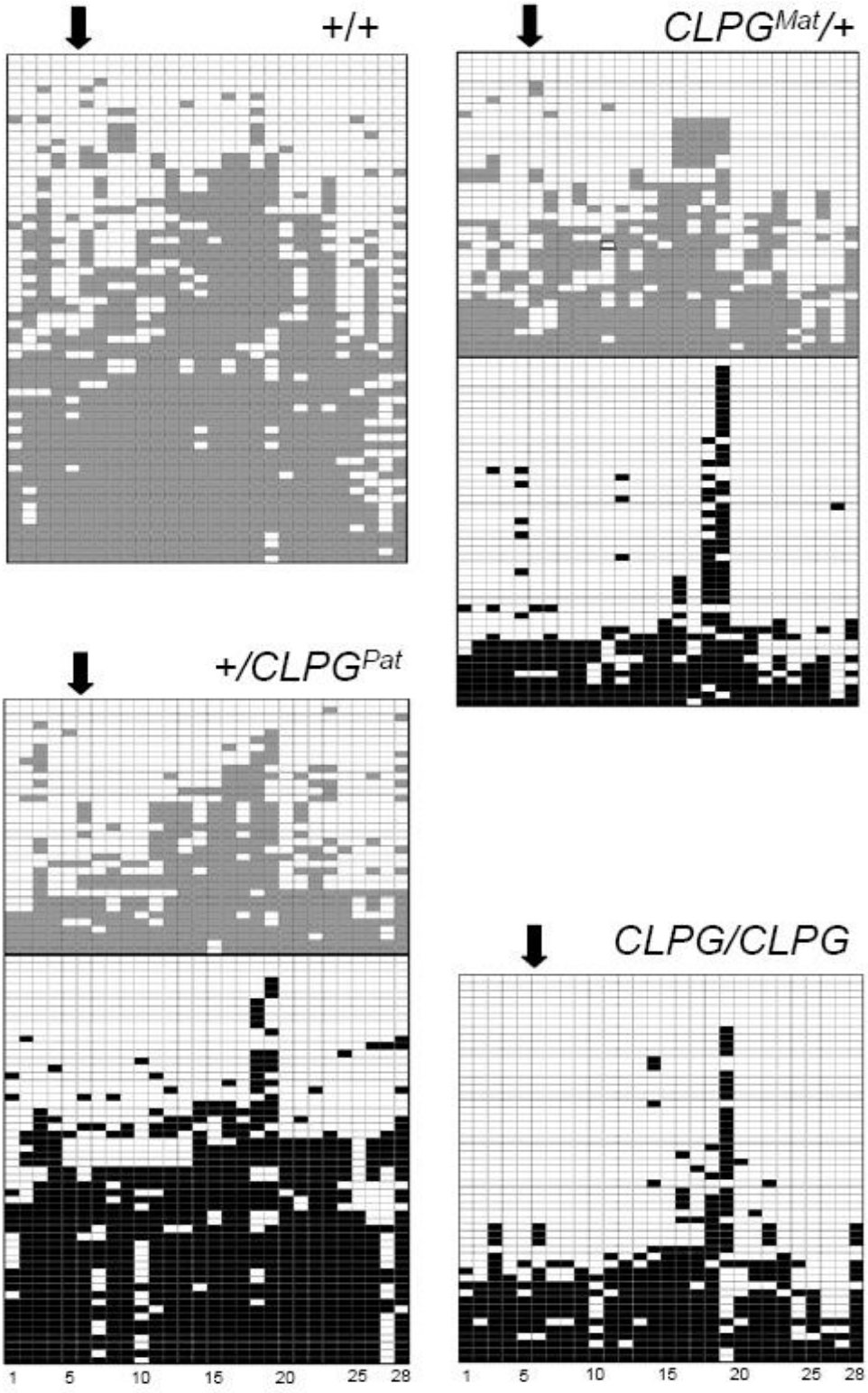
References

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.

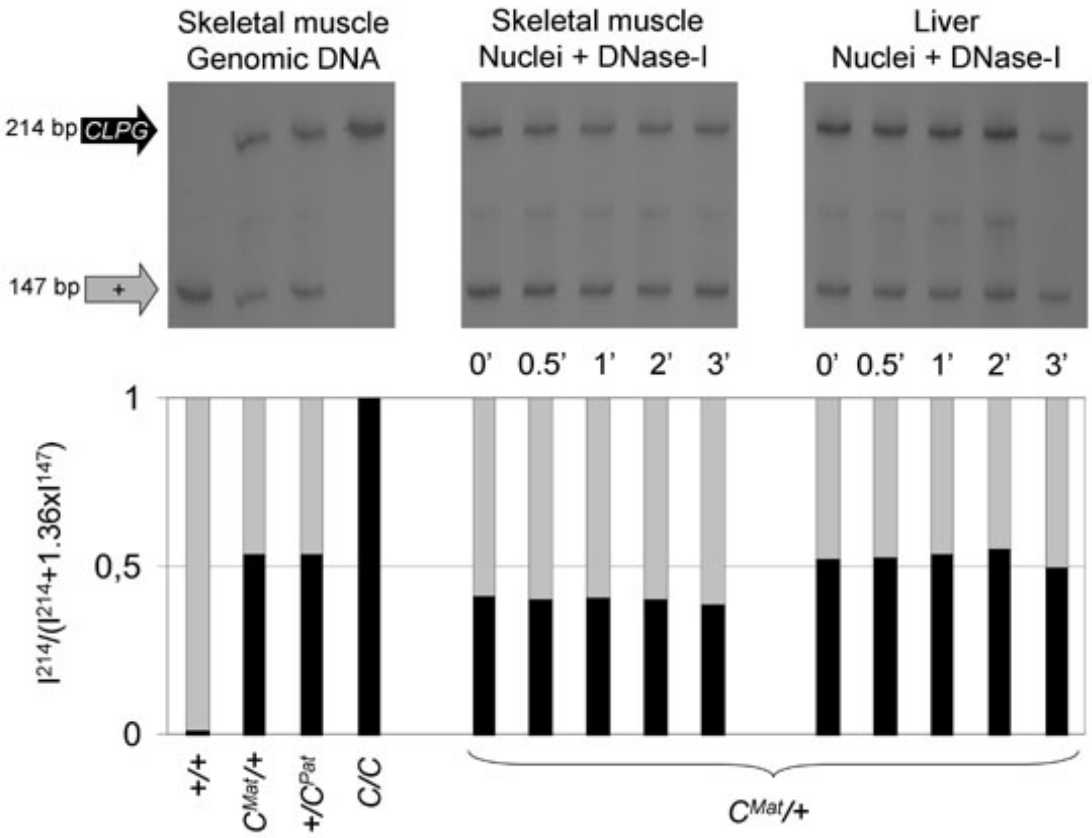
6. Charlier, C., Segers, K., Karim, L., Shay, T., Gyapay, G., Cockett, N. & Georges, M. (2001) *Nat. Genet.* 27, 367—369.
7. Murphy, S. K., Freking, B. A., Smith, T. P., Leymaster, K., Nolan, C. M., Wylie, A. A., Evans, H. K. & Jirtle, R. L. (2005) *Mamm. Genome* 16, 171—183.
8. Charlier, C., Segers, 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.

SUPPLEMENTAL DATA

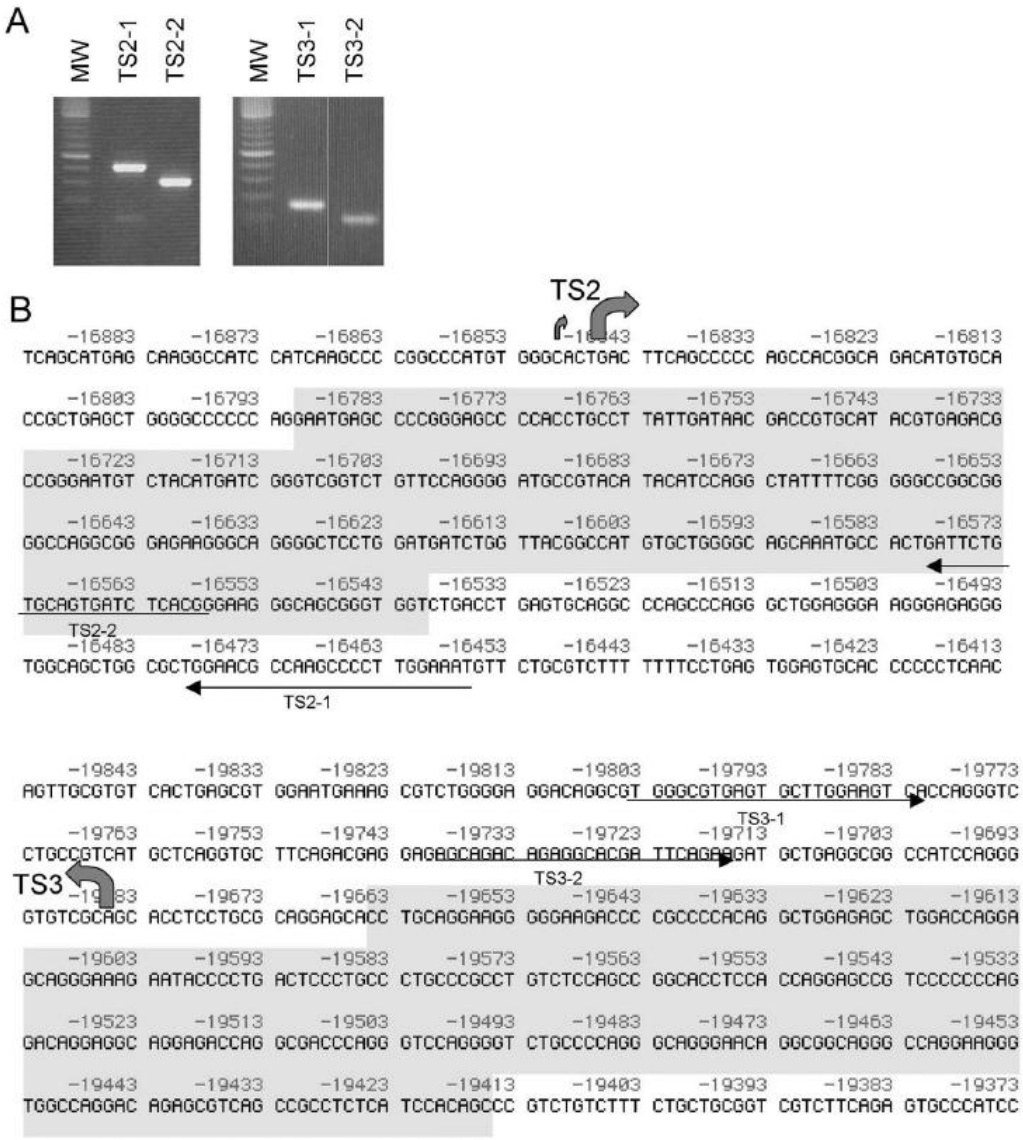
Supplemental Figures



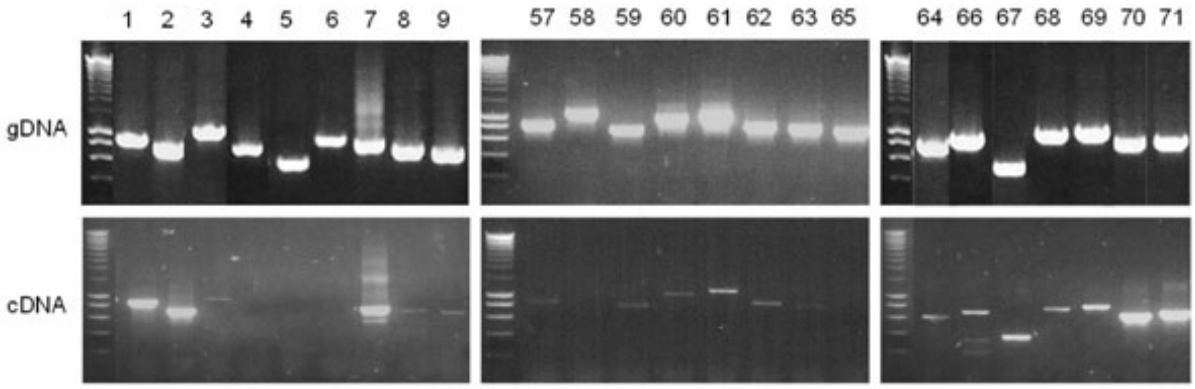
Supporting Figure 5. DNA methylation analysis. Bisulfite sequencing results are shown for an 828-bp amplicon spanning the *CLPG* mutation and for the other four animals representing the possible genotypes at the *CLPG* locus: *+/+, C^{Mat}/+, +/C^{Pat},* and *C/C* (data for the first set of animals is shown in Fig. 2*A*). 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.



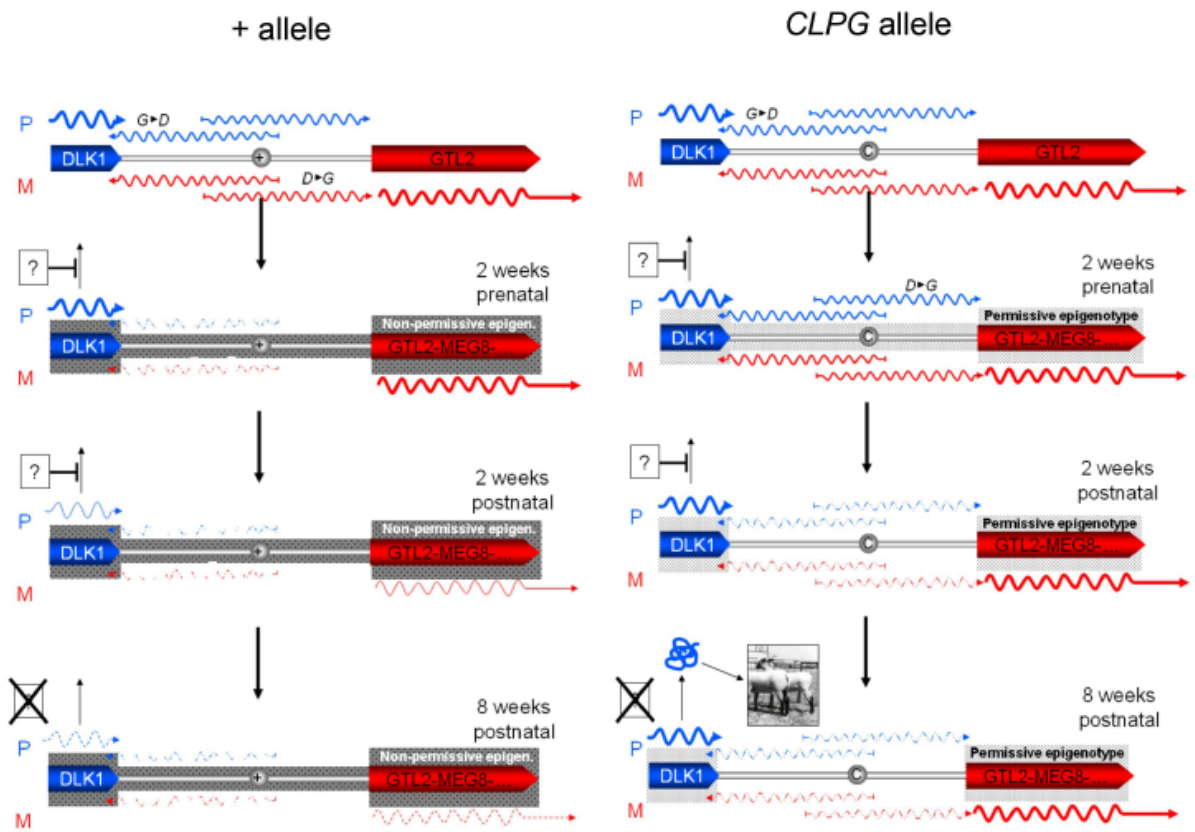
Supporting Figure 6. Demonstration of an increased general sensitivity to DNase-I of the *CLPG* allele in skeletal muscle but not in liver. (*Upper*) The results of hot-stop PCR-RFLP analysis of a ^{32}P-dCTP-labeled 361-bp amplicon spanning the *CLPG* mutation by using DNA extracted from C^{Mat}/+ nuclei (skeletal muscle and liver) treated for increasing incubation times (0', ½', 1', 2', and 3') with 500 units/ml DNase-I. The black arrow points toward the AvaII undigested 214-bp fragment corresponding to the *CLPG* allele; the gray arrow to the 147-bp AvaII digestion product corresponding to the + allele. (*Lower*) The intensity ratio of the 214-bp and 147-bp fragments. To approach the estimation of an allelic ratio, the intensity of the 147-bp band is multiplied by 1.36 because it contains 1.36 times less C residues than the 214-bp fragment. We analyzed genomic DNA extracted from skeletal muscle of animals representing the four possible *CLPG* genotypes as control. It can be seen that the *CLPG* to + allelic ratio is the same in DNase-I treated liver nuclei as it is in genomic DNA of heterozygous animals; in DNase-I treated skeletal muscle nuclei, however, the *CLPG* to + allelic ratio is clearly decreased.



<u>Supporting Figure 7.</u> 5' RACE detection of putative transcription start sites in the *DLK1-GTL2* intergenic region. (*A*) PCR products obtained by 5' RACE corresponding to transcription starts sites TS2 and TS3. Lanes labeled TS2-1 and TS2-2 or TS3-1 and TS3-2 correspond to products obtained by using the primers labeled accordingly in *B* in the nested PCR round. (*B*) The precise locations of TS2 and TS3 as determined by direct sequencing of the 5' RACE products are marked by arrows. The shaded areas correspond to Proscan promoter predictions.



Supporting Figure 8. Probing intergenic transcription between *DLK1* and position –16,846 and between position +478 and *GTL2*. Results of PCR amplification from genomic DNA (gDNA) and random-primed skeletal muscle (*gluteus medius*) cDNA (cDNA) of amplicons located in the *DLK1-GTL2* IG region between *DLK1* and position –16.8 Kb (amplicons 1–9) and between position +478 bp and *GTL2* (amplicons 57–71). The position of the amplicons is as indicated in Fig. 1 (RP-RT-PCR-2).



Supporting Figure 9. Working model for the mode of action of the *CLPG* mutation. (*Left*) A muscle-specific *DLK1-GTL2* intergenic silencer down-regulates *D®G* and *D¬G* intergenic transcription, which causes progressive down-regulation of *DLK1* and *GTL2*. The corresponding nonpermissive chromatin status is maintained epigenetically. (*Right*) Inactivation of the intergenic silencer by the *CLPG* mutation maintains high levels of intergenic transcription in skeletal muscle throughout later stages of development, thereby establishing a permissive chromatin status that is epigenetically maintained, thus allowing high levels of *DLK1* and *GTL2* transcription in this tissue. Release of an as-of-yet-unidentified translational blocking mechanism after birth results in ectopic expression of DLK1 protein and, hence, to the callipyge phenotype in specific muscles of the

hindquarter in *+/C*^{Pat} individuals. In *C/C* individuals, trans inhibition mediated by the madumnal noncoding RNAs prevents this release.

Supporting Materials and Methods

Bisulfite Sequencing

Genomic DNA was extracted from skeletal muscle (*gluteus medius*) of two 8-week-old animals of each callipyge genotype by using standard phenol/chloroform extraction and isopropanol precipitation procedures. The DNA was digested with NsiI and subjected to a bisulfite modification using the CpGenome DNA modification kit (Chemicon International, Temecula, CA). An 828-bp DNA fragment encompassing the *CLPG* mutation and TS1 was amplified by using primers 5'-GGTTGTTTAGAGAGGTTAGATGTTGT-3' and 5'-TAAAACCTAAAATACCCTCCACACT-3' that were designed with the help of METHPRIMER (www.urogene.org/methprimer). PCR was performed in a volume of 120 ml with 240 ng of the bisulfite-treated genomic DNA, 0.6 mM of each primer, 200 mM of each dNTP, 5 units of HotStartTaq DNA polymerase (Qiagen, Valencia, CA), and 1× PCR buffer supplied with the polymerase. PCR conditions were as follows: an initial denaturation of 95°C for 15 min; 10 cycles of 94°C for 45 s, 60°C for 30 s, and 72°C for 2 min; 33 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 2 min; and a final extension of 72°C for 10 min. The PCR product was subcloned by using the TA cloning kit (Invitrogen) to analyze allele specific methylation patterns of multiple DNA molecules. At least 48 clones from each genotype were subjected to a colony-PCR by using M13 forward and reverse primers, and the PCR products that contained the expected 828-bp inserts were sequenced by using the same primers with the BigDye Terminator v3.1 Cycle Sequencing Kit and the 3100 Genetic Analyzer (Applied Biosystems).

DNase-I Hypersensitivity

DNase-I hypersensitive assay was done by using nuclei isolated from skeletal muscle (*gluteus medius*) of two animals of each callipyge genotype and liver of *+/C*^{Pat}(*n* = 1) and *C*^{Mat}*/+* (*n* = 1) animals at 8-week-old according to procedures described by Gregory *et al.* (1) with some modifications. Briefly, »4 g of tissue that had been preserved at –80°C was crushed in

liquid nitrogen by using a mortal and pestle. The resulting powder was immersed in 30 ml of ice-cold 0.3 M sucrose buffer: 0.3 M sucrose/60 mM KCl/15 mM NaCl/5 mM $MgCl_2$/0.1 mM EGTA/0.5 mM DTT/15 mM Tris·HCl (pH 7.5)/2 ml/ml protease inhibitor mixture (Sigma-Aldrich), homogenized in a Potter-Elvehjem homogenizer, and filtered through a fine gauze to remove debris. The cell suspension was centrifuged at 700 × *g* for 5 min at 4°C. The pellet was resuspended in 12 ml of the 0.3 M sucrose buffer and filtrated again by using cell strainers (pore size 100 and 40 mm, BD Biosciences). Four milliliters of the cell suspension was layered on 12.5 ml of ice-cold 1.2 M sucrose buffer and centrifuged at 10,000 × *g* for 20 min at 4°C. The nuclei pellet was resuspended in 4 ml of the 0.3 M sucrose buffer without the protease inhibitor mixture. Then, 800 ml of the nuclei suspension was mixed with 200 ml of the 0.3 M sucrose buffer containing each 0, 25, 50, 100, and 150 units of DNase-I (Roche, Penzberg, Germany) and incubated at 25°C for 15 min. The digestion was stopped by adding 1 ml of stop buffer: 20 mM EDTA (pH 8)/1% SDS, and incubated with 200 mg/ml proteinase K overnight at 50°C. DNA was purified twice with phenol-chloroform and once with chloroform and precipitated with isopropanol. Ten micrograms of the DNase-I treated DNA was digested with SspI. Control genomic DNA was digested with SspI or SspI/BstEII. The resulting fragments were separated on a 0.8% agarose gel, transferred to Hybond *N* + membrane (Amersham Pharmacia Biosciences, Little Chalfont, U.K.) under neutral conditions, and hybridized with a ^{32}P-labeled probe at 55°C overnight in ULTRAhyb buffer (Ambion, Austin, TX). The probe was generated by PCR from ovine genomic DNA with a primer set: 5'-GGATCAGCCACCCATACAGTTC-3' and 5'-CAGGCTGTGCTTTCTGATCAGG-3', gel-purified with the Geneclean kit (Qbiogene, Irvine, CA), random-labeled by using the Random Primers DNA Labeling System (Invitrogen), and purified with the ProbeQuant G50 micro column (Amersham Pharmacia Biosciences). The membrane was washed once with 2´ SSC, 0.1% SDS at room temperature for 10 min, twice with 2´ SSC, 0.1% SDS at 65°C for 15 min, and three times with 0.1´ SSC, 0.1% SDS at 65°C for 15 min, and exposed to an x-ray film for 3–7 days. The size of each DNase-I digested band was calculated by using Duggleby's DNA Size Mapper (2). The position of each DHS was shown as a range from minimum to maximum data or as a point on the average of data that were obtained from several experiments.

DNase-I General Sensitivity

DNase-I sensitive assay was done according to procedures described by Gregory and Feil (3) by using nuclei isolated from skeletal muscle (*gluteus medius*) of $C^{Mat}/+$ (n = 1) and liver from $C^{Mat}/+$ (n = 1) and $+/C^{Pat}$ (n = 1) animals at 8-week-old. Briefly, about 4×10^3 nuclei were permeabilized in a volume of 50 ml with the 0.3 M sucrose buffer containing 0.2% Nonidet P-40 on ice for 5 min and then incubated with 500 units/ml of DNase-I at 25°C for various time (0, 0.5, 1, 2, and 3 min). After heat inactivation of the DNase-I at 95°C for 1 h, the sample was treated with 200 mg/ml of proteinase K at 50°C overnight and heat-inactivated at 95°C for 1 h. Control genomic DNA was extracted from skeletal muscle of the same animals by using standard phenol/chloroform extraction and isopropanol precipitation procedures. A "hot stop" PCR-restriction fragment length polymorphism (RFLP) procedure was used to differentiate the *CLPG* and + allele (4). A 214-bp DNA fragment was PCR amplified in a volume of 30 ml, with 3 ml of the DNase-I treated sample or 60 ng of control genomic DNA, 0.25 mM of each primer encompassing the *CLPG* mutation site: 5'-TGTCCTGGTCTATTTTCGGGC-3' and 5'-GCAAGGGTCTGTTTGGTCCTAA-3', 200 mM of each dNTP, 1.5 mM $MgCl_2$, 0.2 ml of AmpliTaq Gold DNA polymerase (Applied Biosystems), and 1× PCR buffer supplied with the polymerase. PCR conditions were as follows: an initial denaturation of 95°C for 10 min; 33 cycles (for the DNase-I treated sample) or 28 cycles (for the control genomic DNA) of 94°C for 30 s, 58°C for 30 s, and 72°C for 1 min; and a final extension of 72°C for 10 min. Then, 10 ml of the primary PCR product was mixed with 15 ml of a hot-stop PCR mixture containing 0.25 mM of the same primers, 200 mM of each dNTP, 1.5 mM $MgCl_2$, 0.1 ml of the polymerase, and 0.1 uCi of ^{32}P-dCTP (Amersham Pharmacia Biosciences). The reaction was performed with 95°C for 15 min, 58°C for 1 min, and 72°C for 15 min. The hot-stop PCR product was digested with AvaII that discriminated between the *CLPG* (214 bp) versus + allele (147 bp) and resolved on an 8% polyacrylamide gel. The band intensities were quantified by using the Molecular Imager FX system (Bio-Rad).

RT-PCR

Total RNA was extracted from skeletal muscle (*gluteus medius*) of two animals of each callipyge genotype at 2 weeks before birth and 8 weeks after birth by using TRIzol reagent (Invitrogen). Contaminating DNA was removed with the TURBO DNA-free kit (Ambion).

For strand-specific RT-PCR, cDNA was synthesized from 3.5 mg of total RNA with a 0.1 mM strand-specific primer by using the EndoFree RT kit (Ambion) to minimize primer-independent cDNA synthesis. Mock RT-PCR was performed with cDNA that was synthesized without added primer to detect contamination of cDNA caused by the endogenous priming and without reverse transcriptase to detect contamination of genomic DNA. The primer sequences for cDNA synthesis of the *D®G* and *D¬G* transcripts and following PCR amplification of the A, B, C, D, E, F, G, and H amplicons (SS RT-PCR in Fig. 1), and b-actin are in a Table 1. The PCR was performed in a volume of 30 ml with 3 ml of the cDNA or 2 ng of control genomic DNA, 0.2 mM of each primer, 200 mM of each dNTP, 0.2 ml of HotStartTaq DNA polymerase (Qiagen), and 1× PCR buffer supplied with the polymerase. PCR conditions were as follows: for *D®G* and *D¬G* transcripts, an initial denaturation of 95°C for 15 min; 39 or 40 cycles of 94°C for 30 s, 61°C for 30 s, and 72°C for 90 s; and a final extension of 72°C for 10 min; for b-actin, annealing temperature at 53°C and 26 PCR cycles were used. Half of the PCR product was used for detection by an electrophoresis on a 1.5% agarose gel with ethidium bromide. The rest of the PCR product was purified by using the Multiscreen PCR u96 filter plate (Millipore, Billerica, MA) and sequenced using the same primers to detect the parental origin of transcripts when polymorphism was available.

For random primed RT-PCR (RP-RT-PCR in Fig. 1), cDNA was synthesized from 10 mg of total RNA in a volume of 50 ml with 125 ng of random hexamers with or without reverse transcriptase by using the SuperScript-III First-strand synthesis system (Invitrogen). PCR was performed in a volume of 30 ml with 3 ml of the cDNA or 60 ng of control genomic DNA, 0.2 mM of each primer, 200 mM of each dNTP, 0.2 ml of HotStartTaq DNA polymerase, and 1× PCR buffer supplied with the polymerase. PCR conditions were as follows: an initial denaturation of 95°C for 15 min; 42 cycles (for cDNA) or 33 cycles (for genomic DNA) of 94°C for 30 s, 61°C for 30 s, and 72°C for 2 min; and a final extension of 72°C for 10 min. The corresponding primer sequences are in Table 1.

5' RACE

To gain additional information of transcription start sites, 5' RACE was performed by using the GeneRacer kit (Invitrogen) according to the manufacturer's protocol. Briefly, 5 mg of total RNA from skeletal muscle (*gluteus medius*) of *+/C*^{Pat} (*n* = 1) and *C*/*C* (*n* = 1) animals at 2 weeks before birth was dephosphorylated to eliminate 5' phosphates from truncated mRNA and non-mRNA, treated with tobacco acid pyrophosphatase to remove the 5' cap structure from intact mRNA, ligated the GeneRacer RNA oligo to the 5' end of the mRNA, and reverse transcribed by using intergenic transcript-specific primers; 5'-ATTTCCAAGGGGCTTGGCGTTCCA-3' for TS2, and 5'-TGGGCGTGAGTGCTTGGAAGTCA-3' for TS3. These primers were designed to target regions predicted by Promoter Scan (http://thr.cit.nih.gov/molbio/proscan). To amplify 5' ends of the cDNA, the first PCR was performed by using GeneRacer 5' primer (homologous to the GeneRacer RNA oligo) and the same transcript-specific primer for the cDNA synthesis in a volume of 30 ml with 1 ml of the cDNA, 0.6 mM of the GeneRacer 5' primer, 0.2 mM of the transcript-specific primer, 200 mM of each dNTP, 2 mM $MgCl_2$, 0.2 ml of AmpliTaq Gold DNA polymerase, and 1× PCR buffer supplied with the polymerase. PCR conditions were as follows: an initial denaturation of 95°C for 5 min; 8 cycles of 94°C for 30 s and 72°C for 1 min; 8 cycles of 94°C for 30 s and 70°C for 1 min; 25 cycles of 94°C for 30 s, 63°C for 30 s, and 72°C for 1 min; and a final extension of 72°C for 10 min. The nested PCR was performed by using GeneRacer 5' nested primer and the same primer as for cDNA synthesis or a nested transcript-specific primer; 5'-CCGTGAGATCACTGCACAGAATCA-3' for TS2 and 5'-AGCAGACAGAGGCACGATTCAGAA-3' for TS3 in a volume of 30 ml with 0.6 ml of the first PCR products, 0.2 mM of each primer, 200 mM of each dNTP, 2 mM $MgCl_2$, 0.2 ml of AmpliTaq Gold DNA polymerase, and 1× PCR buffer supplied with the polymerase. PCR conditions were as follows: an initial denaturation of 95°C for 10 min; 25 cycles of 94°C for 30 s, 65°C for 30 s, and 72°C for 1 min; and a final extension of 72°C for 10 min. The PCR products were separated by an electrophoresis on a 1.5% agarose gel, cut off from the gel, purified with the Geneclean kit, and sequenced by using the transcript-specific primers.

Supplemental References

1. Gregory, R. I., Khosla, S. & Feil, R. (2001) *Methods Mol. Biol.* 181, 269–284.

2. Duggleby, R. G., Kinns, H. & Rood, J. I. (1981) *Anal. Biochem.* 110, 49–55.

3. Gregory, R. I. & Feil, R. (1999) *Nucleic Acids Res.* 27, e32.

4. Uejima, H., Lee, M. P., Cui, H. & Feinberg, A. P. (2000) *Nat. Genet.* 25, 375–376.

CHAPITRE III

RNAi-Mediated Allelic trans-Interaction at the Imprinted RTL1/PEG11 Locus

Davis E, Caiment F, Tordoir X, Cavaille J, Ferguson-Smith A, Cockett N, Georges M, Charlier C. 2005. Current Biology 15(8): 743-749.

RNAi-Mediated Allelic *trans*-Interaction at the Imprinted *RTL1/PEG11* Locus

Erica Davis,[1] Florian Caiment,[1] Xavier Tordoir,[1] Jérôme Cavaillé,[2] Anne Ferguson-Smith,[3] Noelle Cockett,[4] Michel Georges,[1,*] and Carole Charlier[1]

[1] Department of Genetics, Faculty of Veterinary Medicine, University of Liège, 20 Boulevard de Colonster 4000 Liège , Belgium

[2] Laboratoire de Biologie, Moléculaire Eucaryote, Œntre National de la Recherche Scientifique (Unité Mixte de Recherche 5099), IFR-109, Université Paul Sabatier , 31062 Toulouse , France

[3] Department of Anatomy , University of Cambridge , Cambridge CB2 3DY, United Kingdom

[4] Animal, Dairy, and Veterinary Sciences , College of Agriculture , Utah State University , Logan, Utah 84322-4700

SUMMARY

The *DLK1-GTL2* imprinted domain, encompassing the callipyge (*CLPG*) locus in sheep, has recently been shown to harbour a large number of maternally expressed miRNA genes [1-2]. Two of these (miR-127 and miR-136) are processed from a transcript (*anti-PEG11*) that is antisense to *RTL1*/*PEG11*, a paternally expressed intronless gene with homology to the gag and pol polyproteins of Sushi-like retroelements [3]. We herein demonstrate that several additional miRNAs are processed from *anti-PEG11* and that these regulate *RTL1*/*PEG11* in trans by guiding RISC-mediated cleavage of its mRNA. This is the first demonstration of miRNA-mediated RNAi involving imprinted genes in mammals.

RESULTS AND DISCUSSION

In silico prediction of novel *(anti)PEG11*-hosted miRNA genes

We first aligned the human, mouse, rat, sheep and dog *RTL1*/*PEG11* gene sequences using ClustalW. To improve the quality of the alignment, we performed a protein sequence alignment which was then back-translated into the respective nucleotide sequences. A cluster of tandem repeats (referred to as TRA, TRB and TRC) at the 5' end of *RTL1*/*PEG11* was aligned manually (see Figures S1 and S2).

We then generated a multispecies similarity profile by sliding an 80-bp window across the multiple alignment and by computing an average pair-wise identity score for each window. This allowed us to identify four hyper-conserved regions of respectively 91 (region I), 124 (region II), 114 (region III), and 92 (region IV) base pairs within which all 80-bp windows exhibited average pair-wise similarities ≥ 98 % (Figure 1). All four regions were shown to exhibit highly significant deficits in synonymous substitutions (See Supplemental Experimental Procedures) indicating that the corresponding gene segments undergo evolutionary constraints other than the protein sequence.

Hyper-conserved regions III and IV coincide respectively with miR-127 and miR-136, which are processed from the *anti-PEG11* strand. To test whether hyper-conserved regions IIV might encode additional miRNA genes, we analyzed the corresponding sequences using

RNAfold [4]. As expected conserved hairpin loops were detected on the *anti-PEG11* strand for regions III and IV, corresponding to miR-127 and miR-136. In addition, RNAfold also predicted conserved hairpin loops on the *anti-PEG11* strand for regions I and II, and on the *RTL1/PEG11* strand for regions III and IV (but not I and II) suggesting that both *anti-PEG11*

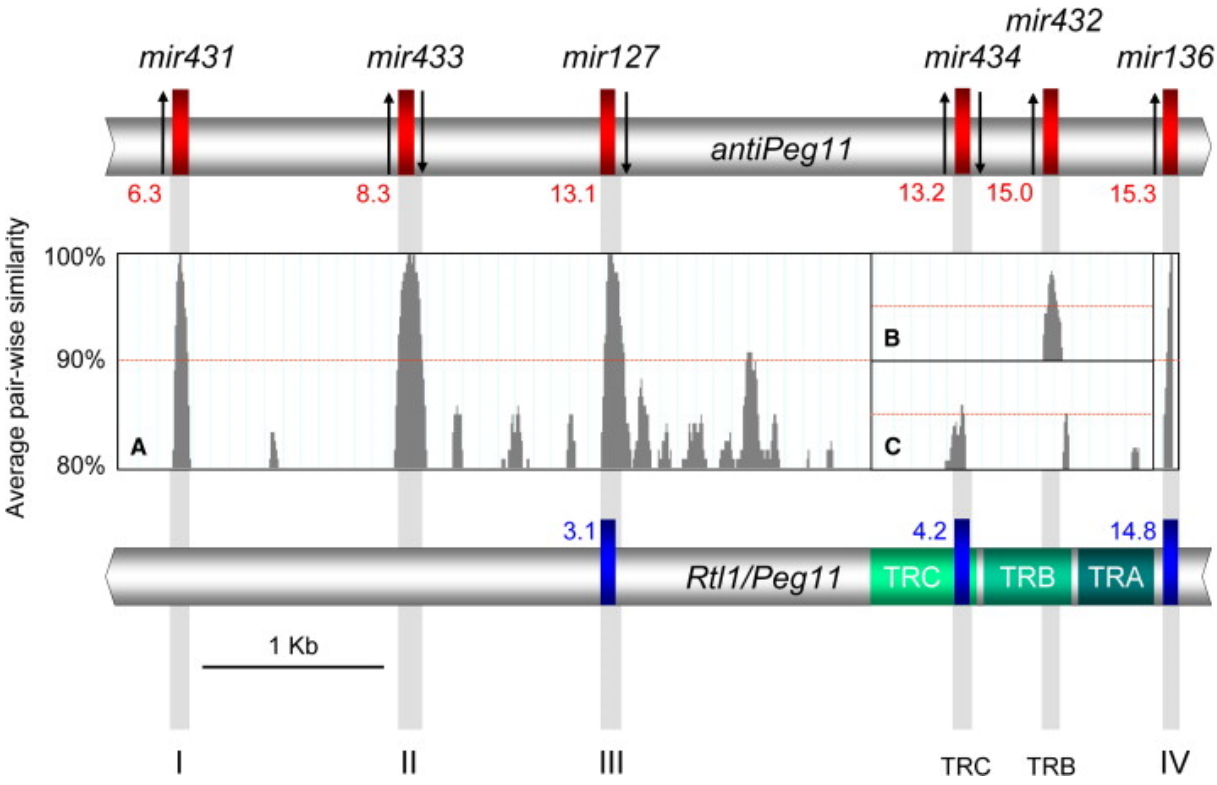


Figure 1: Bioinformatic prediction of *(anti)Peg11*-hosted miRNA genes.

Schematic representation of the *(anti)PEG11* locus showing (i) multispecies similarity profiles (**A**: human, mouse, rat, sheep, dog; **B**: human, sheep, dog; **C**: mouse, rat) identifying four regions of extreme conservation (I, II, III, IV), and two regions of high conservation (TRB, TRC), harbouring evolutionary conserved hairpin structures; (ii) Pre-miRNAs predicted by MirScan (+ scores) in the conserved regions, highlighted in red when experimentally confirmed, in blue when not; (iii) arrows indicating from which arm the mature miRNAs are processed, and (iv) a cluster of tandem repeats (TRA, TRB and TRC) highlighted as green boxes at the 5' end of the *RTL1/PEG11* open reading frame (See Figure S2).

and RTL1/PEG11 might encode additional miRNA genes (Figure S3). The MiRscan scores [5] obtained with the human and mouse sequences for the corresponding RNA stem-loops are shown in Figure 1.

By screening sequence databases we identified a human miRNA (AY785934; hereafter referred to as miR-432) mapping to the TRB tandem repetitions (*anti-PEG11* strand). Its position coincides with a 90-bp window exhibiting an average pair-wise similarity of 95% when comparing human, sheep and dog. Analyzing the corresponding sequences

using RNAfold identified a strand-specific hairpin structure conserved in all three species (Figure S3), and yielded a MiRscan score (human-ovine) of 15.0 (Figure 1). We found no evidence for a stable hairpin loop in the corresponding rodent windows, which only exhibit an average pair-wise similarity of 62 % with their human, ovine and canine orthologues. However, when analyzing the rodent TRC repeats (which are absent in human, sheep and dog), we identified conserved hairpins in both the *RTL1/PEG11* and *anti-PEG11* strands in a 94-bp window with 90% similarity between mouse and rat and yielding MiRscan scores (mouse - rat) of respectively 4.2 and 13.2 (Figure 1).

Hence, bioinformatic analysis predicts four *anti-PEG11* miRNA precursors shared by all analyzed mammals in regions I to IV, one rodent-specific *anti-PEG11* miRNA precursor in region TRC, one *anti-PEG11* miRNA precursor in region TRB shared by non-rodent mammals, two *RTL1/PEG11* miRNA precursors shared by all analyzed mammals in regions III and IV, and one rodent-specific *RTL1/PEG11* miRNA precursor in region TRC.

Expression analysis confirms the *antiPeg11*-, but not *Rtl1/Peg11*-hosted miRNAs

To verify which of these putative, *in silico* predicted, pre-miRNAs are processed into mature miRNAs, we performed primer extensions using total RNA isolated from a range of murine tissues. Four to eight probes were initially tested for each of the eight pre-miRNAs predicted in rodents (thus excluding the TRB pre-miRNA), targeting both stem-loop arms and offset by one to six base pairs within a given arm (Table S1). As expected we detected extension products for the 3' arm of miR-127 and for the 5' arm of miR-136. In addition, we detected extension products corresponding to the 3' arm of the *anti-PEG11*/region I stem-loop (hereafter referred to as miR-431), for the 5' and 3' arms of the *anti-PEG11*/region II stem-loop (hereafter referred to as miR-433-5p and mirR-433-3p), as well as for the 5' and 3' arms of the *anti-PEG11*/TRC repeat stem-loop (hereafter referred to as mir-434-5p and mir-434-3p). The specificity of the extension products was supported by the fact that each mature miRNA was detected by at least two probes yielding products of sizes compatible with the offset of the corresponding primers (Table S1 and Figure S4). We were not able to detect extension products for either of the pre-miRNAs predicted in the sense *RTL1/PEG11* transcript, indicating that miRNAs are exclusively processed from *anti-PEG11*.

All detected miRNAs shared an expression profile similar to that of *RTL1/PEG11*, i.e. preferentially expressed in embryo, placenta, brain and skeletal muscle (Table S2) (Figure 2A). miR-127 and miR-136 were previously shown to be imprinted and preferentially expressed from the maternal allele as they were detected in tissue samples from mUPD12 but not pUPD12 mice [1]. The same was demonstrated in this work for miR-431 (data not shown).

Anti-PEG11-hosted miRNAs guide RISC mediated cleavage of *RTL1/PEG11 in vivo*

We then tested whether *RTL1/PEG11* is indeed a target for the detected *anti-PEG11*-hosted miRNAs *in vivo*. To that end we used mouse placental total RNA to perform RNA ligase mediated (RLM) 5' RACE experiments [6] aimed at isolating the predicted RISC (RNAinduced silencing complex [7])-mediated *RTL1/PEG11* mRNA cleavage products.

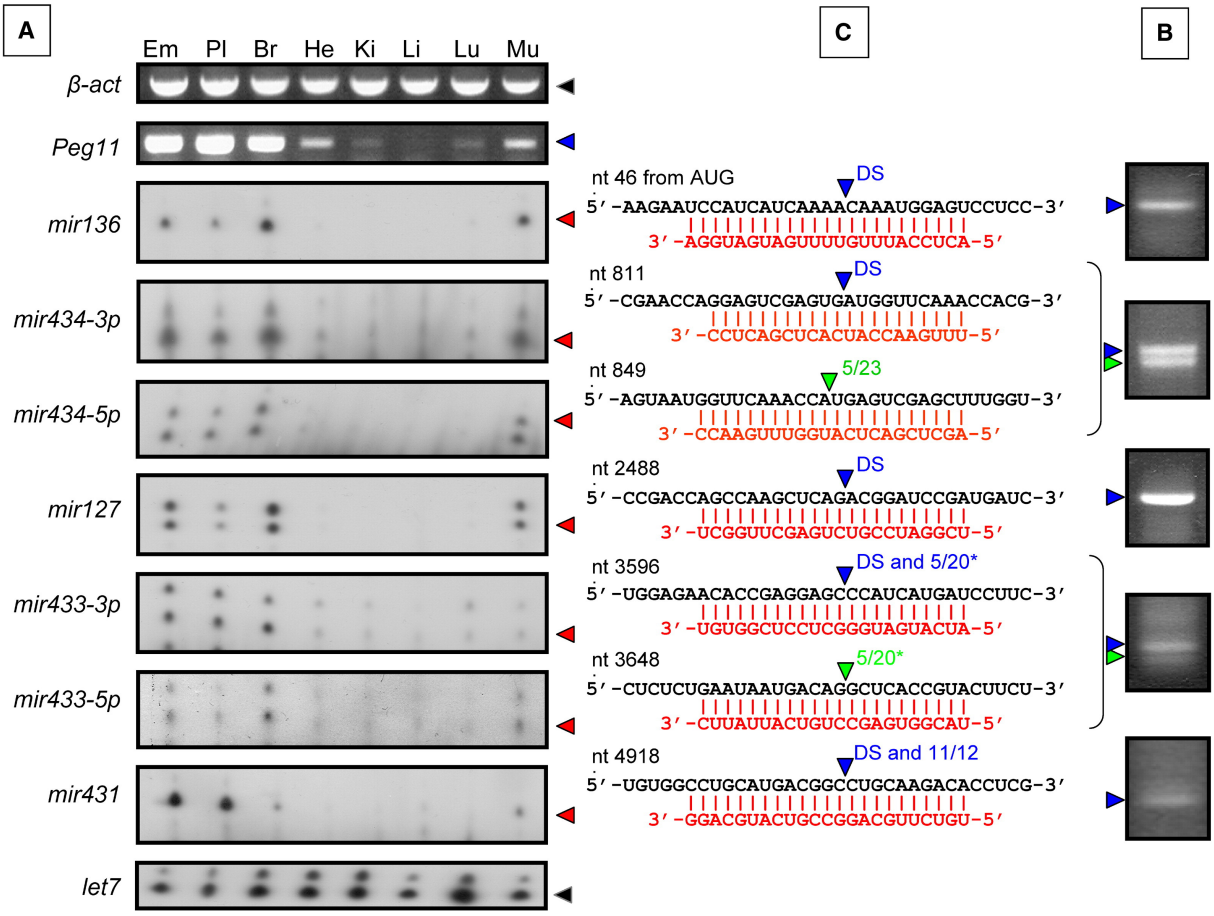


Figure 2: Detection of mature *anti-PEG11* miRNAs and their corresponding *RTL1/PEG11* cleavage products.
A. Pre- (embryo (Em), placenta (Pl)) and post-natal (brain (Br), heart (He), kidney (Ki), liver (Li), lung (Lu), skeletal muscle (Mu)) expression profile for *RTL1/PEG11* and *β-actin* (RT-PCR), as well as let-7 and the seven *anti-PEG11* miRNAs (primer extension).**B**. *RTL1/PEG11* RLM 5'RACE products identifying cleavage products directed by all seven *anti-PEG11* miRNAs. **C**. Sequences of the seven *anti-PEG11* miRNAs (red) hybridized to their respective *RTL1/PEG11* targets (black); the arrows correspond to the cleavage sites identified by RLM 5'RACE either by direct sequencing of the PCR products (DS) or by sequencing individual cloned products (numbers indicate the fraction of clones that identify the predicted cleavage site; both *miR-433* PCR products were cloned jointly (*)).

One round of RT-PCR was sufficient to obtain clean RLM 5'RACE products corresponding to regions I, III, TRC (two bands) and IV. The products of regions I, III, IV and the largest band of region TRC were directly sequenced and shown to correspond to *RTL1/PEG11* mRNA cleavage products ending as expected at the nucleotide complementary to the 10th position of the respective mature miRNAs (miR-431, miR-127, miR-136 and mir-434-3p).

The smaller TRC product was cloned and the insert of 23 random clones sequenced. Five of these were shown to correspond to the *RTL1/PEG11* cleavage product expected for mir-434-5p, ending in this case at the nucleotide complementary to the 11th position of this miRNA. The remaining clones correspond presumably to random *RTL1/PEG11* degradation products. A second round of nested PCR was necessary to obtain two distinct bands corresponding to region II. These were cloned together and the sequence of 20 randomly picked clones was determined. Five of these corresponded to the expected miR-433-5p-guided *RTL1/PEG11* cleavage products, five others to the expected miR-433-3p-guided cleavage products. All were ending at the nucleotide complementary to the 10th position of the respective mature miRNAs. The remaining ten clones corresponded to distinct, presumably random *RTL1/PEG11* degradation products. (Figures 2B and C).

These results unambiguously demonstrate the *in vivo trans*-inhibition of paternally expressed *RTL1/PEG11* by miRNAs processed from the maternally expressed *anti-PEG11* precursor. They also demonstrate that mir-434 and miR-433 are unusual in that both stem-loop arms of the corresponding pre-miRNAs generate a mature miRNA incorporated in RISC. The *in vivo* demonstration of *anti-PEG11*-miRNA mediated degradation of *RTL1/PEG11* satisfactorily accounts for the previously reported observation of a four- rather than two-fold increase in *RTL1/PEG11* mRNA levels in mice inheriting a deletion of the *DLK1-GTL2* imprinting

control element (Δ-IG-DMR) on their maternal chromosome. Such mice inherit two chromosomes with paternal epigenotype, and are thus expected to show a two-fold increase in the expression levels of the paternally expressed protein encoding genes including *RTL1/PEG11*. Our results indicate that the observed four-fold increase is indeed due to the absence of the maternally expressed non coding RNA genes, including the *anti-PEG11*- miRNAs, and thus lack of RNAi-mediated *RTL1/PEG11* degradation - as initially surmised [8].

Identification of Drosha cleavage products of *anti-PEG11*-hosted pri-miRNAs

We then applied the same RLM 5' RACE technique, using murine placental RNA, to the *anti-PEG11* strand in order to detect putative cleavage products mediated by mature miRNAs processed from the *RTL1/PEG11* strand that might not have been detected by primer extension.

Primers were initially selected to explore hyper-conserved regions III and IV (Table S3) Strong RACE products were obtained in one round of RT-PCR and directly sequenced. They where shown to correspond to cleavage sites that mapped at the base of the miR-127 and miR-136 *anti-PEG11* stem-loops (3' arm) at positions agreeing perfectly with those expected from the action of DROSHA, the nuclear RNase III endonuclease that catalyzes pri-miRNA processing [7] (Figure 3). We thus designed primers allowing us to scan the entire anti-Peg11 strand by RLM 5' RACE (Table S3). Using the same procedure, we readily detected DROSHA cleavage products corresponding to the 3' extremity of the mir-434, miR-433 and miR-431 stem-loops. We also detected a cleavage product ending at the 5' extremity of the miR-136 stem-loop (thus not cleaved at the 3' extremity) and corresponding supposedly to an abundant processing intermediate. There was no evidence for DROSHA processing of any other, as of yet undetermined *anti-PEG11* miRNA precursor, even in the TRB region (Figure 3).

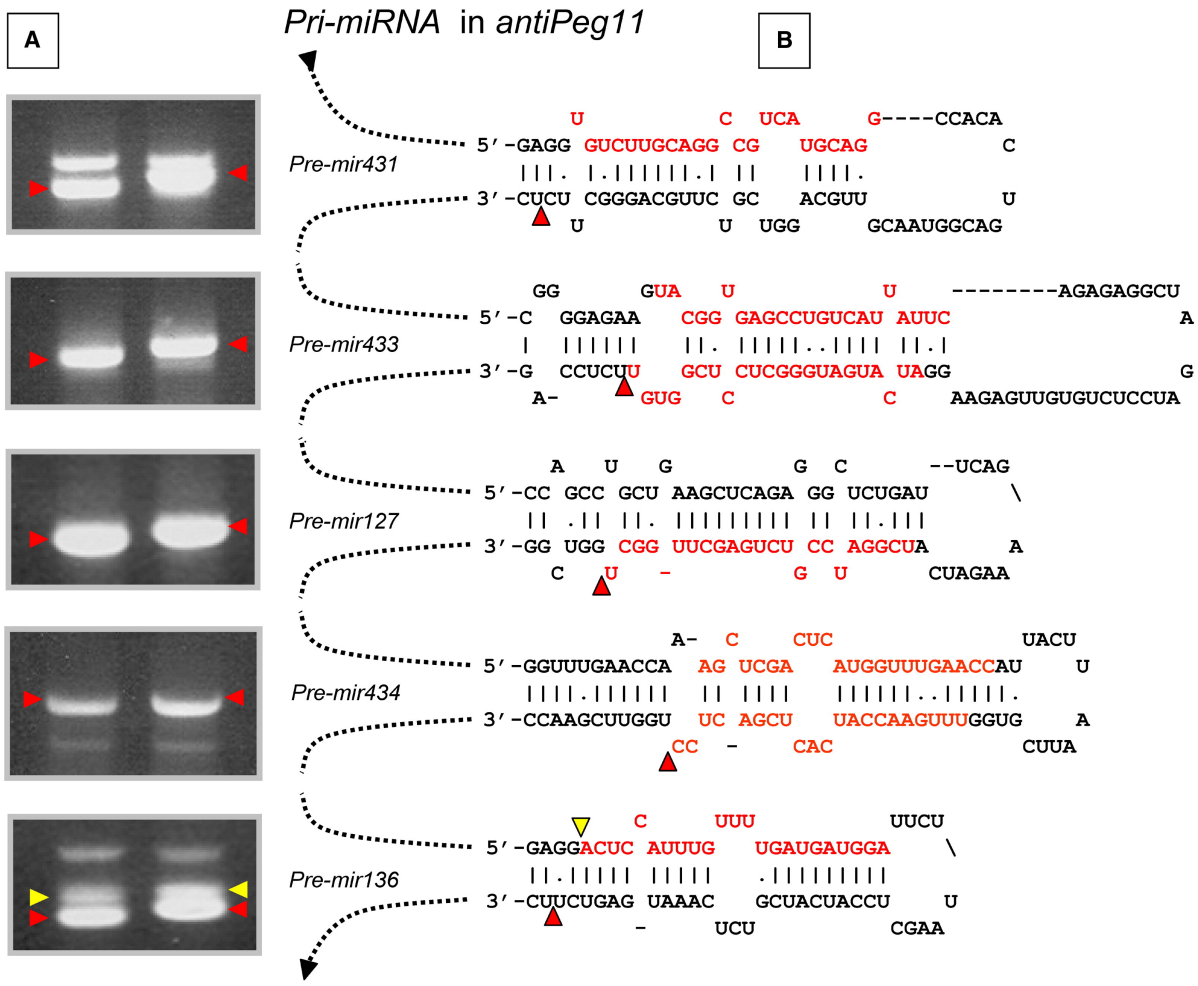


Figure 3: RLM 5' RACE experiments targeting *anti-PEG11* identify putative DROSHA-catalyzed cleavage products. A. Agarose gel electrophoresis of the RLM 5'RACE products targeting the four hyperconserved regions as well as the TRC region on the *anti-PEG11* strand. The two lanes correspond to RACE products obtained with two distinct primer sets per region (see Table S3). **B**. Schematic representation of the five *anti-PEG11* pre-miRNA stem-loops highlighting the experimentally identified DROSHA cleavage sites (arrows) and the mature miRNAs (in red).

These results thus provide additional independent confirmation of the genuine nature of the five *anti-PEG11*-hosted murine miRNA genes, yet the absence of functional *RTL1/PEG11*-hosted miRNA genes.

RNAi-mediated allelic *trans*-interaction and the conflict hypothesis of parental Imprinting

The fact that *RTL1/PEG11* is strongly expressed in the placenta, as well as the most striking symptom associated with *RTL1/PEG11* overexpression (placentomegaly - [9]), sug-

gest that *RTL1/PEG11* promotes placental supply of maternal nutrients akin to other paternally expressed imprinted genes, including *Igf2* [10]. It is interesting in this regard that we were not able to identify the orthologue of *RTL1/PEG11* in 5.7 genome equivalents of Fugu genome, 6.6 genome equivalents of chicken genome, and 7.2 genome equivalents of opossum genome, suggesting that *RTL1/PEG11* is eutherian-specific.

The identification of maternally expressed *trans*-inhibitors of *RTL1/PEG11* reveals a striking resemblance with the *trans*-inhibition of *Igf2* by the maternally expressed *Igf2r* (following binding of *Igf2* to *Igf2r* at the cell surface, the ligand-receptor complex is internalized and targeted to lysosomes in which *Igf2* is degraded [10]). It suggests that the same evolutionary forces at the heart of the parental conflict theory [11] have recruited RNAi in regulating foetal growth by selecting for mutations that create strand-specific pre-miRNAs in *anti-Peg11* while leaving *RTL1/PEG11* unaltered. The occurrence of multiple *anti-PEG11* miRNAs suggests that they are individually incapable of fully counteracting *RTL1/PEG11*.

Preliminary evidence for an unusual similarity amongst the miRNA precursors suggests that they might be paraloguous and provides a glimpse in how this multilayered blocking system might have evolved (data not shown).

RNAi-mediated allelic *trans*-interaction and polar overdominance

Our findings are particularly intriguing in light of the unusual inheritance pattern of the callipyge phenotype ("polar overdominance") in which only heterozygous individuals inheriting the *CLPG* mutation from their father express the muscular hypertrophy [12]. We have previously hypothesized that polar overdominance might involve the *trans* inhibition of paternally expressed protein encoding genes by maternally expressed non coding RNA genes, possibly miRNAs uncovered within the *DLK1-GTL2* domain [13-14].

To test whether the predicted *trans* interaction occurs between *PEG11* and the *anti-PEG11*-hosted miRNA genes in skeletal muscle of sheep, we first examined the expression of the *anti-PEG11* miRNAs in this tissue by means of primer extension. miR-127, miR-136, miR-431 and a miRNA corresponding to the 3'arm of the *anti-PEG11*/TRB repeat stem loop

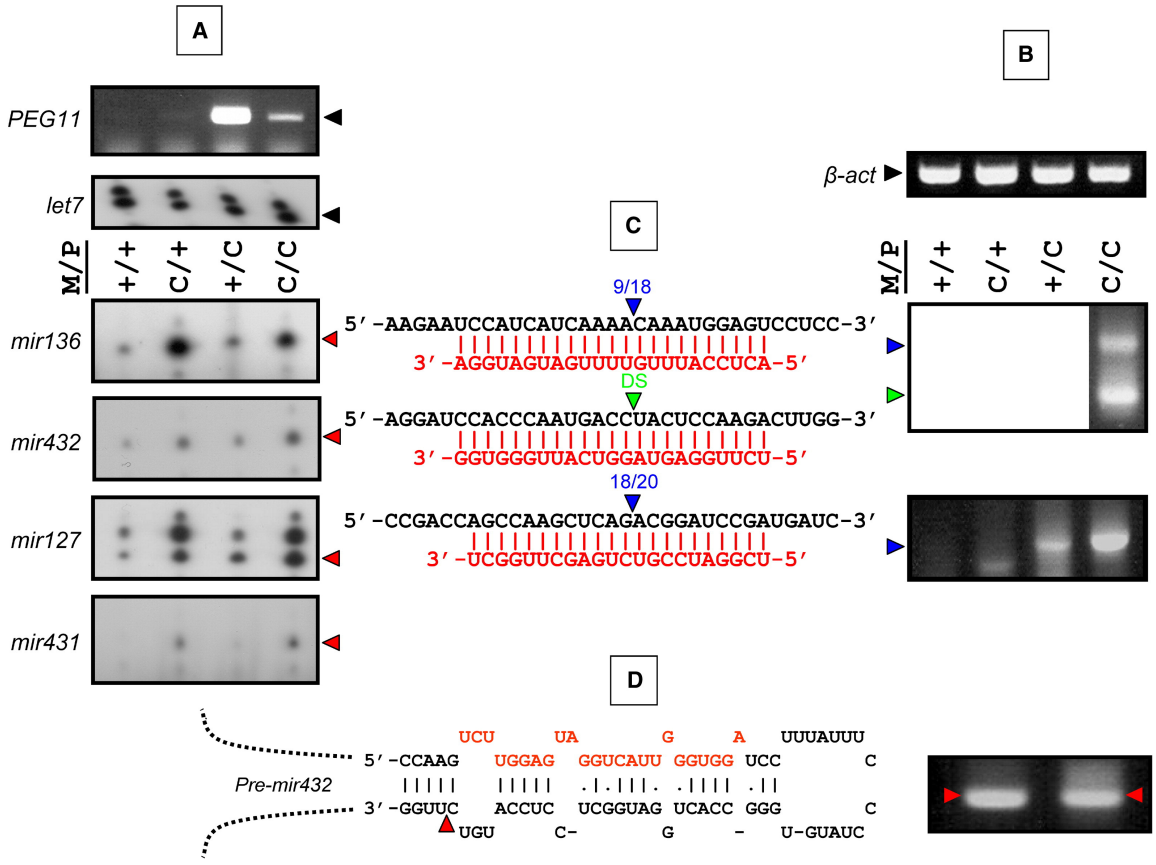


Figure 4: *anti-PEG11*-miRNA and *PEG11* expression, RISC-mediated *PEG11* cleavage and Drosha catalyzed pri-*miR-432* processing in ovine skeletal muscle.
A. Detection, in ovine *longissimus dorsi*, of mature let-7, miR-136, miR-432, miR-127, miR-431 (primer extension), and *PEG11* (RT-PCR; Table S2), showing the *cis*-effect of the *CLPG* mutation on expression levels. **B**. RLM 5' RACE products targeting intact *β-actin* as well as miR-136, miR-432 and miR-127-guided cleavage products in ovine skeletal muscle. **C**. Sequences of *anti-PEG11* miRNAs (red) hybridized to their *PEG11* target (black); the arrows correspond to the cleavage sites identified by RLM 5'RACE either by direct sequencing of the PCR product (DS) or by sequencing individual cloned products (numbers indicate the fraction of clones that identify the predicted cleavage site). **D**. Agarose gel electrophoresis of the RLM 5'RACE products targeting the TRB region on the *anti-PEG11* strand. The two lanes correspond to RACE products obtained with two distinct primer sets (see Table S3). Schematic representation of the pre-*miR-432* stem-loop highlighting the experimentally identified DROSHA cleavage site (arrow) and the mature miR-432 (in red).

(hereafter referred to as *miR-432*) were detected in *longissimus dorsi* of 8-week old sheep representing the four possible genotypes at the *CLPG* locus ($+^{Mat}/+^{Pat}$, $CLPG^{Mat}/+^{Pat}$, $+^{Mat}/CLPG^{Pat}$, $CLPG^{Mat}/CLPG^{Pat}$). The genuine nature of the newly identified *miR-432* miRNA was confirmed by the use of offset primers (Figure S4) and by the identification of the corresponding 3' DROSHA cleavage product (Figure 4D). The expression levels of all these miRNAs were affected by the *CLPG* mutation as expected, being most abundant in $CLPG^{Mat}/+^{Pat}$ and $CLPG^{Mat}/CLPG^{Pat}$ individuals sharing the *CLPG* mutation on their maternal chromosome (Figure 4A).

We then performed RLM 5' RACE experiments to detect putative *PEG11* cleavage products. We first used RNA from *CLPG^{Mat}/CLPG^{Pat}* animals. Indeed, as a result of the *cis* effect of the *CLPG* mutation, these animals are overexpressing both *PEG11* and the *anti-PEG11*-hosted miRNAs and therefore cleavage products are predicted to be most abundant in *CLPG^{Mat}/CLPG^{Pat}* individuals. Using a primer annealing downstream of the *miR-127* complement, we obtained a band after two rounds of nested PCR. It was cloned and the sequence of 18 out of 20 clones shown to correspond to the expected *miR-127* mediated *PEG11* cleavage product. Using a primer annealing downstream of the *miR-432* complement, we obtained two RACE products after two rounds of nested PCR. The smallest one was directly sequenced and shown to correspond to the expected *miR-432* mediated *PEG11* cleavage product. The larger one was cloned and the sequence of nine out of 18 clones shown to correspond to the expected *miR-136* mediated *PEG11* cleavage product (Figures 4B and 4C). We then performed the same RLM 5'RACE experiments using RNA extracted from skeletal muscle of eight week old *+^{Mat}/CLPG^{Pat}*, *CLPG^{Mat}/+^{Pat}* and *+^{Mat}/+^{Pat}* animals. A weak *miR-127* cleavage product was obtained in *+^{Mat}/CLPG^{Pat}* animals, but not in the two other genotypes (Figure 4B). Neither of the *miR-432* nor *miR-136* cleavage products were detected in any of these animals (data not shown). The higher abundance of the RACE products in *CLPG^{Mat}/CLPG^{Pat}* animals when compared to the three other genotypes is thus in agreement with the known *cis* effect of the *CLPG* mutation. The miRNA mediated degradation of *PEG11* transcripts in *CLPG^{Mat}/CLPG^{Pat}* animals also satisfactorily explains why the *PEG11* RNA levels were systematically found to be lower in these animals when compared to *+^{Mat}/CLPG^{Pat}* individuals (Figure 4A and [15]).

We are presently examining whether ectopic expression of *PEG11* might contribute to the induction of the callipyge phenotype and whether miRNA-mediated repression is involved in the previously reported *trans*-inhibition of *DLK1* whose ectopic expression was shown to cause a callipyge-like phenotype in transgenic mice [16].

ACKNOWLEDGMENTS

This project was supported by grants from (i) the FRFC (n° 2.4525.96), (ii) Crédit aux Chercheurs (n° 1.5.134.00) from the FNRS, (iii) Crédit à la Recherche from the ULg, (iv) the SSTC (n° 0135), (v) the European Union (Callimir), (vi) the Utah Center of Excellence Program, (vii) the USDA/NRICGP Grants #94-04358, #96-35205 and #98-03455), (viii) the Utah Agricultural Experiment Station, USU, and (ix) the ACI (Biologie cellulaire, moléculaire et structurale). Erica Davis is a fellow of the Belgian American Educational Foundation. Carole Charlier is Chercheur Qualifié from the FNRS.

REFERENCES

1. Seitz, H., Youngson, N., Lin, S.P., Dalbert, S., Paulsen, M., Bachellerie, J.P., Ferguson-Smith, A.C., and Cavaillé, J. (2003). Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. Nat. Genet. *34*, 261-262.

2. Seitz, H., Royo, H., Bortolin, M.L., Lin, S.P., Ferguson-Smith, A.C., and Cavaille, J. (2004) A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. Genome Res.*14*, 1741-1748.

3. Youngson, N., Kocialkowski, S., Peel N., and Ferguson-Smith A.C. (2005) A small family of sushi-class retrotransposon-derived genes in mammals and their relation to genomic Imprinting. J. Mol. Evol., in press.

4. Hofacker, I.L., Fontana, W., Stadler, P.F., Bonhoeffer, S., Tacker, M., and Schuster, P. (1994) Fast Folding and Comparison of RNA Secondary Structures. Monatshefte f Chemie *125*, 167-188.

5. Lim, L.P., Lau, N.C., Weinstein, E.G., Abdelhakim, A., Yekta, S., Rhoades, M.W., Burge, C.B., and Bartel, D.P. (2003). The microRNAs of Caenorhabditis elegans. Genes Dev. *17*, 991-1008.

6. Yekta, S., Shih, I.H., and Bartel, D.P. (2004) MicroRNA-directed cleavage of HOXB8 mRNA. Science *304*, 594-596.

7. Bartel, D.P. (2004). MicroRNAs: genomics, biogenesis, mechanism, and function. Cell *116*, 281-297.

8. Lin, S.P., Youngson, N., Takada, S., Seitz, H., Reik, W., Paulsen, M., Cavaille, J., and Ferguson-Smith, A.C. (2003) Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome Nat. Genet. *35*, 97-102.

9. Georgiades, P., Watkins, M., Surani, M.A., and Ferguson-Smith, A.C. (2000) Parental origin-specific developmental defects in mice with uniparental disomy for chromosome 12. Development *127*, 4719-2478.

10. Reik, W., Constancia, M., Fowden, A., Anderson, N., Dean, W., Ferguson-Smith, A., Tycko, B., and Sibley, C. (2003) Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. J. Physiol. *547*, 35-44.

11. Wilkins, J.F., Haig, D. (2003) What good is genomic imprinting: the function of parent-specific gene expression. Nat. Rev. Genet. *4*, 1-10.

12. Cockett, N.E., Jackson, S.P., Shay, T.L., Farnir, F., Berghmans, S., Snowder, G.D., Nielsen, D.M., and Georges, M. (1996). Polar overdominance at the ovine callipyge locus. Science. 273, 236-238.

13. Georges, M., Charlier, C., and Cockett, N. (2003). The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. Trends Genet. *19*, 248-252.

14. Georges, M., Charlier, C., Smit, M., Davis, E., Shay, T., Tordoir, X., Takeda, H., Caiment, F., and Cockett, N. (2004). Toward molecular understanding of polar overdominance at the ovine callipyge locus. In Cold Spring

Harbor LXIX Symposium on Quantitative Biology, B. Stillman and D. Stewart, eds. (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory).

15. Charlier, C., Segers, K., Karim, L., Shay, T., Gyapay, G., Cockett, N., Georges, M. (2001). The callipyge (CLPG) mutation enhances the expression of the coregulated *DLK1*, *GTL2, PEG11* and *MEG8* genes in *cis* without affecting their imprinting status. Nat. Genet. *27*, 367-369.

16. Davis, E., Harken Jensen, C., Schroder, H.D., Farnir, F., Shay-Hadfield, T., Kliem, A., Cockett, N., Georges, M., Charlier, C. (2004). Ectopic expression of DLK1 protein in skeletal muscle of padumnal heterozygotes causes the callipyge phenotype. Curr. Biol. 14, 1858-1862.

SUPPLEMENTAL DATA

EXPERIMENTAL PROCEDURES

Identification of R*TL1/PEG11* segments exhibiting significant deficits in synonymous substitution rate.

To identify *RTL1/PEG11* segments exhibiting a lower than expected rate of synonymous substitutions, we first aligned the five (human, mouse, rat, sheep and dog) RTL1/PEG11 protein sequences using ClustalW. For each position in the protein alignment we then generated the score *s = round 2*(E-O-1)* that compares the observed number of codons across the five species, *O*, with the expected number, *E*, given the five amino-acids and assuming that synonymous codons are sampled according to their usage frequency in mammals. Aminoacid positions with a gap in one or more species were given a score *s = -10*. Adjacent positions with -10 score were collated into one. This yielded 10 distinct scores, s_i, with corresponding frequencies p_i. These parameters were then used to compute the likelihood of an aggregate (additive) score for a given segment of size *n*, under the hypothesis of random succession of residues with score s_i and frequency of occurrence p_i following Karlin and Altschul [S1]. This identified four highly significant *RTL1/PEG11* segments at positions (starting from ATG in the multispecies alignment) 10 to 97 (p = 7 x 10-5; *miR-136*: 16-77), 3061 to 3223 (p = 3 x 10-9; *miR-127*: 3140-3209), 4210 to 4336 (p = 1 x 10-3; *miR-433*: 4249- 4339) and 5524-5602 (p = 6 x 10-3; *miR-431*: 5534-5594). By comparison, the fifth most significant segment has a score which is expected to occur in a sequence of the length of *RTL1/PEG11* 2.8 times by chance alone.

Primer extension assay to detect mature miRNAs:

Total RNA was extracted from day-17 p.c. (embryo and placenta) and 6 weeks post-natal (brain, heart, kidney, liver, lung, and skeletal muscle) mouse tissues and from 8 weeks postnatal sheep *longissimus dorsi* employing Trizol® reagent (Invitrogen) according to manufacturer's instructions. Each oligoprobe (Table S1) was labelled by a kination reaction. Briefly, 350ng of primer was labelled with 25 µCi of γ^{32}-ATP by 10U of T4 Polynucleotide

Kinase (New England Biolabs) using the buffer provided by the manufacturer in a 15µl reaction placed at 37°C for 1 hour. The labelled oligoprobe was gel purified in a 15% acrylamide/7M urea gel, eluted over night at 42°C in a final volume of 300µl of SDS 0.1 %, 0.5mM EDTA and 0.5 M ammonium acetate. The eluted probe was ethanol-precipitated after a phenol-chloroform extraction and resuspended in DEPC treated water to reach an activity of 10,000cpm/µl. Ten µg of total RNA (1µg/µl) and 3µl of probe (30,000cpm) were mixed together, heated at 75°C for 5 min, and then an annealing temperature of 42°C was reached by gradual step down (5°C/5min). Reverse transcription was carried out with 5U of AMV reverse transcriptase (Promega) at 42°C for 2 hours in a 20µl final volume of AMV buffer 1x and 1µM dNTP. The resulting cDNA products were separated in a 15% acrylamide / 7 M urea gel. The gel was dried and exposed at -80°C on an autoradiography (Hyperfilm, Amersham) for one night to one week.

RT-PCR analysis of *PEG11/RTL1*

Mouse and sheep total RNA (see above) was DNase treated with TURBO DNA-free™ (Ambion) to remove DNA contamination. The EndoFree RT™ primer specific reverse transcription kit (Ambion) was used to generate cDNA from 3 µg total RNA primed with a *PEG11/RTL1* strand specific primer and a *β-actin* specific primer (Table S2). *PEG11/RTL1* transcript levels were then assessed by cDNA amplification and *β-actin* specific primers were used to confirm initial total RNA integrity and demonstrate equal amounts of RNA across samples. All PCR reactions were conducted with Hot Star Taq polymerase (Qiagen), 10% DMS0 and cycling conditions of 94°C for 12 min; 30 cycles of 92°C for 60 sec, 60°C (*Peg11/Rtl1*) or 55°C (*β-actin*) for 90 s, 72°C for 90 s; and 72°C for 7 min.

RNA Ligase Mediated (RLM) 5' RACE experiments

Total RNA was extracted from day-17 p.c. mouse placenta and from 8 weeks post-natal sheep *longissimus dorsi* . Modified RLM 5' RACE was carried out with the GeneRacer™ Kit (Invitrogen). Briefly, 2.5µg total RNA was ligated to a 44 base synthetic RNA oligonucleotide and cDNA synthesis was primed with an (anti)*PEG11* strand specific primer (500 nM final concentration) and a *β-actin* specific primer (10 nM final concentration) listed

in Table S3 and was executed with SuperScript™ III reverse transcriptase. Next, cDNA was amplified with a non-specific forward primer (GeneRacer™ 5' Primer) and a gene-specific reverse primer (Table S3) using Hot Star Taq polymerase (Qiagen) with a touchdown PCR approach. First round PCR cycling conditions were 94°C for 12 min; 5 cycles of 92°C for 60 sec, 70°C for 90 s, 72°C for 90 s; 10 cycles of 92°C for 60 sec, 69°C (-1°C/cycle) for 90 s, 72°C for 90 s; 20 cycles of 92°C for 60 sec, 60°C for 90 s, 72°C for 90 s; and 72°C for 7 min. When necessary (mouse mir-433, and sheep miR-127, miR-432 and miR-136), PCR products were further amplified by nested PCR utilizing a second non-specific forward primer (GeneRacer™ 5' Nested Primer) and a gene-specific reverse primer (Table S3) using Hot Star Taq polymerase (Qiagen) and 10% DMSO. Nested PCR cycling conditions were 94°C for 12 min; 35 cycles of 92°C for 60 sec, 55°C for 90 s, 72°C for 90 s; and 72°C for 7 min. PCR products were either directly sequenced, or cloned (TOPO TA Cloning® Kit for Sequencing) and sequenced according to standard procedures.

SUPPLEMENTAL REFERENCES

S1. Karlin, S., and Altschul, S.F. (1990) Methods for assessing the statistical significance of molecular sequence features by using general scoring schemes. Proc. Natl. Acad. Sci. USA *87*, 2264-2268.

SUPPLEMENTAL FIGURES

Figure S1. Multiple Sequence Alignment of Peg11 ORF in Five Mammals via Clustal W

CLUSTAL W(1.81) multiple sequence alignment

This alignment is too big to be printed. You can access it freely online :

http://download.cell.com/current-biology/mmcs/journals/0960-9822/PIIS0960982205002332.mmc1.pdf

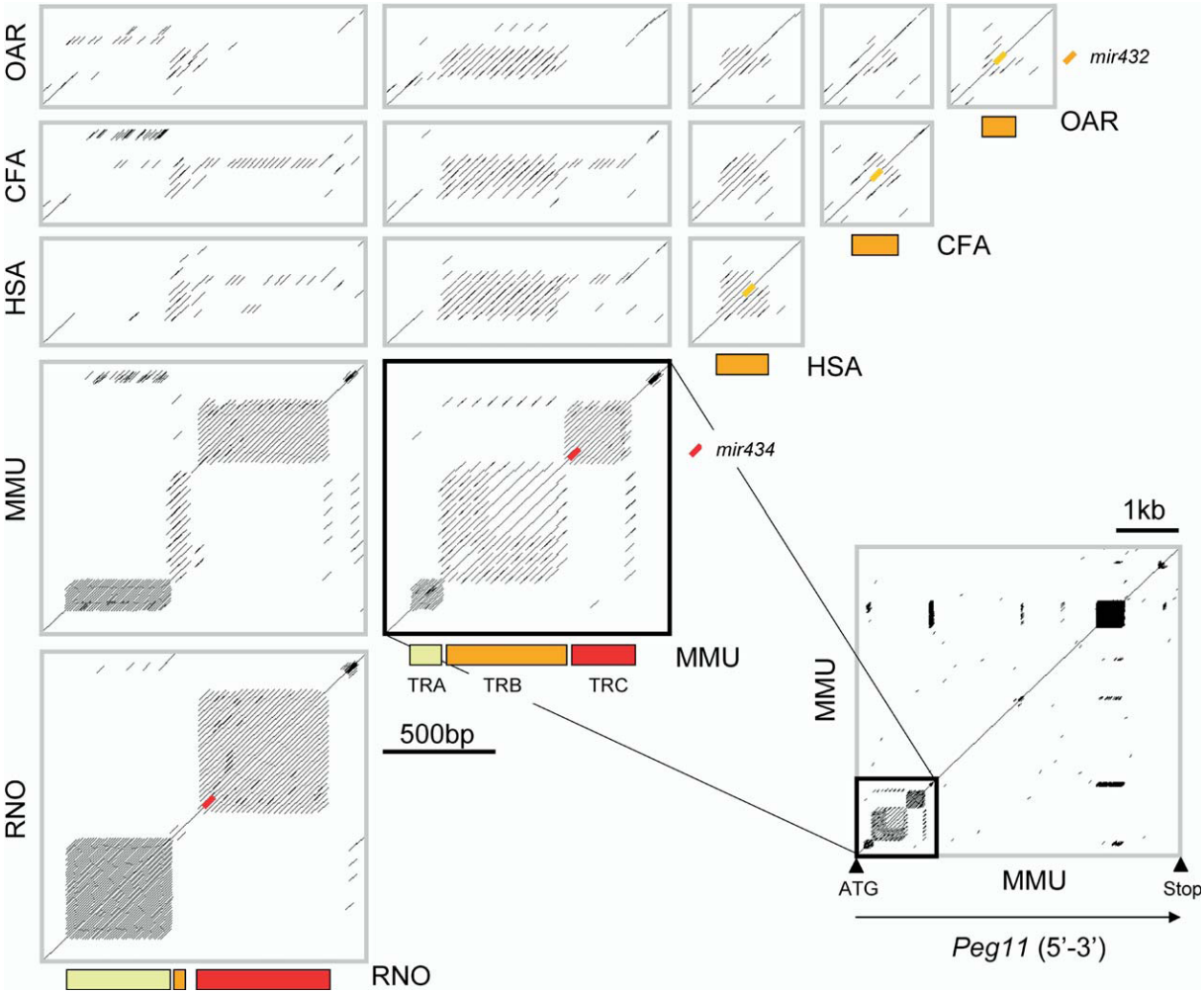


Figure S2. Intra- and interspecies dotplot analyses of the tandem repeat cluster (TR) present at the 5' end of *RTL1/PEG11*.

Dotplot analyses, using Dotmatcher (EMBOSS: http://ngfnblast.gbf.de/EMBOSS/emboss_group.html), were performed on the 5' sequence of *RTL1/PEG11*, beginning at the ATG start codon and ending after the TR, for five mammals: rat (RNO: 1450nt), mouse (MMU: 1270nt), human (HSA: 508nt), dog (CFA: 466nt) and sheep (OAR: 469nt). Intra- and inter-species graphical results, obtained with a window of 33nt and a threshold of 50, are presented. An inset, corresponding to a dotplot analysis of the complete murine gene, shows the relative position of the TR within the *RTL1/PEG11* ORF. Analysing the TR with Equicktandem (EMBOSS) showed that it comprises three adjacent units based on motifs of respectively 12 (TRA: yellow boxes), 33 (TRB: orange boxes) and 24 (TRC: red boxes) base pairs. TRA and TRC are only found in the mouse and rat, TRC encompasses the rodent-specific *miR-434* (red bar). TRB is found in all analyzed species and encompasses the non-rodent-specific *miR-432* (orange bar).

Figure S3. Stem-loops predicted by RNAfold (*anti-PEG11* and *RTL1/PEG11* strands) in the four hyperconserved regions I-IV, in TRC and in TRB.

For each region/strand combination we show the alignment of the nucleotide sequences, with asterisks corresponding to the residues conserved across the species and a line corresponding to the sequence of the mature miRNA. The structure of the stem-loops as determined by RNAfold is then shown in bracket notation, with associated folding free energy. Stem-loops which do not correspond to the most stable possible structure are marked as "suboptimal".

Available online :

http://download.cell.com/current-biology/mmcs/journals/0960-9822/PIIS0960982205002332.mmc1.pdf

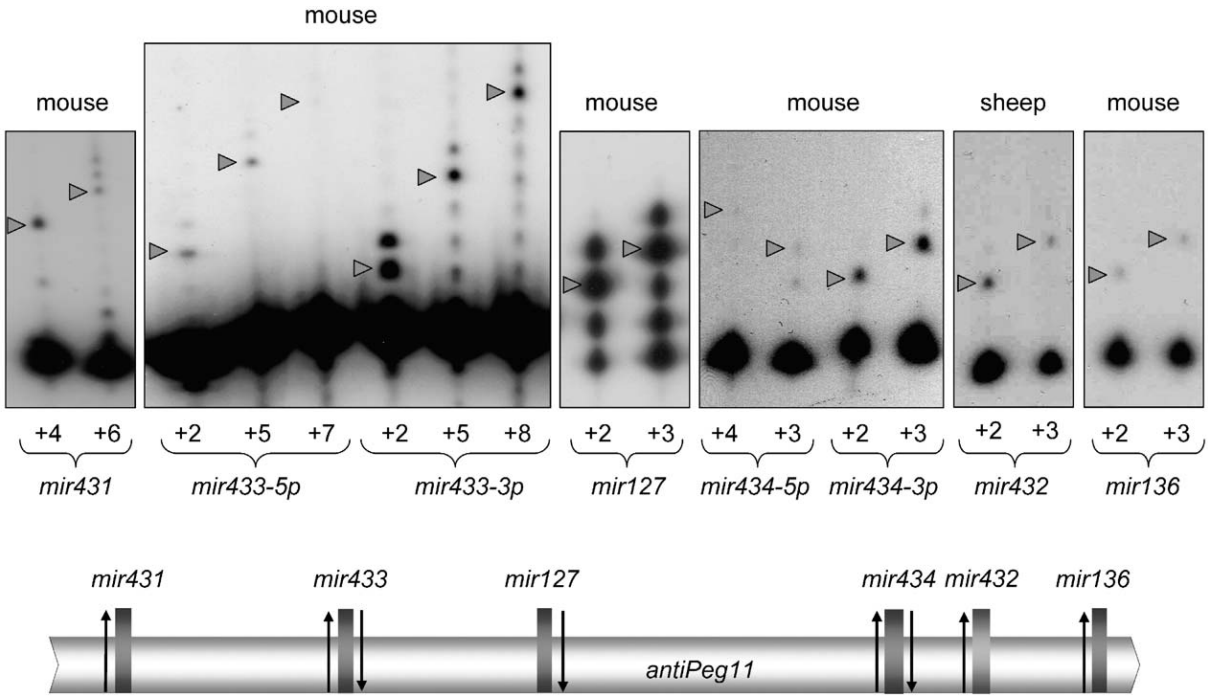


Figure S4. Primer extension assay to detect mature miRNAs: detailed results.
Results of primer extensions targeting seven murine and one ovine *anti-PEG11*-hosted mature miRNA. For each miRNA we show results obtained with two or three primers offset by one to six base pairs. The red arrows mark the position of the extension products in the gel. The number of extended bases are given for each lane with the name of the corresponding miRNA. The relative positions of the corresponding miRNA precursors are shown on the graph.

Table S1 : Oligonucleotide probes used for primer extension
Available online :
http://download.cell.com/current-biology/mmcs/journals/0960-9822/PIIS0960982205002332.mmc1.pdf

Table S2 : Primers for RT-PCR experiments
Available online :
http://download.cell.com/current-biology/mmcs/journals/0960-9822/PIIS0960982205002332.mmc1.pdf

Table S3 : Primers for RLM 5' RACE experiments
Available online :
http://download.cell.com/current-biology/mmcs/journals/0960-9822/PIIS0960982205002332.mmc1.pdf

CHAPITRE IV

Assessing the effect of the CLPG mutation on the microRNA catalogue of skeletal muscle using high throughput sequencing

Caiment F, Charlier C, Hadfield T, Cockett N, Georges M, Baurain D. 2010. Genome Research, in press.

Assessing the effect of the *CLPG* mutation on the microRNA catalogue of skeletal muscle using high throughput sequencing.

Florian Caiment[1], Carole Charlier[1], Tracy Hadfield[2], Noelle Cockett[2],

Michel Georges[1] & Denis Baurain[1]

[1]Unit of Animal Genomics, Department of Animal Production, GIGA-R and Faculty of Veterinary Medicine, University of Liège (B34), 1 Avenue de l'Hôpital, 4000-Liège, Belgium

[2]Department of Animal, Dairy and Veterinary Sciences, Utah State University, Logan, Utah 84322, United States of America

Correspondence: Michel Georges: michel.georges@ulg.ac.be, +32-4-366.41.51 (Tel), +32-4-366.41.98 (Fax)

Abstract

The callipyge phenotype is a monogenic muscular hypertrophy that is only expressed in heterozygous sheep receiving the *CLPG* mutation from their sire. The wild-type phenotype of *CLPG/CLPG* animals is thought to result from translational inhibition of paternally expressed *DLK1* transcripts by maternally expressed miRNAs. To identify the miRNA responsible for this trans-effect we used high-throughput sequencing to exhaustively catalogue miRNAs expressed in skeletal muscle of sheep of the four *CLPG* genotypes. We have identified 747 miRNA species of which 110 map to the *DLK1-GTL2* or callipyge domain. We demonstrate that the latter are imprinted and preferentially expressed from the maternal allele. We show that the *CLPG* mutation affects their level of expression in cis (~3.2-fold increase) as well as in trans (~1.8-fold increase). In *CLPG/CLPG* animals, miRNAs from the *DLK1-GTL2* domain account for ~20% of miRNAs in skeletal muscle. We show that *CLPG* genotype affects the levels of A to I editing of at least five pri-miRNAs of the *DLK1-GTL2* domain, but that levels of editing of mature miRNAs are always minor. We present suggestive evidence that the miRNAs from the domain target the ORF of *DLK1* thereby causing the trans-inhibition underlying polar overdominance. We highlight the limitations of high throughput sequencing for digital gene expression profiling as a result of biased and inconsistent amplification of specific miRNAs.

Introduction

The callipyge phenotype is an inherited muscular hypertrophy of sheep. It is characterized by an unusual inheritance pattern referred to as polar overdominance: only heterozygous individuals having received the *CLPG* mutation from their father express the phenotype (Cockett et al. 1996). The *CLPG* point mutation inactivates a muscle specific silencer controlling the expression of a subset of imprinted genes in the *DLK1-GTL2* domain (i.e. the paternally-expressed protein-encoding *DLK1* and *PEG11* (also known as *RTL1*) genes, and the maternally-expressed non-coding *GTL2* (also known as *MEG3*), anti-*PEG11* (also known as anti-*RTL1*), *MEG8* (also known as *RIAN*) and *MIRG* genes) (Charlier et al. 2001a; Freking et al. 2002; Smit et al. 2003). Hence padumnal heterozygotes ($+^{Mat}/CLPG^{Pat}$) are

characterized by ectopic expression of PEG11 (Byrne et al., 2010) and DLK1 (Davis et al., 2004) in skeletal muscle. DLK1 is thought to contribute to the callipyge phenotype as its ectopic expression increases muscle mass in transgenic mice (Davis et al. 2004). Whether ectopic expression of PEG11 is also involved in phenotypic expression remains to be established.

While showing increased levels of *DLK1* mRNA in muscle - as their $+^{Mat}/CLPG^{Pat}$ counterparts -, no DLK1 protein is observed in muscle of *CLPG/CLPG* animals, accounting for their wild-type phenotype (Davis et al. 2004). The absence of DLK1 protein despite increased levels of *DLK1* mRNA in these animals is thought to result from the ectopic expression of the madumnal non-coding RNAs, as this feature distinguishes *CLPG/CLPG* from $+^{Mat}/CLPG^{Pat}$ individuals (Georges et al. 2003; Georges et al. 2004). The madumnal long non-coding RNA genes are hosting a large array of small C/D snoRNAs and miRNAs of unknown function (Seitz et al. 2004). We have postulated that these small RNAs are the mediators of the *trans*-effect downregulating *DLK1* in *CLPG/CLPG* animals thus causing polar overdominance (Georges et al. 2003; Georges et al. 2004) (Suppl. Fig. 1). This hypothesis received strong support from the demonstration, in the same *DLK1-GTL2* locus, of RNAi-mediated trans-inhibition of the paternally expressed *PEG11* by miRNAs processed from the maternally expressed anti-*PEG11* transcript (Seitz et al. 2003; Davis et al. 2005).

To identify small RNAs that might be involved in the trans-inhibition of *DLK1* in skeletal muscle of *CLPG/CLPG* animals we have performed high-throughput sequencing (HTS) of small RNA libraries generated from skeletal muscle of sheep of the four possible *CLPG* genotypes. To qualify as mediators of the trans-effect underlying polar overdominance (Georges et al. 2003; Georges et al. 2004), the corresponding small RNAs should (i) map to the *DLK1-GTL2* domain, (ii) be imprinted with expression from the maternal allele, (iii) be subject to the cis-effect of the *CLPG* mutation (i.e. being ectopically expressed in skeletal muscle on maternal transmission of the mutation), and (iv) have the ability to guide the RISC complex to *DLK1* transcripts for inhibition.

Results

A catalogue of miRNAs expressed in skeletal muscle of sheep.

The callipyge phenotype is most pronounced in muscles of the hind quarters and manifests itself at ~ one month of age. At eight weeks of age, DLK1 protein is detected in skeletal muscle of $+^{mat}/CLPG^{pat}$ but not of *CLPG/CLPG* animals, suggesting that the *trans*-effect operates at that stage (Davis et al. 2004). Thus, we elected to extract RNA from longissimus dorsi (LD) of two 8-week old animals per C*LPG* genotype. Small RNA (~18 to 30 bp) libraries were generated and sequenced on a Genome Analyzer I (Illumina, San Diego, California). We obtained an average of 6,324,668 reads per animal (range: 5,222,920 – 6,685,342). Filtered, adaptor-trimmed sequences (94.7%) were aligned to the bovine genome used as reference, with the exception of 390 Kb of ovine sequence corresponding to the *DLK1-GTL2* domain (GenBank AF354168). The resulting alignments were used to predict miRNA precursors using miRDeep (Friedlander et al. 2008). This yielded 472 precursors, capturing 98.3% (range: 98.15-99.32%) of trimmed reads. Sequence comparison with precursors in miRBase combined with mapping data, indicate that 228 and 87 are the orthologues and paralogues, respectively, of previously reported bovine miRNAs (Glazov et al. 2009), while nine are the orthologues of miRNAs described in non-ruminant mammals. Thus, 148 precursors might correspond to previously unknown miRNAs (Suppl. Table 1). The chromosomal distribution of miRNA precursors is shown in Suppl. Fig. 2. Chromosome 21, harboring the *CLPG* locus, stands out with 61 precursors (chromosomal average: 14.8).

As the genomic sequence of sheep is not completed, we used the bovine sequence as reference. The effect of this substitution was estimated by comparing the output of miRDeep using either the bovine or the ovine sequence of the *DLK1-GTL2* domain as reference. miRDeep predicts 49 precursors in this domain when using the ovine sequence, of which three are missed when using the bovine reference. Sensitivity is thus decreased by ~6%, but specificity does not seem affected.

We aligned all precursor pairs using BLASTN and used the bitscores (> 35 bits) to identify precursors families using the MCL algorithm (Enright et al. 2002). The unique inflation parameter was set to the most aggregative value (2.0) for which the miR-376 family was recovered without contaminants (Seitz et al. 2004). Using this approach, 256 of the 472

precursors (=54%) clustered in 62 families. The largest family (miR-2284 family) comprised 99 members. The remaining 61 families counted 2.6 members on average (range: 2-6).

While for 197 (=41.7%) precursors we observed reads mapping either to the 5p or 3p arm of the pre-miRNA, both types of reads were observed for the remaining 275 (=58.3%), jointly defining 747 distinct miRNA "species". The fraction (F) of 5p over total reads distinguishes five types of precursors (Landgraf et al. 2007): 5p-mature/3p-star (1>F> 0.87), 5p > 3p (0.87>F>0.50), 5p=3p (F=0.5), 5p < 3p (0.50>F>0.13) and 5p-star/3p-mature (0.13>F>0). The frequency distribution of F-values is shown in Suppl. Fig. 3. The five types represent respectively 46.0, 10.2, 1.2, 10.2 and 32.4% of precursors. Precursors spawning miRNAs preferentially from the 5p arm were 1.4 times more abundant than those with 3p excess.

Aligning the reads with the identified precursors revealed considerable 3' length variability. As this might reflect trimming artifacts due to decreased sequencing fidelity towards the 3'end, we will not elaborate further on it. 5' ends were in general more consistent, with nevertheless considerable evidence for the occurrence of isomirs (Morin et al. 2008). For 65% of the miRNAs ≥ 90% of reads shared the same 5' extremity, for 27% ≥ 90% of reads shared one of two 5' extremities, and for 6% ≥ 90% of reads shared one of three 5' extremities. In general, more than 91% of alternative 5' extremities were within 4-bp of the most common one.

Annotating miRNAs expressed from the *DLK1-GTL2* domain.

Forty-nine of the precursors identified by miRDeep mapped to the *DLK1-GTL2* domain. Of these, 39 corresponded to known miRNAs reported in miRBase, while 10 were unknown. Detailed examination of the SOAP (Li et al. 2008) alignments, revealed 5,729 reads mapping to 14 regions not recognized by miRDeep as miRNA precursors. Six of these corresponded to miRNAs reported in miRBase and were included in the catalogue. In addition, 487 reads mapped to 12 predicted C/D snoRNAs within *MEG8* (of note, bona fide C/D snoRNAs are ~80bp long and would therefore have been excluded from the small RNA libraries). We found no reads for fourteen miRNAs reported either in human and/or in mice (of which five conserved in sheep).

5’ ends showed the level of variability observed in the genome-wide catalogue, i.e. respectively 58%, 34% and 8% of miRNAs with one, two and three isomirs representing ≥ 90% of the reads. Remarkably, 49 precursors (89.1% of the expressed precursors) had reads mapping to both 5p and 3p arms, to be compared with the genome-wide 58.3%.

A summary of all miRNA precursors identified in the *DLK1-GTL2* domain is given in Fig. 1. A total of 110 distinct small RNA species (not distinguishing isomirs) were identified, mapping to 61 miRNA and 12 C/D snoRNA precursors. All detected small RNAs derive from the same strand as *GTL2*, anti-*PEG11*, *MEG8* and *MIRG*. Using a 10-way mammalian sequence alignment of *MIRG*, we generated a plot of sequence conservation within 8-nt windows (Suppl. Fig. 4A). We observed a striking coincidence between the peaks of conservation and the positions of the miRNAs, supporting miRNA-generation as primary function of *MIRG*. A similar colocalisation of conservation peaks and C/D snoRNAs is not observed for *MEG8* (Suppl. Fig. 4B).

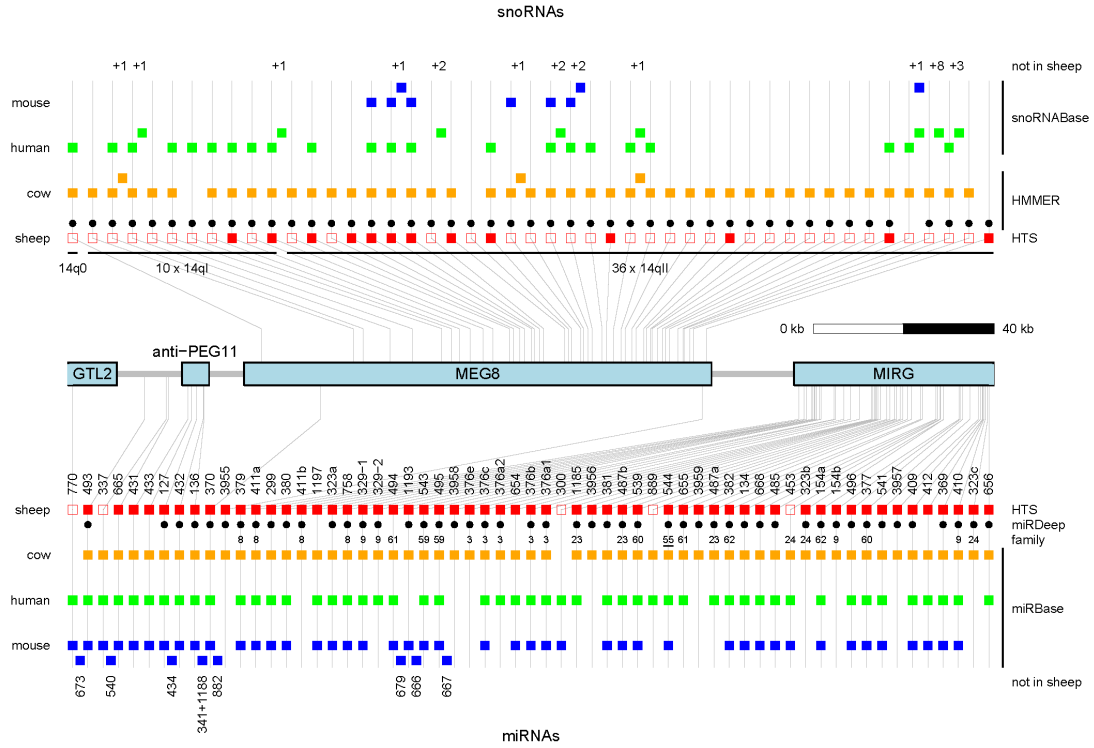


Figure 1: Comparative map of the small RNA genes in the *DLK1-GTL2* domain: snoRNAs (upper panel) and miRNAs (lower panel). Square boxes correspond to small RNAs detected in sheep (red), cow (orange), human (green) and mouse (blue). Gray lines connect orthologues in the four species and indicate their chromosomal position with respect to the four long non-coding RNA genes in the domain: *GTL2*, *anti-PEG11, MEG8* and

MIRG. The position of precursors not detected in sheep is indicated by squares nested between vertical gray lines. The numbers above and below snoRNA and miRNA columns, respectively, correspond to numbers of additional paralogues for the snoRNAs (from +1 to +8) and names of additional miRNAs. The red squares are filled when reads for the corresponding small RNA were found in the conducted HTS experiments, empty when not. miRNAs predicted by miRDeep (Friedlander et al. 2008) in sheep are marked by black dots. Numbers below the black dots identify the cluster/family to which the corresponding miRNA was assigned using the BLAST/MCL algorithm (Enright et al. 2002). The family number of miR-544 is underlined as the other members map outside of the DLK1-GTL2 domain. SnoRNAs predicted by HMMER (Durbin et al. 1998) in sheep are marked by black dots. snoRNAs in mouse and human correspond to predictions made by Cavaille et al. (2002). snoRNAs in the cow were predicted by HMMER (Durbin et al. 1998). miRNAs in cow, human and mouse were extracted from miRBase (Griffiths-Jones 2006).

Limits of HTS for the quantitative assessment of miRNA expression.

Read numbers are assumed to faithfully reflect expression levels, allowing for accurate digital gene expression profiling. However, recent data indicate that the amplification steps during library construction may introduce substantial, protocol-specific biases (Linsen et al. 2009). To evaluate accuracy and precision of our HTS data in measuring miRNA expression, we (i) repeated the HTS experiment for seven of the eight animals (including RNA extraction, library construction and sequencing on an Illumina GA-II instrument), (ii) hybridized skeletal muscle (LD) RNA from the eight animals on Exiqon miRCURY™ LNA (Version 9.2 - updated to miRBase 11.0) arrays (GEO GSE24146), and (iii) performed QRT-PCR for eight miRNAs spanning a broad range of expression levels as determined by HTS. While the Exiqon arrays allow interrogation of 569 human miRNAs, we restricted the analysis to 265 for which the LNA probes were perfectly complementary to the orthologous ruminant miRNA.

The main conclusions of this experiment can be summarized as follows:

i. Spearman rank correlations (r_S) between sequencing replicates were 0.80 on average, thus suggesting adequate reproducibility of HTS (Suppl. Fig. 5). Note that correlations were slightly higher when comparing pairs of animals with same *CLPG* genotype within sequencing runs (average r_S = 0.83; data not shown). Examination of specific miRNAs, however, highlighted limitations of digital expression profiling by HTS. Hence, while miR-127 accounted on average for 4% of reads originating from the *DLK1-GTL2* domain in the first experiment, its contribution increased to 25% on average in the second, pointing towards systematic discrepancies between the two experiments for some miRNAs.

Moreover, while miR-1 represented ≥ 83% of reads (average 86%) in the first series of eight libraries, and ≥ 80% of reads (average 82%) in 5/7 libraries of the second series, it only reached 32% and 49% in the two remaining ones, thus showing substantial discrepancies even within experiment. Finally, within sequencing experiments, the 5p/3p ratio differed significantly between individuals for nearly all miRNA precursors (chi squared test). In extreme cases, different individuals would appear to have inverted 5p/3p ratios despite sequence depths of hundreds and even thousands. In no case were these opposite 5p/3p ratios confirmed in the second experiment. A representative example (miR-382) is shown in Suppl. Fig. 6. Thus, while the repeatability of HTS may seem satisfactory in general, our findings suggest that amplification efficiency of specific miRNAs may considerably vary between experiments.

ii. r_S values between expression levels (ranks) assessed using the Exiqon miRCURY™ LNA arrays averaged 0.86 between individuals of same *CLPG* genotype. This value has to be compared with a value of 0.90 when restricting the HTS data to the 265 miRNAs interrogated with the Exiqon array. Thus, in these experiments HTS and array hybridization were characterized by comparable reproducibility. Yet, when comparing ranks obtained with the two methods, r_S values dropped to 0.63 (first sequencing experiment) and 0.68 (second sequencing experiment) (Suppl. Fig. 7A&B). Suppl. Fig. 7C&D illustrate the impact of this correlation drop in terms of probability of reversed rank order between alternative methods as a function of observed fold difference in expression level: miRNA pairs showing a five-fold difference in expression level on the Exiqon arrays still have a probability of ~0.15 to be ranked inversely by HTS. Of note, the RNAs hybridized on the arrays were not size-selected (thus potentially including pri-, pre- and mature miRNA molecules), while the RNAs used to construct the libraries for HTS were size-selected to include only mature miRNAs.

iii. Given the observed discrepancies between the HTS and array-hybridization, we performed QRT-PCR for eight miRNAs (let-7d, miR-1, miR-206, miR-127, miR-382-5p, miR-382-3p, miR-3958, miR-3959) spanning a range of expression levels using

the looped RT primer approach (targeting mature miRNAs) (Chen et al. 2005). QRT-PCR experiments were conducted in duplicate on the RNA samples used for HTS. Abundance of miRNA "x" relative to let-7d was estimated as $\varepsilon_x^{Ct_x} / \varepsilon_{let7d}^{Ct_{let7d}}$ where ε's are the experimentally determined amplification efficiencies and *Ct*'s the threshold exceeding cycle numbers. From these analyses it appeared that the QRT-PCR results were more consistent with the array-hybridization than with HTS in terms of expression ranks and estimated fold differences in expression levels (Suppl. Fig. 8A&B). Our data strongly suggest that some miRNAs undergo preferential amplification during the HTS procedure.

Effect of CLPG genotype on relative expression levels of miRNAs in the *DLK1-GTL2* domain.

The previous findings call for caution when interpreting variations in expression levels of individual miRNAs. To overcome this limitation we examined the effect of *CLPG* genotype on the expression level of the miRNAs from the *DLK1-GTL2* domain considered as a group.

We first confirmed the previously described cis-effect of the *CLPG* mutation on neighboring genes in the sequenced RNA samples. QRT-PCR experiments were conducted using primer sets specific for mature *DLK1* and *GTL2* transcripts, and for two internal controls (*RPLP0, RPS18*) selected with geNorm out of five housekeeping genes (Vandesompele et al. 2002). The expected *CLPG* effects were clearly observed (Suppl. Fig. 9). Expression levels of *DLK1* were increased ~12-fold and ~4-fold in, respectively, $+^{Mat}/CLPG^{Pat}$ and *CLPG/CLPG* animals when compared to *+/+*, while being slightly decreased (~0.6) in $CLPG^{Mat}/+^{Pat}$. Expression levels of *GTL2* were increased ~30-fold, ~14-fold and ~5-fold in *CLPG/CLPG,* $CLPG^{Mat}/+^{Pat}$ and $+^{Mat}/CLPG^{Pat}$, when compared to *+/+* animals. These results were undistinguishable from the ones that were previously reported using samples originating from other animals (Charlier et al. 2001a; Davis et al., 2004; Davis et al., 2005). The previously observed ~5-fold increase of *GTL2* expression in $+^{Mat}/CLPG^{Pat}$ when compared to *+/+* animals, and ~2-fold increase of *GTL2* expression in *CLPG/CLPG* when compared to $CLPG^{Mat}/+^{Pat}$ animals remains particularly intriguing and points towards a trans-effect of the

padumnal *CLPG* mutation on the expression level of the madumnal non-coding RNA genes (Charlier et al. 2001a).

We then analyzed the HTS data. Read numbers corresponding to a given miRNA species (i.e. mapping either to the 5p or 3p arm of a precursor) were first adjusted to account for the different numbers of total "mappable" reads per individual. Relative expression level for a given animal was expressed as $\log_2(i/m)$, where *i* corresponds to the adjusted number of reads for that individual and *m* is the experiment-specific average number of adjusted reads for that miRNA across the seven individuals that were sequenced twice. Average $\log_2(i/m)$ across miRNAs differed considerably between individuals, including for miRNAs outside of the *CLPG* locus. This was thought to reflect experimental issues rather than genuine biological differences (Suppl. Fig. 10). Therefore, $\log_2(i/m)$ values were corrected for the average $\log_2(i/m)$ value across miRNAs mapping outside of the *DLK1-GTL2* domain (for that individual). We then tested the effect of *CLPG* genotype on the corrected relative expression levels by ANOVA, using both sequencing experiments jointly. Fig. 2A shows the corresponding log(1/p) values. The effect of *CLPG* genotype on the relative expression level of miRNAs from the *DLK1-GTL2* domain is clearly visible from the localized cluster of significant log(1/p) values. Six (miR-379, miR-411a, miR-495, miR-154b, miR-655 and miR-299) of the 99 "regular" miRNAs (i.e. excluding small RNAs derived from C/D snoRNAs) exhibited p-values < 6×10^{-5}, corresponding to the Bonferroni-corrected 5% threshold, while 47 were characterized by nominal p-values < 0.05. Fig. 2B shows, for each of the eight studied animals, the average relative expression levels over all "regular" miRNAs of the *DLK1-GTL2* domain. The order is identical to that observed for the long non-coding RNA genes including *GTL2*: *CLPG/CLPG* > $CLPG^{Mat}/+^{Pat}$ > $+^{Mat}/CLPG^{Pat}$ > *+/+*. The magnitude of the effect, however, was smaller: expression levels were increased ~6.4 fold, ~4.4 fold and ~2.0 fold in *CLPG/CLPG*, $CLPG^{Mat}/+^{Pat}$ and $+^{Mat}/CLPG^{Pat}$ when compared to *+/+* animals.

The previous figures pertain to "regular" miRNAs from the *DLK1-GTL2* domain. As can be seen from Fig.2A the effect of *CLPG* genotype on the expression level of small RNAs derived from C/D snoRNAs were not significant, suggesting that *MEG8* might escape the *cis*-effect of the *CLPG* mutation, contradicting previous findings (Charlier et al. 2001a). Examination of the effect of *CLPG* genotype on C/D snoRNA-derived small RNAs, however,

revealed the expected trend in all but one $+^{Mat}/CLPG^{Pat}$ individual (Suppl. Fig. 11). Expression levels of C/D snoRNA-derived species (average number of reads: 70; median: 6) were low when compared to miRNAs (average number of reads: 21,300; median: 249). Low levels, combined with the aberrant behavior of one individual, explains the non-significance of the *CLPG* effect on the expression level of small RNAs derived of C/D snoRNAs, which we nevertheless believe exists.

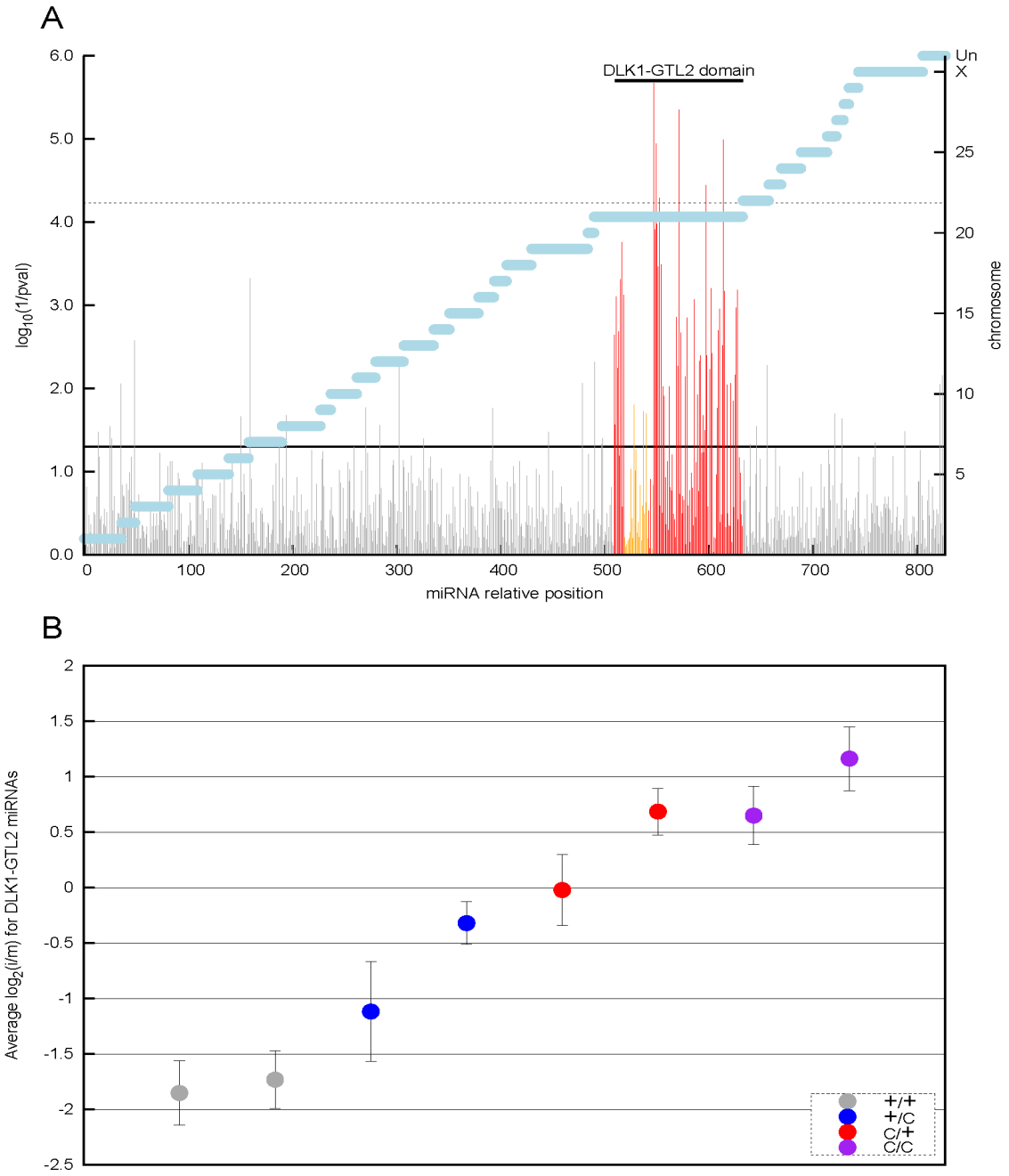


Figure 2: A $Log_{10}(1/p)$ values of the effect of *CLPG* genotype on the expression level of 851 small RNAs in skeletal muscle of eight 8-week old sheep. Expression levels were estimated from the number of Illumina GA reads from two independent HTS experiments**.** The statistical significance of the *CLPG* effect was estimated by ANOVA. Gray vertical bars correspond to miRNAs outside of the *DLK1-GTL2* domain, red vertical bars to miRNAs from the *DLK1-GTL2* domain and orange vertical bars to small RNAs derived from C/D snoRNA precursors. Horizontal black lines correspond to the nominal (plain line) and Bonferroni-adjusted (dotted line) 5% significance thresholds. Horizontal blue bars mark the different chromosomes (right Y-axis). UN: correspond to unassigned sequence contigs. **B.** Average expression level, relative to the mean expression level of seven individuals sequenced twice (HTS1 and HTS2), of 99 "regular" miRNAs (i.e. excluding small RNAs derived from C/D snoRNAs) from the *DLK1-GTL2* domain in skeletal muscle of eight sheep sorted by *CLPG* genotype (gray: *+/+;* blue: $+^{Mat}/CLPG^{Pat}$; red: $CLPG^{Mat}/+^{Pat}$; purple: *CLPG/CLPG*). Error bars correspond to 1.96 x the standard error of the estimate.

The effect of *CLPG* genotype on relative miRNA expression levels was also evaluated from the array data (Suppl. Fig. 12A&B). The effect of *CLPG* genotype was equally clear, manifesting itself as a clustered rise in log(1/p) values. Expression levels were increased ~4.8 fold, ~3.1 fold and ~2.2 fold in *CLPG/CLPG*, $CLPG^{Mat}/+^{Pat}$ and $+^{Mat}/CLPG^{Pat}$, when compared to *+/+* animals. Hence, the ranking was as expected (*CLPG/CLPG* > $CLPG^{Mat}/+^{Pat}$ > $+^{Mat}/CLPG^{Pat}$ > *+/+*), yet the magnitude of the effect slightly lower than the HTS estimates.

Imprinting status of miRNAs in the *DLK1-GTL2* domain.

The obvious interpretation of miRNA expression levels in $+^{Mat}/CLPG^{Pat}$ and $CLPG^{Mat}/+^{Pat}$ intermediate between *+/+* and *CLPG/CLPG* (Fig. 2B) is that most miRNAs from the *DLK1-GTL2* domain are not imprinted, yet affected by the *CLPG* cis-effect. All previous evidence, however, indicates that the maternally expressed non-coding RNA genes, including the embedded C/D snoRNA and miRNA genes, are exclusively expressed from the maternal allele. In the mouse, tested C/D snoRNAs and miRNAs from the *DLK1-GTL2* domain were expressed in mice with maternal uniparental disomies of chromosome 12 (mUPD12) but not with paternal UPD12 (Cavaille et al. 2002; Seitz et al. 2003; Seitz et al. 2004). The same small RNA genes were expressed in mice inheriting a deletion of the IG-DMR imprinting control element when on the paternal allele, but not when on the maternal allele (Lin et al., 2003; Seitz et al., 2004). In human, *MEG8* (also known as *RIAN*) from which the C/D snoRNA genes are processed (Cavaille et al. 2002), was not expressed in patients with pUPD14 (Kagami et al. 2008). We have previously shown that in sheep muscle, *anti-PEG11* and *MEG8* (hosting

miRNAs and C/D snoRNAs, respectively) are exclusively expressed from the maternal allele, irrespective of *CLPG* genotype (Charlier et al. 2001a; Charlier et al. 2001b).

To more directly assess the imprinting of the miRNAs from the *DLK1-GTL2* domain in sheep and the effect of the *CLPG* mutation on it, we searched for SNPs in the vicinity of pre-miRNAs for which at least one of the four studied $CLPG^{Mat}/+^{Pat}$ or $+^{Mat}/CLPG^{Pat}$ animals would be heterozygous. We found nine such SNPs tagging six pre-miRNAs. Seven of these SNPs were within 160 bp from the corresponding pre-miRNA (miR-379, miR-134, miR-485, miR-453, miR-154b), one was in the loop (miR-453) and one at position 20 of the miRNA* (miR-377). One $+^{Mat}/CLPG^{Pat}$ animal was homozygous for all SNPs and hence non-informative, but the other three were heterozygous for most (Table 1). For each SNP the allele associated with the *CLPG* mutation was determined by sequencing a *CLPG/CLPG* animal.

miRNA	miR-379	miR-134		miR-485		miR-453		miR-154b	miR-377
Position	80bp-5'	3'-32bp	3'-160bp	3'-78bp	3'-94bp	57bp-5'	loop	69bp-5'	miR*
CLPG:+	T:C	A:C	A:C	T:C	A:G	T:C	T:C	T:C	T:C
$(CLPG/+)_1$		100/0	100/0	100/0	98/2	99/1	99/1	97/3	14/0
$(CLPG/+)_2$		100/0	95/5	84/16	86/14	100/0		96/4	7/0
+/CLPG	98/2	94/6	97/3	83/17	88/12	95/5		98/2	46/0

Table 1: Imprinting status of miRNAs in the ovine *DLK1-GTL2* domain and effect of the CLPG genotype. Line 1: name of the tested miRNA; line 2: position of the interrogated SNP with respect to the corresponding pre-miRNA; line 3: SNP alleles associated wit the *CLPG* and + allele; line 4-6: miR-379-miR-154b: % of the madumnal/padumnal alleles found in cDNA; miR-377: number of HTS reads corresponding to the madumnal/padumnal allele

For miR-377 and knowing that the SNP mapped to the miRNA*, we determined imprinting status from HTS data. We exclusively detected reads corresponding to the madumnal allele, both in the $CLPG^{Mat}/+^{Pat}$ and $+^{Mat/}CLPG^{Pat}$ animals, thus supporting tight imprinting, exclusive madumnal expression and no effect of *CLPG* genotype on imprinting. For miRNAs with SNPs lying outside of the mature miRNAs, we amplified the pri-miRNA with primers within 179 bp from the pre-miRNA, directly sequenced the resulting amplicons and measured the allelic ratio using PeakPicker (Ge et al. 2005). For four of the five miRNAs the results were identical to miR-377 (Table 1): tight imprinting, near exclusive madumnal expression, no effect of *CLPG* genotype. For miR-485, however, we observed relaxation of

imprinting in one *CLPG*^{Mat}/+^{Pat} and one +^{Mat}/*CLPG*^{Pat} animal, for which ≈20% of transcripts were derived from the padumnal allele. There was no evidence for relaxation of imprinting of miR-485 in the other informative *CLPG*^{Mat}/+^{Pat} animal (Table 1). It is noteworthy that miR-134 located 588 bp upstream, and miR-154b located 4,293 bp downstream of miR-485 did not show evidence for relaxation of imprinting in the same individuals.

Effect of *CLPG* genotype on absolute expression levels of miRNAs in the *DLK1-GTL2* domain.

The previous analyses are not informative about absolute miRNA expression levels: are they just present in minute, biologically irrelevant amounts or do they make a significant contribution to the pool of miRNAs in muscle? Analysis of the two sequencing experiments indicate that, after exclusion of miR-1 whose read numbers were inflated (see above), the percentage of reads originating from miRNA precursors in the *DLK1-GTL2* domain was 4% in *+/+* animals, but increased to 10%, 11% and 21% of the total in +^{Mat}/*CLPG*^{Pat}, *CLPG*^{Mat}/+^{Pat} and *CLPG/CLPG* animals. The Exiqon arrays only measure the abundance of some miRNAs present in a tissue. However, as the proportion of miRNAs mapping to the *DLK1-GTL2* domain was virtually identical in the HTS (99/826=12%) and hybridization data (34/265=12.8%), the ratio of the sum of fluorescence intensities for miRNAs in the domain over the sum of fluorescence intensities over all miRNAs would also provide an estimate of the cellular abundance of miRNAs originating from the *DLK1-GTL2* domain. miRNAs from the domain accounted for 3.5%, 9.0%, 15.6% and 22.3% of the Hy3 fluorescence on the Exiqon arrays in *+/+*, +^{Mat}/*CLPG*^{Pat}, *CLPG*^{Mat}/+^{Pat} and *CLPG/CLPG* animals. Both approaches thus provided comparable estimates, indicating that the *DLK1-GTL2* domain contributes a sizeable fraction of the miRNA population, especially in *CLPG/CLPG* animals in which they are predicted to mediate the trans-inhibition of *DLK1*.

While claims about expression levels of individual miRNAs are hazardous for the reasons mentioned before, the QRT-PCR and array experiments strongly suggest that miRNAs from the *DLK1-GTL2* domain are characterized by an at least 30-fold range of expression levels (Suppl. Fig. 8 & 13).

Effect of *CLPG* genotype on relative expression levels of miRNAs outside the *DLK1-GTL2* domain.

Skeletal muscles that express the callipyge hypertrophy have a profoundly altered physiology. Ectopic expression of DLK1 (Davis et al. 2005) triggers a cascade of secondary events leading to muscular hypertrophy (e.g. Vuocolo et al. 2007). These may involve altered miRNA expression. To detect such secondary miRNA perturbations, we tested the effect of *CLPG* genotype on relative expression levels of miRNAs outside of the domain. We first considered the HTS and array hybridization data separately. When accounting for multiple testing, no miRNA outside of the *DLK1-GTL2* domain appeared to be significantly affected by *CLPG* genotype (Fig. 2 and Suppl. Fig. 12). In an attempt to increase power, we combined HTS and array data for the 265 miRNAs with information on both platforms. Even then, no miRNA outside of the domain was significant (Suppl. Fig. 14). We conclude that the molecular events connecting ectopic expression of DLK1 and muscular hypertrophy do not involve altered miRNA expression.

Editing of miRNAs from the *DLK1-GTL2* domain in skeletal muscle.

It was recently observed that a cluster of miRNAs mapping to the *DLK1-GTL2* domain [human miR-368, miR-376a1 (5' and 3'), miR-376b and miR-376a2 (5' and 3'); murine mir-376a (5' and 3'), mir-376b (5' and 3') and mir-376c] undergo extensive A to I editing in human and mice, particularly in the central nervous system (Kawahara et al. 2007). The most extensively edited sites correspond to position "+3" or "+4" of the 5p miRNAs and position "+6" of the 3p miRNAs. Analyses performed in KO mice suggest that 5p editing is ADAR2 (also known as ADARB1) dependent, while 3p editing is ADAR1 (also known as ADAR) dependent. Editing seemed not to affect processing, as equally high levels were observed in pri-miRNAs and derived mature miRNAs. By changing the seed, editing was predicted to alter the target spectrum.

As editing of miRNA seeds from the *DLK1-GTL2* domain in sheep may likewise alter affinity for *DLK1*, we systematically searched for it. We first examined whether the precursors of mir-376a,b,c undergo editing in skeletal muscle of mice. We RT-PCR amplified the corresponding pri-miRNAs from cDNA of brain, kidney and skeletal muscle (quadriceps

femoris) from an FVB mouse, and sequenced the corresponding PCR products. Strong (~ 80%) and moderate (~ 60%) editing of the "+44" 3p position (residue "+6" of the mature miRNA) was observed for the three studied miRNAs in, respectively, brain and kidney, hence recapitulating part of the results of Kawahara et al. (2007). Contrary to these authors, we found no evidence for editing of the 5p arms, whether in or upstream of the mature miRNA sequence. No editing was observed in skeletal muscle of mice (data not shown).

We then scanned the pri-miRNAs corresponding to 56 precursors from the *DLK1-GTL2* domain using RNA extracted from skeletal muscle (LD) of one *CLPG/CLPG* and one *+/+* sheep. Within the miR-376 cluster, we did observe substantial levels of editing of the "+44" 3p position (corresponding to position "+6" of the mature miRNA) of miR-376e (0-25%), miR-376c (also known as miR-368)(5-45%), miR-376a2 (0-50%), miR-376b (0-95%), but not of miR-654 and miR-376a1. No editing was observed in the 5p arm for any of these pri-miRNAs. Outside of the miR-376 cluster, we observed strong editing of three other pri-miRNAs: at the equivalent "+44" 3p position (corresponding in this case to position "+5" of the mature miRNA) for miR-381 (10-82%), at 5p position "+5" (corresponding to position "+5" of the mature miRNA) for miR-411a (0-20%), and at 5p position -4 outside of the mature miRNA sequence for miR-369 (0-40%). Note that neither miR-381 nor miR-411a show obvious similarity with members of the miR-376 cluster. Based on these results, we evaluated the level of editing of miR-376e, miR-376c, miR-376a2, miR-376b, miR-381 and miR-411a in 16 additional animals representing the four possible *CLPG* genotypes at two and eight weeks of age. Editing of miR-376c, miR-376b and miR-381 was observed in some of the new animals, but not of miR-376e, miR-376a2 and miR-411a. Unexpectedly, we observed a highly significant ($p \leq 1.3 \times 10^{-5}$) effect of *CLPG* genotype on the level of editing of miR-376c, miR-376b and miR-381: *+/+* animals had markedly higher levels of pri-miRNA editing than the three other genotypes (Fig. 3A).

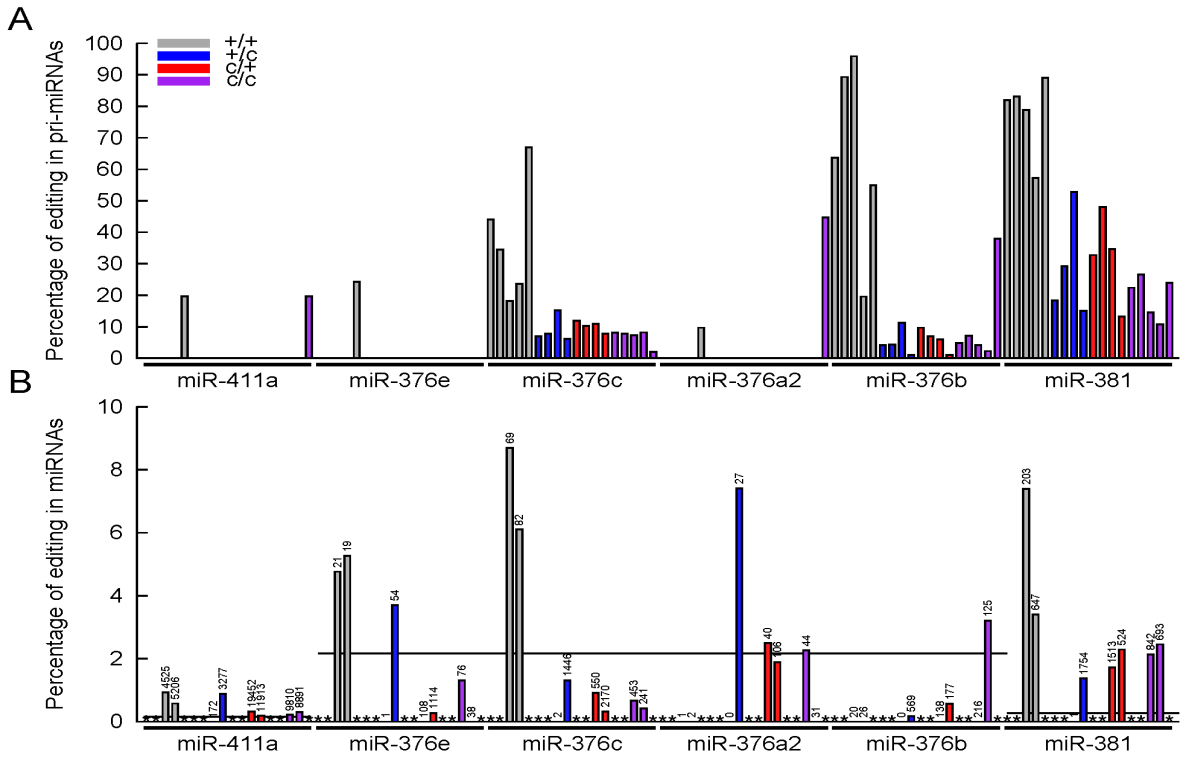


Figure 3: (A) Percentage of A to I editing of pri-miRNAs at "+5" position (pre-miR-411a) or "+44" position (pre-miR-376e, pre-miR-376c, pre-miR-376a2, pre-miR-381) in longissimus dorsi of 18 animals representing the four possible CLPG genotypes (gray: *+/+* (5); blue: $+^{Mat}/CLPG^{Pat}$ (4); red: $CLPG^{Mat}/+^{Pat}$ (4); purple: *CLPG/CLPG* (5)). The first two animals of each CLPG genotype were analyzed at two weeks, the others at eight weeks. (B) Percentage of A to I editing of mature miRNAs at position "+5" (miR-411a = 5p), "+6" (miR-376e, miR-376c, miR-376a2, miR-376b = 3p) and "+5" (miR-381 = 3p)). Animals are ordered as in (A). Animals without HTS data are marked by *. The numbers above each column correspond to the total number of reads (edited + non-edited) available for analysis. The black horizontal lines correspond to the average level of A to G substitution observed for miRNAs derived from the 5p arm at position "+5" (miR-411a), from the 3p arm at position "+6" (miR-376e, miR-376c, miR-376a2, miR-376b) and from the 3p arm at position "+5" (miR-381).

To verify whether editing of the pri-miRNAs resulted in equivalent proportions of edited mature miRNAs (as observed by Kawahara et al. 2007), we evaluated the level of editing of mature miR-376e, miR-376c, miR-376a2, miR-376b, miR-381 and miR-411a in the HTS libraries (Fig. 3B). In general, editing levels of mature miRNAs were below 10%. For the three miRNAs with high levels of pri-miRNA editing (i.e. miR-376c, miR-376b, miR-381), levels dropped considerably in the fully processed miRNAs. This was most striking for miR-376b with virtually total absence of edited reads. For miR-376c and miR-381, editing levels dropped by a factor of ~10 when compared to the precursors. Thus for these miRNAs, editing either inhibits pri/pre-miRNA processing and/or reduces the stability of the miRNA.

The effect of *CLPG* genotype on editing levels was still apparent for miR-376c and miR-381. For miR-376a2, editing levels appeared higher after than before miRNA processing (although still well below 10%). In this case, editing may thus promote processing and/or stability. Finally, for miR-411a, mature editing levels were consistently of the order of 1%, which was well above background (average of 0.02% across 14 5p miRNAs with a A residue at position "+5"). Such levels would not have been reliably detected at the pri-miRNA level.

Evaluating the affinity of miRNAs in the *DLK1-GTL2* domain for *DLK1*.

Having generated an exhaustive catalogue of miRNAs expressed in skeletal muscle of *CLPG/CLPG* animals allowed us to test the miRNA-mediated *DLK1* trans-inhibition hypothesis with unprecedented power.

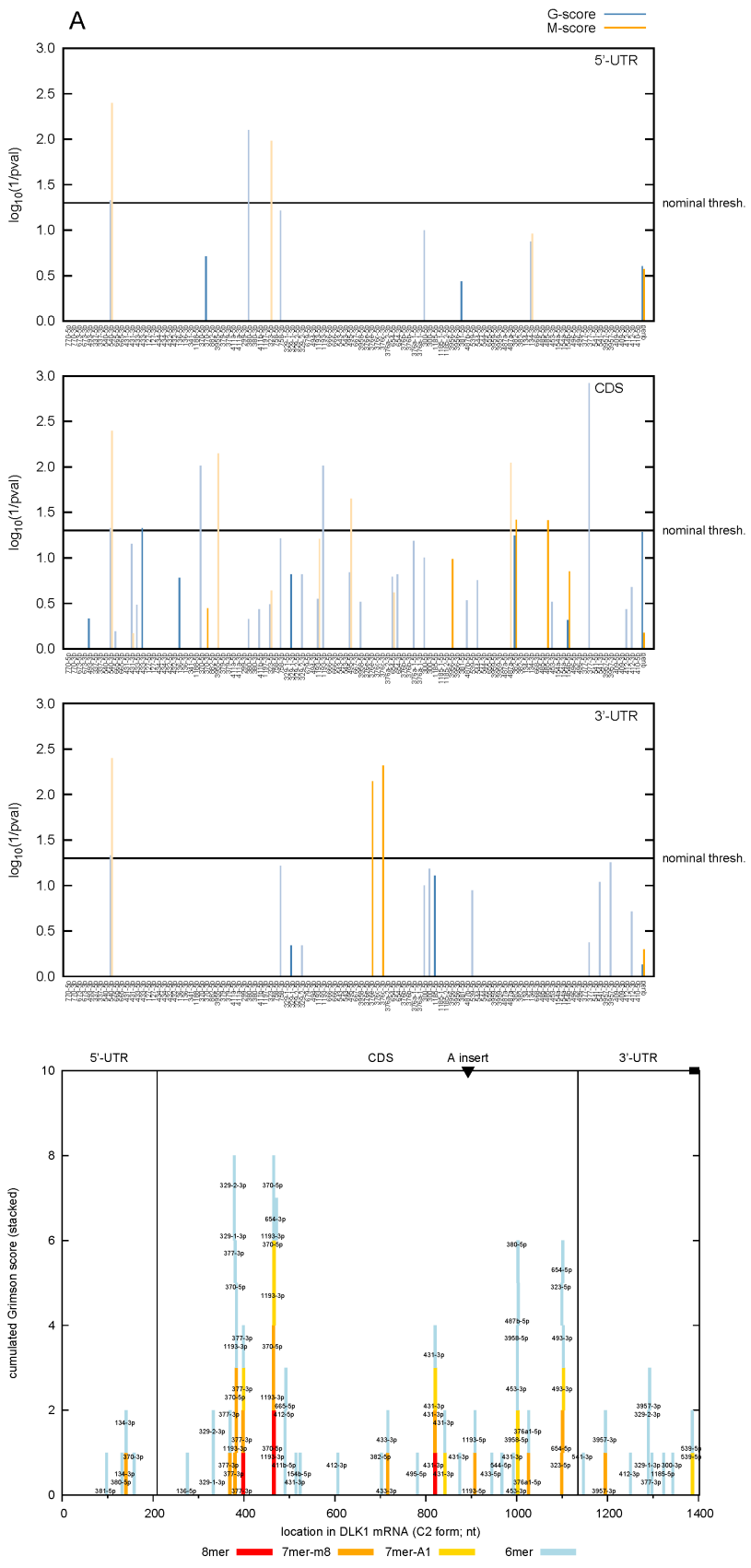
For each of the 114 miRNA species from the *DLK1-GTL2* domain, we singled out the most abundant isomir (or pair of isomirs in ex aequo cases, leading to 127 distinct sequences) and quantified its affinity for *DLK1* using two established metrics. The first one ("G-species-score") follows Grimson et al. (2007) and counts the occurrences of 6mer (Watson-Crick (WC) reverse complement of miRNA residues 2 to 7), 7mer-m8 (WC reverse complement of miRNA residues 2 to 8), 7mer-A1 (WC reverse complement of miRNA residues 2 to 7 plus 3' A anchor) and 8mer matches (WC reverse complement of miRNA residues 2 to 8 plus 3' A anchor) in *DLK1*. Thus a 8mer match would increase the "G-species-score" by four, that of a 7mer (without 8-nt match) by 2, and that of 6mer match by one. The second one ("M-species-score") sums scores (≥ 140) obtained with the more liberal miRNA-target miRanda identification engine (John et al. 2004). G-species-scores were summed to generate a "G-quadrille-score", and M-species-score were summed to generate a "M-quadrille-score. "Quadrille scores" evaluate the affinity for *DLK1* of the miRNAs considered as a team. Moreover, the same scores were generated for human and mouse (using species-specific miRNA sequences reported in miRBase), and corresponding scores summed across species to generate "multiorganism (MO) scores". The latter should be more effective at identifying an unusual affinity for *DLK1* if conserved across species. Whether the mechanisms underlying the trans-inhibition of *DLK1* observed in sheep are shared with other species, remains unknown.

To evaluate the statistical significance of the obtained metrics, we compared them with their distribution obtained on 10,000 random shuffles of the *DLK1* sequence. Shuffling was conducted such as to maintain the original trinucleotide composition of the target gene (cfr. Suppl. methods). While miRanda is expected to inflate the number of target predictions, their statistical significance should be well controlled by this approach (i.e. there is no reason why the true *DLK1* sequence should yield better miRanda scores than the shuffled sequences).

We first tested the approach using the *PEG11* ORF as positive control (1,000 shufflings). *PEG11* is indeed targeted by at least 6 miRNA species derived from five pre-miRNAs in *anti-PEG11* in ovine, human and mouse (miR-431, miR-433, miR-127, miR-432 (human & sheep), mir-434 (mouse), miR-136) (Davis et al., 2005). The targeting of *PEG11* by these miRNAs is "plant-like", relying on WC complementarity over the entire length resulting in target slicing (Davis et al., 2005). For proper comparison with the presumably "animal-like" situation of *DLK1*, we only considered the miRNA seed sequences (residues 1 to 8) when computing the G-species-scores (replacing the 3' A-anchor constraint (applying to 8mer and 7mer-A1 matches) by 3' WC complementarity). miRanda scores were computed as before. When considering the ovine sequence (ORF) alone, both quadrille-scores were significant (G: p=0.003; M: p=0.025), hence detecting the presence of one or more miRNAs targeting *PEG11*. Considered individually, however, none of the miRNA exceeded the Bonferroni-corrected 5% threshold. Two (of twelve) miRNAs processed from *anti-PEG11* (miR-431-5p and miR-136-3p) achieved nominal significance (p=0.003) for both G- and M-species-score. Intriguingly, the ovine-specific miR-3959-3p, although processed from *MIRG*, achieved equivalent significance (p=0.003) (Suppl. Fig.15A). When considering the three species simultaneously (hence exploiting evolutionary conservation known to exist), the significance of the two quadrille scores increased (G&M: p<0.001). Moreover, six (of twelve) miRNA species processed from *anti-PEG11* yielded the highest possible signal (nominal p<0.001; Bonferroni-corrected p ~ 0.10). Interestingly, miR-411a-5p processed from *MIRG* achieved the same top score, while the signal for miR-3959 remained essentially unchanged (nominal p=0.003) as this Laurasiatheria-specific miRNA is not shared with human and mouse (Suppl. Fig. 15B). The high miR-411a scores reflects one 7-mer and ten 6-

mer matches in mouse, one 7-mer and six 6-mers in human, and two 7-mers and two 6-mers in sheep. Three matches were conserved in the three species, and one in two species. The high ovine miR-3959 score is due to one 8mer and one 7-mer match. Thus, application of our method to *PEG11* indicated that: (i) significant quadrille-scores but not species- scores could be obtained without exploiting evolutionary conservation, (ii) significant quadrille- and species-scores could be obtained when exploiting conservation. Most interestingly, this analysis strongly suggests that the paternally expressed *PEG11* is not only targeted by fully complementary miRNAs processed from the maternally expressed *anti-PEG11* but also by miRNAs processed from the maternally expressed *MIRG* pri-miRNAs, which recognize their target via seed-dominated complementarity.

We then applied the same approach to *DLK1* including the 5'UTR, ORF and 3'UTR as it is established that miRNAs may target these different gene compartments (Baek et al. 2008; Selbach et al. 2008; Tay et al. 2008; Chi et al. 2009). When relying solely on ovine information, the most noteworthy result was the nearly significant G-quadrille-score ($p = 0.052$) on the *DLK1* ORF, hence suggesting an unusual affinity of the ovine miRNA team for this segment of *DLK1* (Fig. 4A). This signal was primarily driven by miR-377-3p (nominal $p = 0.0012$; one 8mer and two 7mer-m8 matches), miR-1193-3p (nominal $p=0.0098$; one 8mer and two 7mer-m8 matches) and miR-370-5p (nominal $p=0.0098$) one 8mer and one 7mer-m8 match) (Fig. 4B). Note that none of these miRNAs achieves Bonferroni-adjusted significance. There was no convincing evidence for miRNA targeting of the *DLK1* 5' nor 3' UTR.


A
G-score
M-score
5'-UTR
CDS
3'-UTR
nominal thresh.
log10(1/pval)
cumulated Grimson score (stacked)
location in DLK1 mRNA (C2 form; nt)
8mer
7mer-m8
7mer-A1
6mer

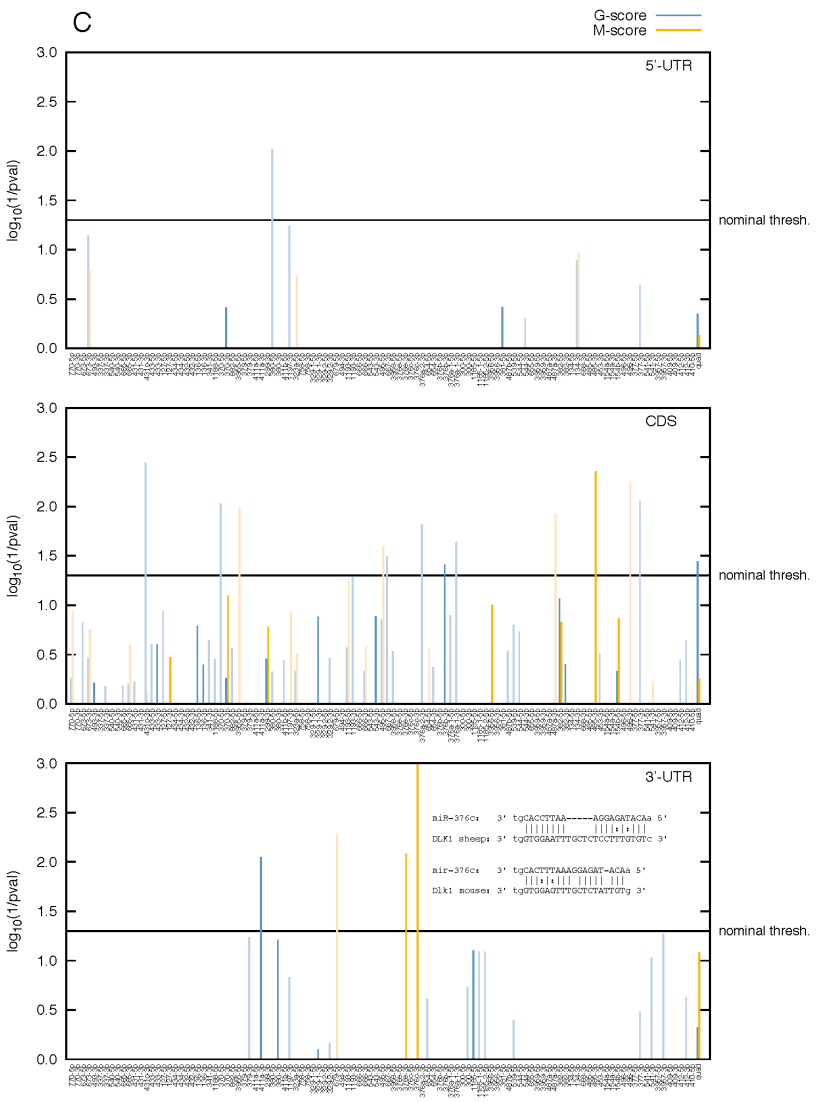


Figure 4: (A) Statistical significance (log(1/p)) of the affinity of ovine miRNAs in the *DLK1-GTL2* domain for the 5'UTR, coding sequence (ORF) and 3'UTR of the ovine *DLK1*. The affinity was measured using either G- (blue) or M-scores (orange) as defined in the text. Bars are dark colored for highly-expressed and light colored for lowly-expressed miRNAs. The last pair of bars ("quad") at the right of the graph corresponds to the quadrille scores, the remaining bars to the species-scores and are labeled accordingly. p-values were determined using the sequence shuffling test described in the text. Species-scores require a Bonferroni correction for 127 independent tests. (B) Position in the DLK1 mRNA of target sites (8-mers, 7-mers and 6-mers as defined by Grimson et al. (2007)) for the same set of miRNA species. (C) Same as in (A) except that the scores are "multi-organism (MO) scores" combining information from sheep, human and mouse.

When adding the human and murine information, the significance of the MO G-quadrille score for the ORF increased slightly (p=0.036), although none of the individual miRNAs clearly stood out (Fig. 4C). When applied to the 3'UTR, the MO M-species-score for miR-376c-3p achieved Bonferroni-corrected significance (nominal p = 0.0004; Bonferroni-

corrected p = 0.05). This signal was due to a miRanda target site shared between mouse and sheep (Fig. 4C).

Weighting miRNA scores by expression level estimated from HTS reads and including all isomirs in such analyses did not yield stronger signals (data not shown).

Discussion

We herein establish a catalogue of miRNAs expressed in skeletal muscle of sheep. Using miRDeep (Friedlander et al. 2008), we detected 747 small RNA species mapping to 472 miRNA precursors. Three hundred twenty four of these were classified as orthologues or close paralogues of known miRNAs, leaving 148 candidate novel miRNAs. It is noteworthy that expression levels of new miRNA were considerably lower than those of known miRNAs (Suppl. Fig. 16).

As the ovine genome sequence is not completed, we used the bovine as reference for most of the genome. Comparison of the miRDeep performances on 390 Kb of contiguous sequence available in both bovine and sheep, indicates a possible loss of ~6% sensitivity but not of specificity. Reanalyzing the sequence data with the ovine reference may thus increase the number of detected miRNAs. Within the *DLK1-GTL2* domain, six known miRNA precursors were missed by miRDeep despite the occurrence of HTS reads. Assuming that the *DLK1-GTL2* miRNA catalogue of sheep is near complete, this corresponds to a sensitivity of 49/55 = 0.89. This figure is identical to that obtained by the miRDeep developers in *C. elegans* (Friedlander et al. 2008).

When focusing on the annotation of miRNAs in the *DLK1-GTL2* domain, we noted ~500 reads mapping to 12 predicted C/D snoRNA genes in *MEG8*. This is reminiscent of Ender et al. (2008) who reported human AGO1-4 associated small RNAs mapping to C/D and H/ACA snoRNA precursors. More specifically, Ender et al. (2008) demonstrated Drosha-independent (also known as RNASEN), Dicer-dependent (also known as DICER1) processing of miRNAs derived from the *bona fide* ACA45 snoRNA (also known as SCARNA15), thereby revealing an alternative pathway for the generation of functional miRNAs. It is not yet known which of the C/D snoRNAs in *MEG8* (Cavaille et al. 2002) are genuine, associating

with core components of the C/D RNP (e.g. FBL, NHP2L1, NOP56, NOP58). Contrary to Ender et al. (2008), in four of six cases in which reads were derived from both arms, the two miRNA species were characterized by 2-nt 3' overhangs compatible with Drosha-dependent processing of the pri-miRNA (Suppl. Fig. 17). Further work is needed to exclude the trivial possibility that the corresponding miRNA derive from erroneously annotated C/D snoRNAs.

We confirm that miRNAs from the *DLK1-GTL2* domain are imprinted in skeletal muscle of sheep and preferentially expressed from the maternal allele. As for the other genes in the domain, the imprinting status of the miRNAs is not affected by *CLPG* genotype. Of note, the imprinting status of the genes in the *DLK1-GTL2* domain cannot be directly tested in *CLPG/CLPG* animals as the padumnal and madumnal alleles are identical. However, we have demonstrated that the IG-DMR is differentially methylated in *CLPG/CLPG* animals as in other genotypes, supporting regular imprinting (data not shown). Previous studies in skeletal muscle of sheep revealed tight imprinting control for both paternally and maternally expressed genes. For one of the miRNAs (miR-485) we observed relaxation of imprinting in two (one $CLPG^{Mat}/+^{Pat}$ and one $+^{Mat}/CLPG^{Pat}$) out of three studied individuals: molecules derived from the paternal allele represented ~15% of the total. Flanking miRNAs (located 600bp upstream (miR-134) and 4,000bp downstream (miR-453)) did not show evidence of relaxation in the same samples. This strongly suggests that miR-485 is at least in part processed from a transcription unit independent of the one generating the two other miRNAs. Rather than being one unique large pri-miRNA, *MIRG* may thus encompass multiple transcription units controlled by distinct promoters. Along the same lines, it has recently been suggested that miR-433 and the adjacent miR-127 are processed from distinct pri-miRNAs (Song and Wang 2008) rather than from a unique *anti-PEG11* precursor shared with miR-431, miR-434 and miR-136 (Davis et al. 2005).

We find that the miRNAs from the *DLK1-GTL2* domain are affected by the *CLPG* mutation in the same manner as the maternally expressed long non-coding RNA genes (Charlier et al. 2001a). The main effect is in *cis* causing a ~3.2-fold increase in miRNA expression from a *CLPG* versus wild-type maternal chromosome ($[CLPG^{Mat}/+^{Pat}]/[+/+]\approx 3,7$; $[CLPG/CLPG]/[+^{Mat}/CLPG^{Pat}]\approx 2,7$). As for the other genes affected by the *CLPG* mutation, this is thought to result from the inactivation of a muscle-specific cis-acting silencer element. In

addition, we confirm a trans-effect consisting in the ~1.8-fold higher expression of miRNAs from the maternal allele when the paternal allele is *CLPG* versus wild-type (([$+^{Mat}/CLPG^{Pat}$]/[+/+]≈2.1; [*CLPG/CLPG*]/[$CLPG^{Mat}/+^{Pat}$]≈1.5). While the MAT->PAT trans-effect is known (*PEG11*; (Davis et al. 2005)) or hypothesized (*DLK1;* (Georges et al. 2003; Georges et al. 2004)) to reflect miRNA-mediated trans-inhibition, the molecular mechanisms underlying this PAT->MAT trans-effect remain elusive. One possible explanation is that the silencer element that is inactivated by the *CLPG* mutation has the capacity to exert its effect in trans on the homologous chromosome. Such mechanism has been attributed to enhancers in Drosophila and other organisms and may underlie transvection (Kennison & Southworth, 2002).

The *cis*-effect on miRNA expression is consistent with that observed for the long non coding RNA genes but is considerably weaker (~3.2-fold versus ~9.5-fold). The reason for this difference is unknown. An explanation might be the saturation of the miRNA processing machinery.

Although HTS-based digital expression profiling may not be as quantitative as initially assumed, the combined HTS, array and QRT-PCR data strongly suggest that miRNA expression levels differ at least ~30-fold. This could be due to differential processing efficiency of the precursors and/or stability of the processed miRNAs, but may also reflect the dependence on distinct promoters of unequal strength. The latter hypothesis is supported by the miR-485 imprinting data (see above), as well as by the recent identification of private host-gene independent promoters for intronic miRNAs (Ozsolak et al. 2008).

Contrary to skeletal muscle of mouse, we observed substantial levels of A to I editing for 4/6 pri-miRNAs from the miR-376 family (miR-376e, miR-376c, miR-376a2, miR-376b) and for three unrelated pri-miRNAs (miR-381, miR-411a, miR-369) from the *DLK1-GTL2* domain in skeletal muscle of sheep. We noted a significant effect of *CLPG* genotype on pri-miRNA editing. Editing in *+/+* animals, characterized by the lowest miRNA expression levels, was ~4.2-fold higher than in the other genotypes. The reasons underlying this observation remain unclear, but could involve saturation of the editing machinery, or downregulation of components of the editing machinery by miRNAs from the domain. Contrary to Kawahara et

al. 2007, editing levels were lower in the mature miRNA population when compared to precursors: edited molecules never made up > 10% of reads.

The primary aim of this study was to lay the grounds for the identification of miRNAs that might account for the translational downregulation of *DLK1* observed in *CLPG/CLPG* animals, i.e. the MAT->PAT trans-effect. We approached this by posing that if an exceptional affinity of miRNAs from the *DLK1-GTL2* domain for *DLK1* could be clearly demonstrated *in silico*, this would very strongly support the hypothesis. The absence of such statistically significant affinity does not preclude actual interaction in vivo, but is neutral with respect to the hypothesis. We would not pretend that our data reveal an unambiguous affinity of the *DLK1-GTL2* miRNAs for *DLK1,* yet it is intriguing that the G- quadrille score for the ORF, and the miR-376c-3p MO M-species-score for the 3'UTR, both achieved 5% significance. This suggests that the *DLK1-GTL2* miRNAs might indeed effectively downregulate *DLK1* in *CLPG/CLPG* animals. It is worthwhile restating in this regard that miRNAs from the domain account for an estimated ~20% of cellular miRNAs in these animals. We are in the process of testing this prediction biochemically using both reporter assays and AGO-based (also known as EIF2C) target coimmunoprecipitation (e.g. Takeda et al., 2010), prioritizing miRNAs on the basis of the results presented in Fig. 4.

While miRNAs from the domain are strong candidate direct mediators of the MAT->PAT trans-effect on *DLK1,* alternative possibilities should not be excluded. Amongst those figure (i) an indirect effect of the miRNAs from the domain, as well as (ii) miRNA-independent mechanisms. It is interesting with regards to the latter that no clear function has yet been assigned to *GTL2*.

Individual miRNAs are predicted to target 200 to 300 genes on average (Grimson et al., 2007). What are the targets of the highly conserved miRNAs in the *DLK1-GTL2* domain? To start addressing this question, we assembled the list of 2,832 bovine genes having at least one conserved 8- or 7-mer target site in their 3'-UTR for any of the 24 out of 153 "conserved miRNA families" with representatives in the *DLK1-GTL2* domain (Friedman et al. 2009). We then looked for the enrichment of specific gene ontology (GO) terms amongst these genes **(**Ashburner et al**.** 2000**)**. To that end we randomly sampled 10,000 (GO Slim analysis) or 200,000 (whole GO analysis) sets of 2,832 genes from the complete TargetScan list of 8,458

bovine miRNA-targeted genes, and compared the hit number of each GO term for the list of target genes of the *DLK1-GTL2* miRNAs, with the distribution of hit numbers across the 10,000 (respectively 200,000) random sets of genes. The resulting p-values were Bonferroni corrected for the 6,930 terms sampled out of the whole GO graph, or for the 55 terms of the GO Slim graph. Suppl. Table 2A shows the eight most enriched terms in both analyses, corresponding to a Bonferroni corrected p-value ≤ 0.027 for the GO slim analysis, or a nominal p-value ≤ 10^{-4} for the GO whole analysis. The outcome of this analysis strongly suggests that the miRNAs from the *DLK1-GTL2* domain are devoted to the targeting of regulators of the gene circuitry operating at the transcriptional, translational and post-translational level, primarily in the nervous system. The list of genes corresponding to these top hits is provided in Suppl. Table 2B&C.

Materials & Methods

Construction of small RNA libraries and high-throughput sequencing.

Small RNA libraries were constructed using the "Small RNA sample preparation kit" following the instructions of the manufacturer (Illumina). Briefly, 10 µg of total RNA extracted with Trizol (Invitrogen) was size fractionated by denaturing polyacrylamide gel electrophoresis (PAGE; 15%) and molecules ranging from 18 to 30 nucleotides eluted. RNA adapters were successively ligated to the 5' then 3' end of the isolated small RNAs, and ligation products of the desired length (40-60bp then 70-90bp) recovered by sequential PAGE (15% then 10%). Small RNAs appended with 5' and 3' adapters were reverse transcribed with Superscript II (Invitrogen) and amplified with Phusion DNA polymerase (Finnzymes Oy). Resulting amplicons were PAGE (6%) gel purified, hybridized on a flow cell lane, clustered and sequenced (36 cycles) using standard procedures (Illumina). Libraries corresponding to eight animals (two of each CLPG genotype) were first sequenced on a GA-I (Illumina) by Fasteris SA (Geneva, Switzerland). The experiment was subsequently repeated for seven animals using a GA-II instrument (Illumina) at the GIGA-R core facilities.

Bioinformatic analysis of small RNA reads.

The bioinformatic procedures applied for preprocessing of HTS reads, prediction, curation and annotation of miRNA precursors, prediction of C/D snoRNAs in the *DLK1-GTL2*

domain, gene annotation and conservation analyses in the domain, quantitative analyses of HTS reads, comparison of HTS, Exiqon and Taqman data, analysis of non-miRNA HTS reads, evaluation of miRNA affinity for *DLK1*, and GO analyses of targets of miRNAs encoded in the domain are described in details in Supplemental Methods.

Exiqon array hybridization.

Skeletal muscle RNA samples from the same eight animals (two of each CLPG genotype) extracted with Trizol (invitrogen) were hybridized on Exiqon miRCURY™ LNA Arrays (v.9.2) at Exiqon (Vedbaek, Denmark). Briefly, RNA quality was evaluated on an Agilent Bioanalyzer 2100. Individual samples were labelled with Hy3 using the miRCURY™ Hy3™/Hy5™ power labelling kit, and cohybridized on the arrays with a Hy5-labelled equimolar mix of the eight samples. Arrays were scanned in an ozone free environment. Fluorescence intensities were normalized with the global Lowess (LOcally WEighted Scatterplot Smoothing) regression algorithm using all probes except those corresponding to miRNAs from the DLK1-GTL2 domain.

Quantitative RT-PCR.

QRT-PCR analyses of miRNAs were conducted using predesigned (miR-1, miR-127, miR-206, miR-382, miR-382*, let-7d) or custom (miR-3958 and miR-3959) Taqman®MicroRNA assays (ABI) on an 9700HT (ABI) instrument. Assay-specific amplification efficiencies were determined using serial RNA dilutions.

Editing.

The level of pri-miRNA editing was determined by sequence analysis of genomic- and cDNA- derived PCR products. Genomic DNA and total RNA were extracted using Trizol (invitrogen). Reverse transcription was carried out using Supercript III (Invitrogen) on 1µg of total RNA pre-treated with Turbo DNase (Ambion) and cDNA PCR amplified using GOLD Taq (ABI)(35 cycles). Primer sequences used for mouse and sheep are provided in Supplementary Table 3. Amplicons were sequenced on a 3730 instrument (ABI) and the degree of editing estimated from the electropherograms using PeakPicker (Ge et al. 2005).

Acknowledgments

This work was funded by grants from (i) the Fonds National de la Recherche Scientifique, (ii) the University of Liège, (iii) the European FW6 program (CALLIMIR), (iv) the Communauté Française de Belgique (ARC Mirage and ARC Biomod), and (v) the Belgian Science Policy Organisation (SSTC Genefunc PAI). Carole Charlier is Chercheur Qualifié au Fonds National de la Recherche Scientifique. We are grateful for the support of the GIGA-R sequencing core facility.

References

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25(1): 25-29.

Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. 2008. The impact of microRNAs on protein output. Nature 455(7209): 64-71.

Byrne K, Colgrave ML, Vuocolo T, Pearson R, Bidwell CA, Cockett NE, Lynn DJ, Fleming- Waddell JN, Tellam RL. 2010. The imprinted retrotransposon-like gene PEG11 (RTL1) is expressed as a full-length protein in skeletal muscle from Callipyge sheep. PLoS One 5(1):e8638.

Cavaille J, Seitz H, Paulsen M, Ferguson-Smith AC, Bachellerie JP. 2002. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. Hum Mol Genet 11(13): 1527-1538.

Charlier C, Segers K, Karim L, Shay T, Gyapay G, Cockett N, Georges M. 2001a. The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status. Nat Genet 27(4): 367-369.

Charlier C, Segers K, Wagenaar D, Karim L, Berghmans S, Jaillon O, Shay T, Weissenbach J, Cockett N, Gyapay G et al. 2001b. 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. Genome Res 11(5): 850-862.

Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR et al. 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. Nucleic Acids Res 33(20): e179.

Chi SW, Zang JB, Mele A, Darnell RB. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. Nature 460(7254): 479-486.

Cockett NE, Jackson SP, Shay TL, Farnir F, Berghmans S, Snowder GD, Nielsen DM, Georges M. 1996. Polar overdominance at the ovine callipyge locus. Science 273(5272): 236- 238.

Davis E, Caiment F, Tordoir X, Cavaille J, Ferguson-Smith A, Cockett N, Georges M, Charlier C. 2005.RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. Curr Biol 15(8): 743-749.

Davis E, Jensen CH, Schroder HD, Farnir F, Shay-Hadfield T, Kliem A, Cockett N, Georges M, Charlier C. 2004. Ectopic expression of DLK1 protein in skeletal muscle of padumnal heterozygotes causes the callipyge phenotype. Curr Biol 14(20): 1858-1862.

Durbin R, Eddy SR, Krogh A, Mitchison G. 1998. Biological sequence analysis: Probabilistic models of proteins and nucleic acids. Cambridge Univ Pr.

Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G. 2008. A human snoRNA with microRNA-like functions. Mol Cell 32(4): 519-528.

Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 30(7): 1575-1584.

Freking BA, Murphy SK, Wylie AA, Rhodes SJ, Keele JW, Leymaster KA, Jirtle RL, Smith TP. 2002. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. Genome Res 12(10): 1496-1506.

Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N. 2008.Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol 26(4): 407-415.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res 19(1): 92-105.

Ge B, Gurd S, Gaudin T, Dore C, Lepage P, Harmsen E, Hudson TJ, Pastinen T. 2005. Survey of allelic expression using EST mining. Genome Res 15(11): 1584-1591.

Georges M, Charlier C, Cockett N. 2003. The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. Trends Genet 19(5): 248-252.

Georges M, Charlier C, Smit M, Davis E, Shay T, Tordoir X, Takeda H, Caiment F, Cockett N. 2004. Toward molecular understanding of polar overdominance at the ovine callipyge locus. Cold Spring Harb Symp Quant Biol 69: 477-483.

Glazov EA, Kongsuwan K, Assavalapsakul W, Horwood PF, Mitter N, Mahony TJ. 2009. Repertoire of bovine miRNA and miRNA-like small regulatory RNAs expressed upon viral infection. PLoS One 4(7): e6349.

Griffiths-Jones S. 2006. miRBase: the microRNA sequence database. Methods Mol Biol 342: 129-138.

Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol Cell 27(1): 91-105.

John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. 2004. Human MicroRNA targets. PLoS Biol 2(11): e363.

Kagami M, Yamazawa K, Matsubara K, Matsuo N, Ogata T. 2008. Placentomegaly in paternal uniparental disomy for human chromosome 14. Placenta 29(8): 760-761.

Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. 2007. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. Science 315(5815): 1137-1140.

Kennison JA & Southworth JW 2002. Transvection in Drosophila. in Homology Effects. Advances in Genetics 46: 399-420.

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. Cell 129(7): 1401-1414.

Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: short oligonucleotide alignment program. Bioinformatics 24(5): 713-714.

Lin SP, Youngson N, Takada S, Seitz H, Reik W, Paulsen M, Cavaille J, Ferguson-Smith AC. 2003. Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on use chromosome 12. *Nat Genet* **35**(1): 97-102.

Linsen SE, de Wit E, Janssens G, Heater S, Chapman L, Parkin RK, Fritz B, Wyman SK, de Bruijn E, Voest EE et al. 2009. Limitations and possibilities of small RNA digital gene expression profiling. Nat Methods 6(7): 474-476.

Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M et al. 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res 18(4): 610-621.

Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG, Zhang X, Song JS, Fisher DE. 2008. Chromatin structure analyses identify miRNA promoters. Genes Dev 22(22): 3172-3183.

Seitz H, Royo H, Bortolin ML, Lin SP, Ferguson-Smith AC, Cavaille J. 2004. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. Genome Res 14(9): 1741-1748.

Seitz H, Youngson N, Lin SP, Dalbert S, Paulsen M, Bachellerie JP, Ferguson-Smith AC, Cavaille J.2003. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. Nat Genet 34(3): 261-262.

Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. 2008. Widespread changes in protein synthesis induced by microRNAs. Nature 455(7209): 58-63.

Smit M, Segers K, Carrascosa LG, Shay T, Baraldi F, Gyapay G, Snowder G, Georges M, Cockett N, Charlier C. 2003. Mosaicism of Solid Gold supports the causality of a noncoding A-to-G transition in the determinism of the callipyge phenotype. Genetics 163(1): 453-456.

Song G, Wang L. 2008. MiR-433 and miR-127 arise from independent overlapping primary transcripts encoded by the miR-433-127 locus. PLoS One 3(10): e3574.

Takeda H, Charlier C, Farnir F, Georges M. 2010. Demonstrating polymoprhic miRNA_mediated gene regulation in vivo: application to the g+6223G-A mutation of Texel sheep. RNA, in the press.

Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. 2008. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. Nature 455(7216): 1124-1128.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol 3(7): RESEARCH0034.

Vuocolo T, Byrne K, White J, McWilliam S, Reverter A, Cockett NE, Tellam RL. 2007. Identification of a gene network contributing to hypertrophy in callipyge skeletal muscle. Physiol Genomics 28(3): 253-272.

Supplemental Data

Supplemental Methods

Preprocessing of HTS reads

As the two first of the eight libraries sequenced by Fasteris SA (Geneva) had reads limited to 35 nt in length, all HTS reads were truncated to 35 nt prior to analysis. To speed up subsequent computations, HTS reads from each library were first pooled by sequence and absolute counts encoded in sequence identifiers. Notably, quality values were discarded.

To remove sequencing adapters, each unique sequence was globally aligned with the Gex Adapter 2 (5'-TCGTATGCCGTCTTCTGCTTG; Illumina) using NEEDLE (EMBOSS package 6.1.0; Rice et al. 2000) with gap penalties of 10.0 and 0.5 (opening and extension, respectively). NEEDLE output was then automatically processed to trim sequences with a 3'-aligned region ≥ 10 nt having at most 3 edits (mismatch or gap) relatively to the adapter. These somewhat liberal thresholds were selected to maximize the number of usable reads.

After a repooling step carried out for each individual library, a data set combining all trimmed reads from the eight libraries was assembled (~240,000 unique sequences representing ~48,000,000 reads). This merged data set was then used to predict miRNA and snoRNA precursors, thus ensuring that all animals shared the same list of predictions.

Prediction of miRNA precursors

Ovine miRNA precursors were predicted genome-wide using core algorithms of MIRDEEP (Friedlander et al. 2008). For the *DLK1-GTL2* domain, we had our own ovine sequence (Charlier et al. 2001), whereas the cow genome (bosTau4 assembly; Liu et al. 2009) was used as a proxy for the remaining of the genome (due to the incompleteness of the sheep genome).

In practice, unique sequences were first mapped on the ovine domain or on the unmasked cow contigs using SOAP.CONTIG 1.11 (Li et al. 2008). The following parameters were selected as a reasonable compromise between speed and sensitivity considering that genome-wide predictions would be based on heterologous mapping: seed size (8 nt),

maximum number of mismatches (2), maximum gap size (2 nt), maximum number of equal best hits (10 and all reported).

SOAP output was then filtered to discard sequences mapping to more than five genomic locations (as suggested in miRDeep protocol) and re-formatted to match the requirements of the excise-candidates component of miRDeep pipeline. For the *DLK1-GTL2* domain, this step resulted in the excision of ~1,000 candidate precursors, while ~180,000 candidates were excised from bovine contigs.

Following miRDeep protocol, candidate precursors were submitted to RNAfold 1.6.5 (McCaskill 1990) using default parameters. In parallel, SOAP.short was used to re-map unique sequences on candidate precursors with the same parameters as above except that the maximum number of equal best hits was raised to 10,000. RNAfold structures and SOAP output re-formatted as a 'signature' file were finally input to miRDeep core component, along with a list of metazoan miRNAs extracted from miRBase 12 (Griffiths-Jones 2006) for scoring conserved seeds. miRDeep was configured to make use of RandFold 2.0 (Bonnet et al. 2004) and the score cut-off was kept to the default value (1).

Curation of miRNA precursors

For the ovine *DLK1-GTL2* domain, 49 precursors were predicted by miRDeep, of which 9 were unknown (miR-154b, miR-323b, miR-323c, miR-411b and miR-3955 to 3959) and one absent from miRBase (miR-376e) yet described in (Seitz et al. 2004). Moreover, careful examination revealed that six precursors had been missed by miRDeep in spite of corresponding reads in our HTS libraries (see below for detailed protocol). In two cases (miR-494 and 665), precursors were not even excised because they would have been > 140 nt, which is the internal threshold of the excise-candidates component. Three other candidates (miR-431, 433, 1197) were rejected because they failed to fold in a canonical stem-loop, while the last one (miR-412) had a score just below the cut-off. Finally, manual analysis of a MLAGAN (Brudno et al. 2003) multiple alignment of the domain showed that six more ovine precursors were conserved at ≥ 70% with their human orthologues (with no deletion in the mature sequence), though no read could be associated with these (miR-300, 337, 453, 654,

770, 889). Hence, the curated catalogue for the *DLK1-GTL2* domain comprises 61 miRNAs (49 predicted by miRDeep and 12 unpredicted yet very likely real precursors).

Of the 540 bovine precursors retained by miRDeep, 71 were discarded as false positives resulting from widespread mapping of mutated reads related to highly expressed miRNAs (miR-1, 206, 378, and the let-7 family). Specifically, taking advantage of the BLASTN tool of the miRBase 12 web interface, we filtered out precursors for which the first hit to the predicted mature sequence was any of these mature miRNAs with a bitscore < 35 bits. Using standalone NCBI BLASTN 2.2.22 (Altschul et al. 1997) to map the remaining bovine precursors on the ovine *DLK1-GTL2* domain, we further removed 46 precursors that were perfectly syntenic with their ovine orthologues. This let us with 423 bovine precursors identified by heterologous mapping of our ovine HTS reads. Even if miR-1 mature dominates our libraries (see main text), we noticed that only one (on BTA13) of its two known loci is actually predicted by miRDeep. Again, this failure is due to the candidate precursor on BTA24 being > 140 nt.

To assemble a genome-wide miRDeep catalogue of ovine miRNAs, the reduced bovine catalogue was combined to the raw ovine catalogue for the *DLK1-GTL2* domain (without the 12 unpredicted precursors), which resulted in a list of 472 precursors. The latter was used for all statistical analyses pertaining to miRNA expression, whereas the curated ovine catalogue limited to the domain (61 precursors) was used in Fig. 1 and to study miRNA affinity for *DLK1*.

Annotation of miRNA precursors

To annotate our catalogue, we compared the 472 ovine precursors to the 615 bovine precursors present in miRBase 14. As the largest part of our catalogue had been predicted by heterologous mapping on the cow genome, orthologues of known bovine miRNAs were straightforward to identify by comparison of genomic locations. For all these precursors, we nevertheless checked that the ovine mature sequence (deduced from HTS reads) had at most two mismatches with its bovine counterpart. The remaining precursors were subdivided into three categories according to the outcome of two successive standalone BLASTN analyses against bovine then non-ruminant miRNA species present in miRBase 14

(using a word length of 12 nt). If any of the ovine "species" (i.e., mature, star, 5p, 3p; see below) aligned to the first hit with a bitscore > 32 bits, the corresponding ovine precursor was annotated either as a paralogue of a known bovine miRNA or as an orthologue of a non-ruminant miRNA, depending on the genomic source of the hit. Otherwise, the ovine precursor was considered unknown. In the *DLK1-GTL2* domain, manual annotation was preferred to automated annotation for miR-376e and miR-411b.

Clustering of the ovine precursors into families based on sequence similarity was carried out with the MCL algorithm (Enright et al. 2002). In a first step, all precursors were compared to each other with standalone BLASTN, again using a word length of 12 nt for maximum sensitivity. Then, the MCLBLASTLINE component of the MCL package was used to build and partition a graph from the pairwise bitscores. To limit false homologies, a bitscore cut-off (> 35 bits) was set after visual inspection and mining of the BLASTN report. The unique inflation parameter *I* was set to 2.0 as described in the main text. The 484 precursors (including the 12 unpredicted precursors) clustered into 287 families, of which 224 were singletons. Note that family #61 involving one unpredicted precursor from the ovine *DLK1-GTL2* domain is not mentioned in the main text but shown in Fig. 1.

Prediction of C/D snoRNAs in the *DLK1-GTL2* domain

Ovine and bovine snoRNAs in the *DLK1-GTL2* domain were predicted with HMMER 2.3.2 (Durbin et al. 1998) using HMM profiles built from human snoRNAs. Human sequences for 14qI and 14qII families were downloaded from snoRNABase 3 (Lestrade and Weber 2006) and separately aligned with CLUSTALW 1.83 (Thompson et al. 1994). The HMMBUILD component of HMMER was then used to compute the two profiles and HMMSEARCH to scan the two ruminant loci. Domain-specific E-value cut-offs for 14qI and 14qII profiles were set to 1e-10 and 1e-04, respectively. We selected these values empirically as those yielding the closest set of snoRNAs in ruminants (10 x 14qI and 35/36 x 14qII) in comparison to the human domain (9 x 14qI and 31 x 14qII). As some snoRNAs were predicted by both profiles, the boundary between the two families that was eventually drawn might be slightly off. The lone 14q0 was localized in the bovine domain by mapping the known ovine sequence with standalone BLASTN. Finally, using the MLAGAN multiple alignment of the domain, we

manually identified orthologues among ovine (including an additional copy of 14qII missed by HMMER), bovine, human and murine snoRNAs based on synteny and sequence similarity.

Gene annotation and conservation analyses in the *DLK1-GTL2* domain

To determine exon/intron boundaries for protein-coding and start/end positions for long non-coding transcripts in the ovine domain, we combined experimental evidence obtained in sheep with human, murine and bovine transcripts annotated in public databases. Briefly, non-ovine unspliced transcripts were downloaded from Ensembl 57 using BioMart (Haider et al. 2009) and the Convert tool of the UCSC genome browser to target the region orthologous to the ovine domain in hg19 (chr14:101,136,416-101,545,413), mm9 (chr12:110,645,537-110,991,765) and bosTau4 (chr21:65,669,039- 66,065,774). All transcripts were then jointly mapped on our ovine contigs using standalone BLASTN with an E-value threshold of 0.001. Finally, individual hits with an identity percentage > 80% that belonged to the same annotated gene were merged *in silico* to deduce crude ovine coordinates.

Sequence conservation for *MIRG* and *MEG8* was computed with plotcon (EMBOSS package 6.1.0) using a window size of 8 nt. Aligned blocks spanning the whole set of human miRNAs (miRBase 14) or snoRNAs (snoRNABase 3), respectively for *MIRG* and *MEG8*, were assembled using the Galaxy portal (Giardine et al. 2005) based on the 46-way multiZ alignment (UCSC genome browser). Aside from hg19, mm9 and bosTau4, blocks also contained sequences from panTro2, ponAbe2, rheMac2, rn4, cavPor3, equCab2, and canFam2 (all built with the Syntenic Net method). To the exception of miR-1193 that was manually added (as it is missing from miRBase), these analyses included only short RNA species described in human, yet generally conserved across mammals.

Quantitative analyses of HTS reads

To estimate miRNA expression from HTS data, unique sequences from each individual library were first re-mapped on the 472 precursors of the curated catalogue with SOAP.short (see above for parameter values). Then, SOAP output was parsed to filter out

reads with indels and those mapping on the reverse strand (relatively to the precursor sequence). Further, when a unique sequence had more than one best hit (with at most two mismatches), its absolute count (as encoded in the identifier) was uniformly distributed among equal best hits. Two main outputs were generated from the automated parsing of this controlled mapping step: (i) empirical base counts at each position of each precursor and (ii) a quantitative inventory of all unique molecules (i.e., with a unique combination of start and end positions) spawned by each precursor. In both cases, HTS libraries were also parsed separately so that comparisons between genotypes (or individual animals) remained possible in addition to merged analyses. Empirical base counts were used to deduce genome-wide ovine sequences from HTS reads (see above) as well as to examine imprinting and editing, whereas the inventory was processed as follows.

Unique molecules from each precursor were first sorted by descending count and by location on the precursor (to break ties). Starting with the most abundant, molecules were then assigned to growing "stacks" based on their location. Stack height and width enlarged as required to accommodate new molecules except when this would result in merging two neighbouring "stacks". Instead, those rare molecules overlapping two stacks were discarded. At the end of this process, each precursor had from one to three non-overlapping stacks of molecules of defined height and width. Stacks were classified in "species" according to their relative height: if the 2nd highest was >= 0.15 as abundant as the highest, both were qualified as '5p' and '3p', depending on their location; otherwise, the highest was qualified as 'mature' and the 2nd highest (if present) as 'star'. The few additional stacks were considered minor (e.g., loop) and discarded. Though this standard nomenclature (Landgraf et al. 2007) is used in Suppl. Table S1, it had no practical influence on statistical analyses, with all species from each precursor equally considered in computations (see below). Within each stack, molecules were ranked according to their abundance. Except for isomir statistics reported in the main text, the most abundant molecule was selected to represent its whole species (in terms of both sequence and start/end positions), while all molecules contributed to the height of the stack corresponding to the species.

To assemble input tables for statistical analyses across libraries, stack heights for each precursor (at most two) were combined using species from the merged data set as a

guide. Specifically, a stack from a particular library was included only if start/end positions of its most abundant molecule fell within 3 nt of those of the merged data set for the considered precursor. This approach allowed us to ignore species types during combination while discarding on a case-by-case basis libraries in which a given species was not comparable with those from other libraries. Resulting tables were normalised and further analysed with the R statistical software package (R Development Core Team, Vienna, Austria) as described in the main text. For the analyses restricted to the *DLK1-GTL2* domain, sub-tables were generated by filtering complete tables on precursor locations. Similar tables were assembled for snoRNAs by separately mapping HTS reads to the 47 "precursor" sequences identified in the ovine *DLK1-GTL2* domain.

Analyses of additional HTS libraries

Seven new HTS libraries were sequenced and analysed in house using the Illumina Sequencing Analysis Software 1.3. Contrary to the first batch of libraries, quality values were taken into account to filter reads using default parameters of the pipeline (i.e., those with a "chastity" less than 0.6 on two or more bases among the first 25 bases were discarded). After filtering, the average number of reads per library was very close to that found in the original experiment (6,452,524 vs. 6,324,668) yet with a much wider range (2,309,996—8,494,025 vs. 5,222,920—6,685,342). New reads were trimmed to 35 nt and their adapter removed before further processing. The corresponding merged data set had ~430,000 unique sequences representing ~43,500,000 reads (for seven libraries).

To allow comparison with the eight libraries already analysed, new sequences were simply re-mapped on the previously established miRNA/snoRNA catalogues and the output parsed as above. Since statistical analyses actually required the use of the 15 HTS libraries, input tables from the first and second batch had to be combined in some way. This was carried out through a 'full join' between the two input tables where miRNA species rows from a given precursor were combined whenever start/end positions of the most abundant molecule of their respective merged data sets fell within 3 nt of each other. Even if reasonable, this strategy had the somewhat undesirable effect of generating two complementary 'star' rows for 28 precursors with unstable 'star' species. Given that 51 more

miRNA species were only found in the second batch (37 in the other way around), this explains the difference between the numbers of species mentioned in the main text when respectively considering the first batch alone (747) or the two batches at the same time (747+28+51=826).

Comparison of HTS, Exiqon and Taqman data

As Exiqon probes had been designed for human miRNAs, it was crucial to identify those that could hybridize to ovine miRNAs to avoid analysing false negatives. To this end, we only considered Exiqon probes for which the target human mature sequence (as provided by Exiqon) was found unaltered in our ruminant precursors and at a location within 3 nt of the start position of the corresponding miRNA species (averaged across HTS1 and HTS2 merged data sets). In an attempt to limit many-to-many relationships in combined tables, we further discarded probes that were reported by Exiqon as targeting more than a single miRNA. This led to a combined table of 265 miRNA species. The eight Taqman probes were selected to ensure proper hybridization to ovine miRNAs and manually assigned to their HTS and/or Exiqon counterparts.

For the analyses of relative expression of domain-encoded miRNAs broken by genotype, HTS and Exiqon data were combined as above. However, for four Exiqon probes rejected by our stringent combination strategy (miR-377-3p, miR-412-3p, miR-654-3p, miR-655-3p), we nevertheless observed the expected genotype effect, which indicates that these probes specifically hybridize to their targets. Therefore, they were manually added to the corresponding data sets.

Analyses of non-miRNA HTS reads

In contrast to other studies dealing with high throughput sequencing of short RNAs (e.g., Rathjen et al. 2009), an overwhelming fraction (~90—93%) of our HTS reads could be affiliated to one or more miRNA precursors. The remaining reads consisted either in 35-nt sequences lacking a recognizable sequencing adapter or in shorter (apparently non-miRNA) sequences followed by a regular adapter. To ensure that we did not miss any interesting

short RNA species, both classes of "orphan" reads were pooled (within each batch of HTS libraries) and sequentially analyzed as summarized in Suppl. Fig. 18.

First, single base tracts were filtered out using a very simple heuristic: HTS reads (without adapter when recognizable) having >= 60% N's or >= 90% of either A, C, G, T (excluding N's) were tagged as poly-N (or A, C, G, T) and dropped from further analyses. For simplicity, these are reported as 'poly-A' since the latter amounts for 86—98% of the HTS reads identified with this strategy. Next, we assembled a custom database compiling the two sequencing adapters, all miRBase 14 precursors, human snoRNAs from snoRNABase 3, as well as the non-redundant content of Repbase 14.10 (Jurka et al. 2005). Using standalone BLASTN with an E-value threshold of 0.01, the remaining reads were then annotated by sequence similarity to the first database hit (when available). As this BLAST analysis was deliberately less stringent than the original SOAP mapping, it is not surprising to retrieve 41-49% of orphan reads corresponding to known miRNAs. In many cases, these new "miRNA" reads were either too short (< 19 nt) or too mutated (> 2 mismatches; indels) to be effectively mapped by SOAP, while a third class consisted of reads mapping to real miRNA precursors missed by miRDeep. Among reads matching Repbase entries, we identified fragments of tRNAs, of both small- and large-subunit rRNAs, and of various other repeated sequences. Finally, the reads still to be annotated were likewise compared to a database of bovine cDNAs (built from Ensembl 56 using BioMart). Beside the expected fragments of highly-expressed protein genes, this third analysis yielded a few additional tRNAs and snoRNAs, which were obviously absent from Repbase. At the end of the annotation process, 21—29% of the orphan reads remained unidentified (reported as 'unknown').

As shown in Suppl. Fig. 18, the qualitative breakdown of orphan reads is very similar between the two batches of HTS libraries, albeit relative abundance differs quite markedly. This is especially striking for two pairs of categories that appear to be somehow in competition for cloning or sequencing: adapters vs. poly-A and miR-1 vs. miR-378.

To look for endo-siRNAs (Nilsen 2008) in our orphan reads (in the two batches of HTS libraries), we used FindPeaks 4.0.8 (Fejes et al. 2008) followed by a manual analysis of the resulting peaks. In practice, orphan reads identified (in the BLASTN analysis) as poly-A tracts, sequencing adapters, miRNAs or snoRNAs were removed before proceeding. Reads were

also filtered on length (19—25 nt) to ensure efficient mapping on the bovine genome. Though this eliminated all the remaining 35-nt sequences, visual inspection of randomly selected read batches confirmed that most dropped sequences were altered beyond recognition. Using SOAP.CONTIG with the same parameter values as for the prediction of miRNA precursors (see above), the qualifying reads were mapped on the unmasked cow contigs. Then, SOAP output was filtered to discard sequences mapping to more than five genomic locations (to avoid repetitive elements) and those mapping to unknown chromosome fragments (too numerous to handle with FINDPEAKS). After conversion to BED format, the filtered SOAP output was further processed by two accessory components of FINDPEAKS (SEPARATEREADS and SORTFILES). FINDPEAKS itself was finally invoked with the following parameters: distribution type (3 = native or sequence coverage), minimum peak size (1), maximum PET (paired end tag) size (1,000 nt), and BEDGRAPH output. For the manual analysis, BEDGRAPH files for the two HTS experiments were simultaneously uploaded to the UCSC genome browser and all peaks above 400 in windows of 20,000 bp were examined by eye (89 chromosomal segments in total). This analysis did not provide any evidence for endo-siRNAs, and evaluated peaks corresponded to previously described categories of RNA, namely tRNAs, rRNAs, other repeat sequences and small pieces of protein coding genes that are known to be highly expressed in skeletal muscle .

A distinct analysis was carried out for the *DLK1-GTL2* domain. Briefly, the two initial SOAP mappings used to excise candidate precursors (see above) were directly processed by the FINDPEAKS pipeline using the same parameter values as genome-wide, except for the minimum peak size (10). This yielded 111 and 146 peaks, respectively for the first and second batch of HTS libraries. Peak locations were then compared to those of the 49 precursors predicted by MIRDEEP to remove peaks corresponding to already identified miRNAs. Finally, the remaining peaks were examined by eye and a coverage cut-off was set at 50, which resulted into 6 expressed precursors not predicted by MIRDEEP to be added to the domain miRNA catalogue (see above), as well as 8 more "stacks" that might correspond to additional miRNAs but were not further considered.

Evaluation of miRNA affinity for *DLK1*

To assess miRNA affinity for the *DLK1* transcript, we assembled two different catalogues of miRNA species encoded by the *DLK1-GTL2* domain. The first one was limited to ovine miRNAs and contained the most abundant representative (when expressed) of the 110 species identified after the first HTS experiment, as well as 4 species specific to the second HTS experiment, 6 alternative molecules reaching > 70% of the most expressed molecule, and 7 edited variants, thus amounting to 127 distinct sequences, all considered equally. In contrast, the second catalogue included 70 human and 77 murine "orthologues" in addition to the 114 (110+4) most abundant ovine miRNA species. When ovine species displayed an alternative 5p extremity relative to miRBase annotation, we considered both possibilities by extracting additional shifted orthologues from human and murine precursors (and in the other way around). This concerned miR-127 (5p), 381 (3p), 409 (3p), 411a (5p), 431 (3p), 487a (3p), 539 (5p), and 654 (5p). Altogether, the second catalogue included 140 triplets of sequences (114 "ovine" species, 8 shifted species, 2 human- and 15 murine-specific species, as well as the human/murine version of miR-668) and was used to compute multiorganism (MO) scores summed across the three orthologous sequences of the *DLK1* transcript (see main text for details).

Two different scoring engines were used. For G-scores, we used a customized version of COMPSEQ (EMBOSS package 6.1.0) allowing to look for k-mers longer than 6 nt. M-scores were computed with MIRANDA 3.0-sept2008 (Betel et al. 2008) using default parameter values, except that we set a conservative score cut-off (140) to avoid spurious weak matches and to speed up computations. Note that using older versions of MIRANDA yielded surprisingly contrasted predictions for some miRNA species (data not shown).

Shuffling was carried out separately for each region of the *DLK1* transcripts (5'-UTR, CDS, 3'-UTR). In all cases, the algorithm was as follows. Using a 3-nt overlapping sliding window, the numbers of occurrences of all nucleotide triplets in the region are recorded. From these counts, we deduce (1) probabilities for starting dinucleotides and (2) conditional probabilities for single base extensions. To generate shuffled sequences, a semi-empirical strategy was preferred to a purely probabilistic approach after comparison of the performances of the two methods (data not shown). First, a starting dinucleotide is

randomly drawn from the empirical distribution. Then, the sequence is extended until the required length by drawing the next nucleotide conditionally on the last two nucleotides. For each nucleotide, we first try to draw from empirical pools containing the true number of single base extensions for each prefix. If this fails (because the current prefix pool has been exhausted before reaching the end of the nascent sequence), we draw from the empirical distribution of single base extensions corresponding to the current prefix. To maximize the odds to get shuffled sequences as close as possible to the true sequence in compositional terms, we generate twice the required number of sequences, rank candidates according to the sum of triplet count differences with the original sequence, and keep the upper half of the list. Hence, to get 10,000 shuffled variants of the 3'-UTR of *DLK1*, we actually generated 20,000 sequences, of which we discarded the 10,000 sequences that were the least similar in trinucleotide composition to the real 3'-UTR.

After scoring all sequences, p-values were computed by comparing each of the species or "quadrille" scores obtained with the true sequence of interest to the corresponding distribution of scores obtained with the shuffled sequences. To account for multiple testing, simple Bonferroni adjustments were achieved by dividing the 0.05 significance threshold by the number of tests.

GO Analysis of miRNA targets

Conserved target sites for conserved miRNA families were downloaded from the human TargetScan 5.1 website (Friedman et al. 2009). From this initial multispecies list, we first discarded all rows corresponding to non-bovine sites. Then, we further filtered rows to keep only those sites targeted by one of the 24 conserved miRNA families with representatives in the *DLK1-GTL2* domain: miR-127, miR-134, miR-136, miR-154, miR-299/299-3p, miR-300, miR-329/362-3p, miR-370, miR-376/376ab/376b-3p, miR-376c, miR-377, miR-379, miR-382, miR-410, miR-411, miR-431, miR-433, miR-487/487b, miR-494, miR-495/1192, miR-496, miR-543, miR-544, miR-758. This resulted into a list of 4,798 sites spread among 2,832 bovine 3'-UTRs.

Control sets of genes of the same sample size were randomly drawn among the 8,458 bovine genes present in the TargetScan list. This is noteworthy as it means that control

genes have all at least one conserved target site for a conserved miRNA family, which could lead to an under-representation of some functional classes (e.g., house-keeping genes). This was carried out either 10,000 or 200,000 times, depending on the total size of GO graph used in the analysis (see main text for details). Whenever a given gene was selected (either randomly or through a cognate miRNA), all the associated GO terms got a hit. This was done without weighting for the number of miRNA target sites in the 3'-UTRs. At the end of the random sampling, the enrichment was statistically evaluated as for *DLK1* affinity, i.e., by comparing the number of hits obtained by each individual GO term with the true set of genes to the distribution of the number of hits obtained by the same term with the sets of control genes. Bonferroni adjustments used the number of GO (or GO Slim) terms actually associated to the bovine TargetScan list (6,930 for the whole GO graph and 55 for the GO Slim graph).

GO associations were fetched from the EBI GOA project (Barrell et al. 2009). We used the human GOA annotation (release 81, 21 January 2010) since TargetScan only provides human gene symbols, even for bovine orthologues. The mapping to GO Slim terms was performed using the corresponding GOA Slim map also available at the EBI FTP server. Annotation of GO terms was achieved by using a local mirror of the GO termdb downloaded from http://www.geneontology.org/ on March 1st, 2010.

Supplemental References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25(17): 3389-3402.

Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, Apweiler R. 2009. The GOA database in 2009-an integrated Gene Ontology Annotation resource. Nucleic Acids Res 37(Database issue): D396-403.

Betel D, Wilson M, Gabow A, Marks DS, Sander C. 2008. The microRNA.org resource: targets and expression. *Nucleic Acids Res* **36**(Database issue): D149-153.

Bonnet E, Wuyts J, Rouze P, Van de Peer Y. 2004. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* **20**(17): 2911-2917.

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* **13**(4): 721-731.

Charlier C, Segers K, Karim L, Shay T, Gyapay G, Cockett N, Georges M. 2001. The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status. *Nat Genet* **27**(4): 367-369.

Durbin R, Eddy SR, Krogh A, Mitchison G. 1998. *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge Univ Pr.

Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* **30**(7): 1575-1584.

Fejes AP, Robertson G, Bilenky M, Varhol R, Bainbridge M, Jones SJ. 2008. FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics* **24**(15): 1729-1730.

Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N. 2008. Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol* **26**(4): 407-415.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**(1): 92-105.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* **15**(10): 1451-1455.

Griffiths-Jones S. 2006. miRBase: the microRNA sequence database. *Methods Mol Biol* **342**: 129-138.

Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A. 2009. BioMart Central Portal--unified access to biological data. *Nucleic Acids Res* **37**(Web Server issue): W23-27.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**(1-4): 462-467.

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* **129**(7): 1401-1414.

Lestrade L, Weber MJ. 2006. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* **34**(Database issue): D158-162.

Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics* **24**(5): 713-714.

Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y et al. 2009. Bos taurus genome assembly. *BMC Genomics* **10**: 180.

McCaskill JS. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* **29**(6-7): 1105-1119.

Nilsen TW. 2008. Endo-siRNAs: yet another layer of complexity in RNA silencing. *Nat Struct Mol Biol* **15**(6): 546-548.

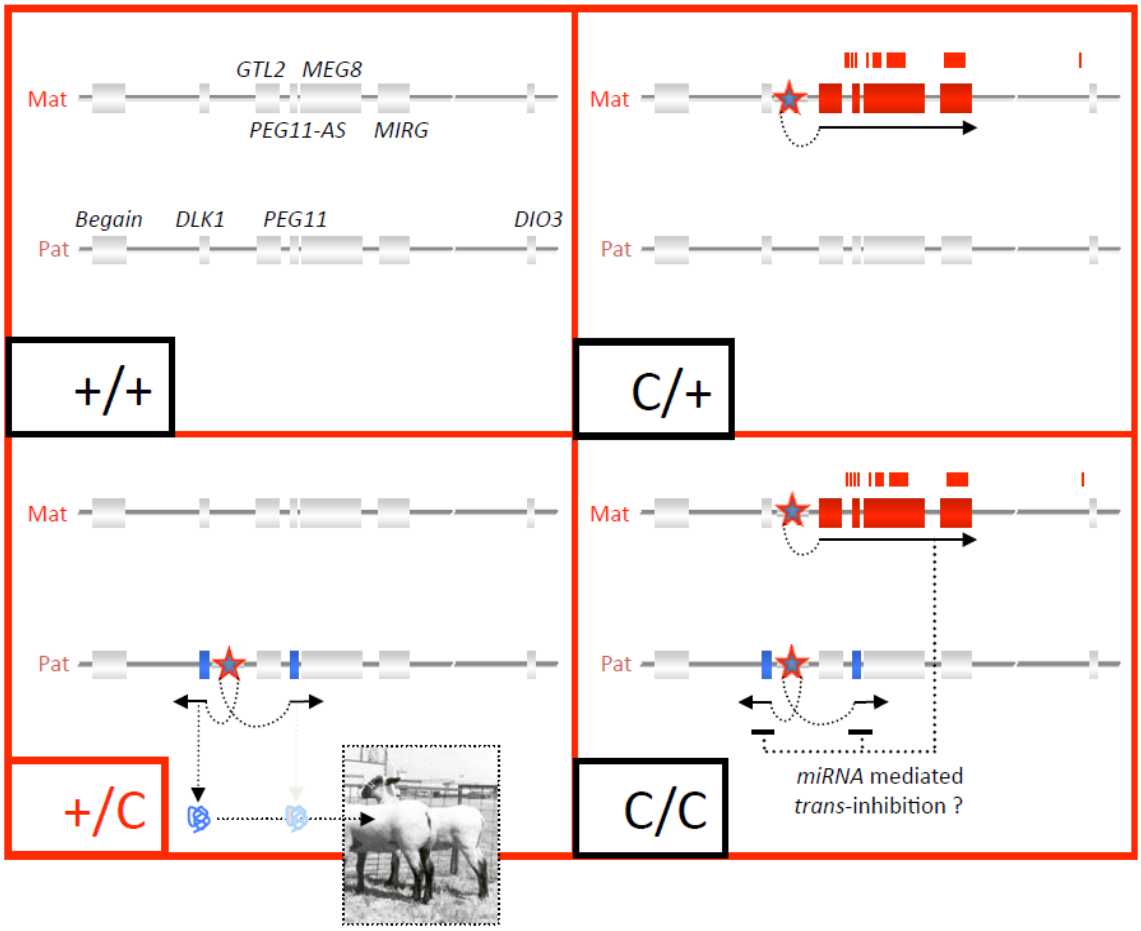
Rathjen T, Pais H, Sweetman D, Moulton V, Munsterberg A, Dalmay T. 2009. High throughput sequencing of microRNAs in chicken somites. *FEBS Lett* **583**(9): 1422-1426.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**(6): 276-277.

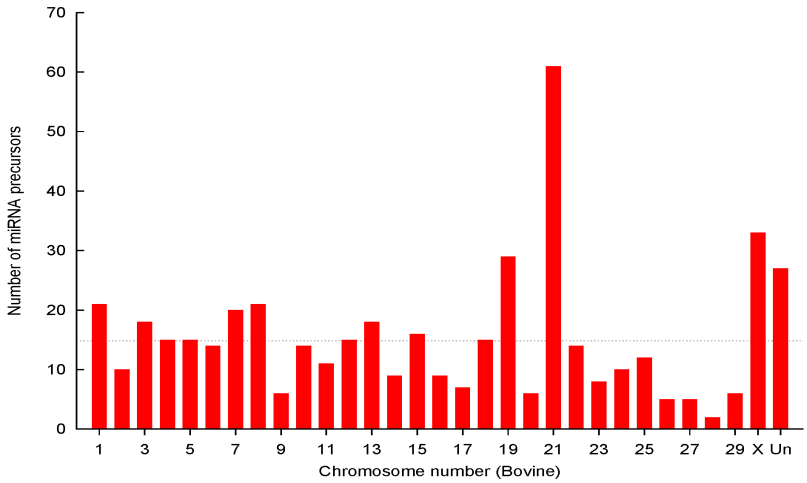
Seitz H, Royo H, Bortolin ML, Lin SP, Ferguson-Smith AC, Cavaille J. 2004. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res* **14**(9): 1741-1748.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**(22): 4673-4680.

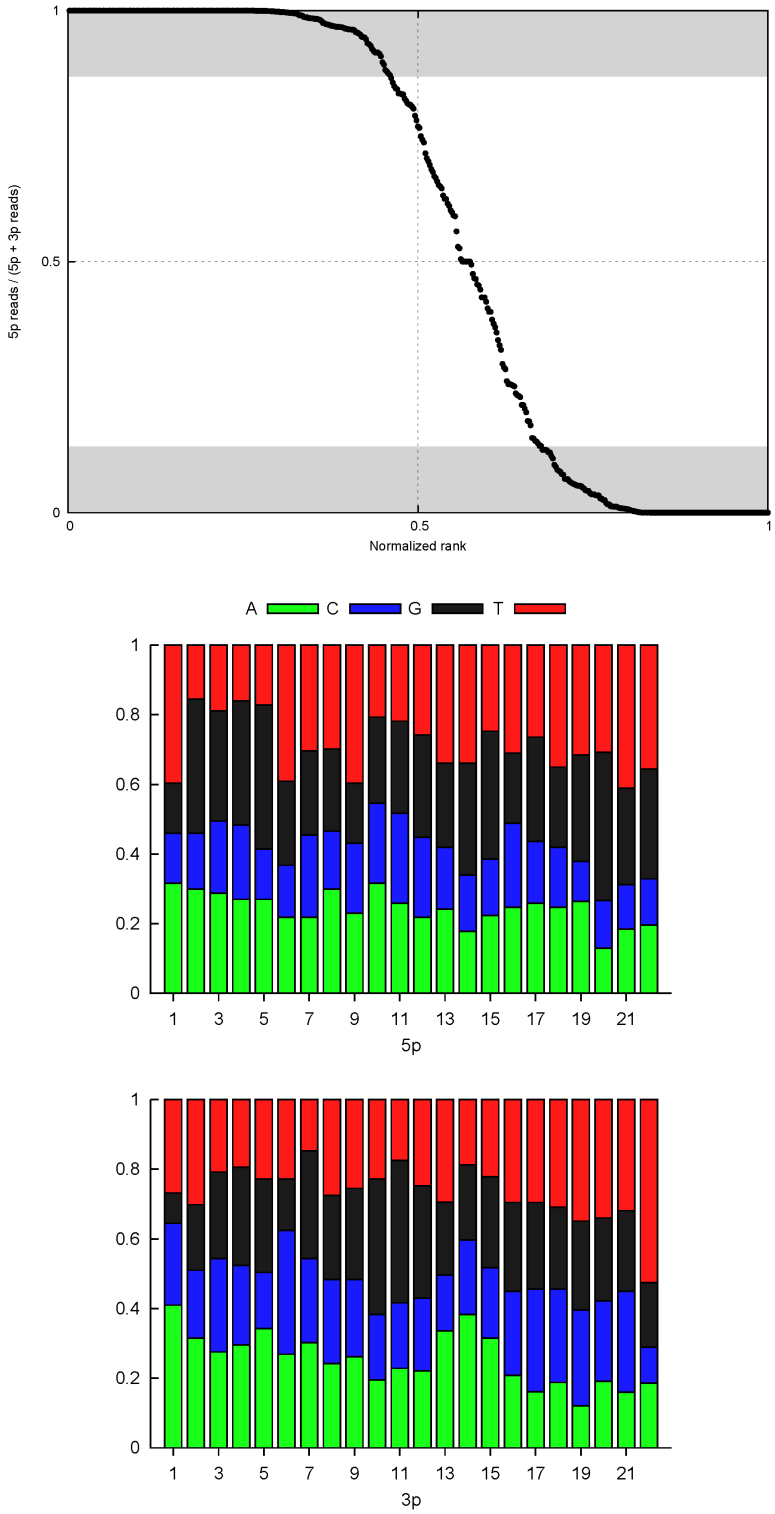
Supplemental Figures



Supplemental Figure 1: Working model for polar overdominance at the ovine *CLPG* locus.

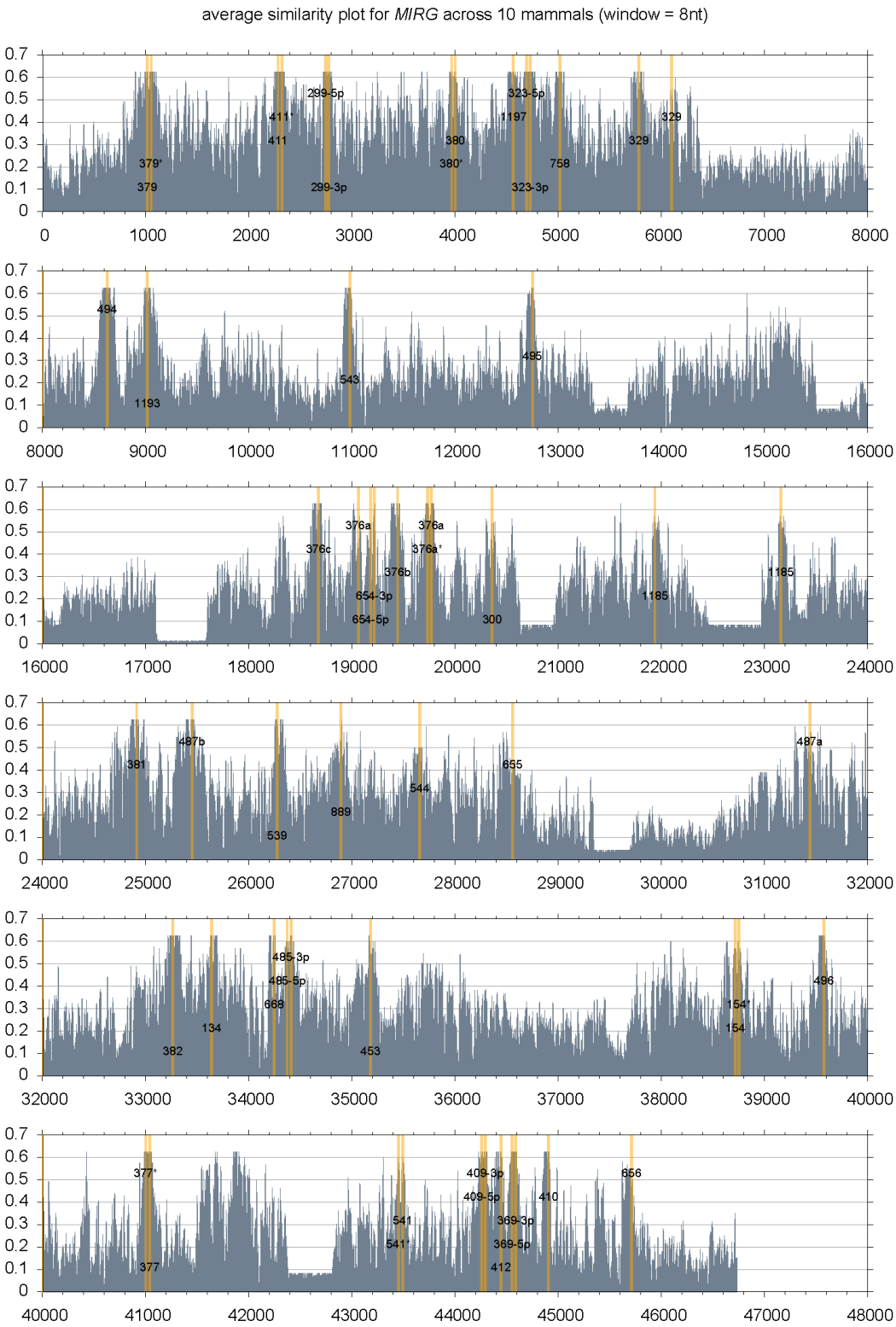


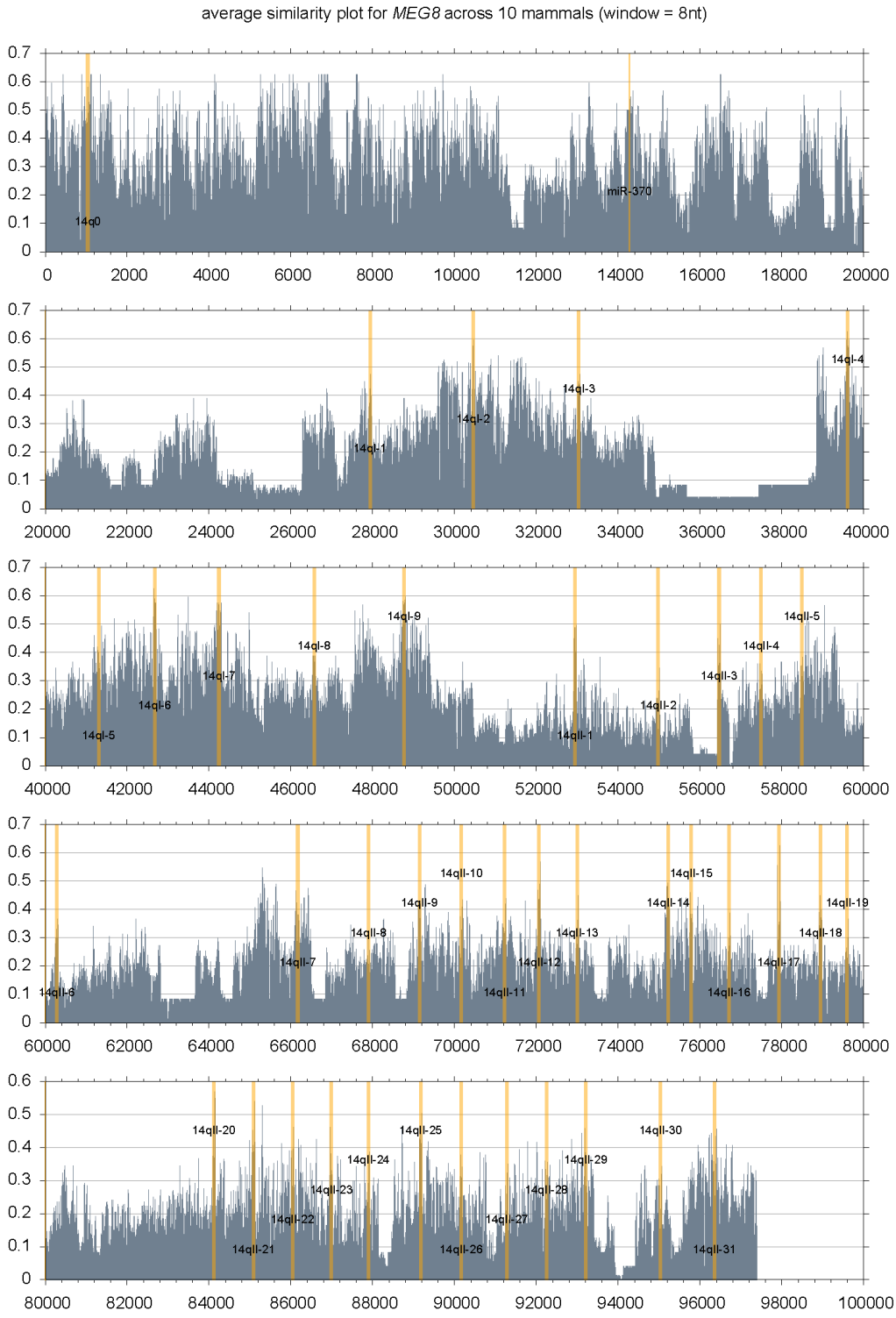
Supplemental Figure 2: Number of microRNA precursors [predicted by miRDeep (Friedlander et al. 2008) based on deep sequencing of small RNA libraries from skeletal muscle of sheep] that map to the different bovine chromosomes. The bovine genome sequence was used as reference for the miRDeep analysis except for 390 Kb of ovine sequence corresponding to the core of the *DLK1-GTL2* domain encompassing the *CLPG* mutation. U: unassigned sequence contigs.



Supplemental Figure 3: (A) Ratio of reads mapping to the 5p arm for the 472 precursors predicted by miRDeep. The graph illustrates the excess of miRNAs spawned from the 5p arm. The shaded areas cover precursors for which either the 5p or 3p arm generates > 87% of the reads, thus defining a star species issued from the other arm. **(B)** Base pair composition of miRNAs derived from the 5p arm (upper panel) or 3p arm (lower panel). miRNA from the populous miR-2284 family were excluded from the analysis. Data are shown for the second HTS experiment, but were in essence identical to that of the first HTS experiment. Position one of the miRNA was defined as the starting position of the most expressed isomir.

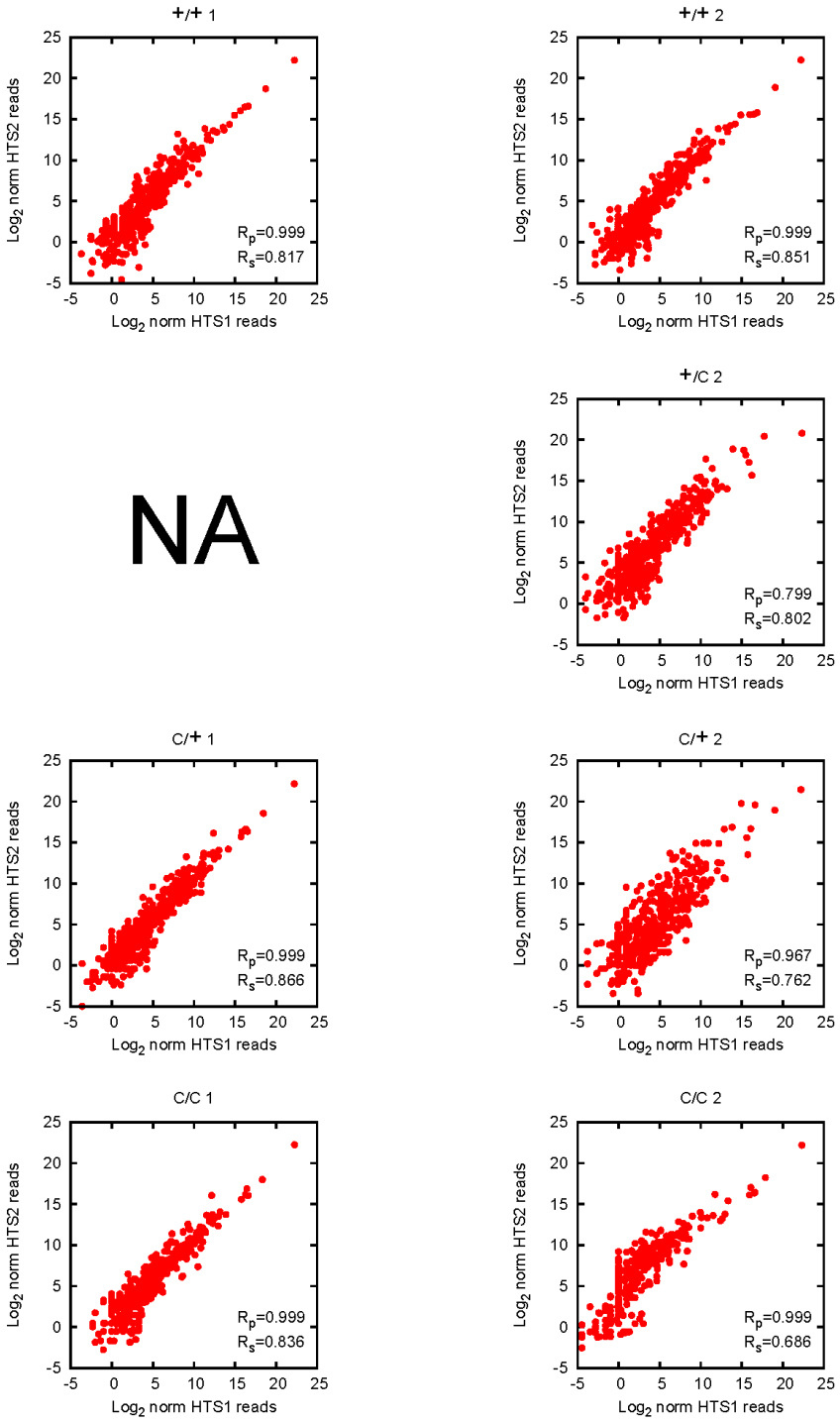
A


average similarity plot for *MIRG* across 10 mammals (window = 8nt)
379*
379
411*
411
299-5p
299-3p
380
380*
323-5p
323-3p
1197
758
329
329
494
1193
543
495
376a
376a
376c
376a*
376b
654-3p
654-5p
300
1185
1185
487b
381
539
889
544
655
487a
485-3p
485-5p
668
134
382
453
154*
154
496
377*
377
541
541*
409-3p
409-5p
369-3p
369-5p
412
410
856

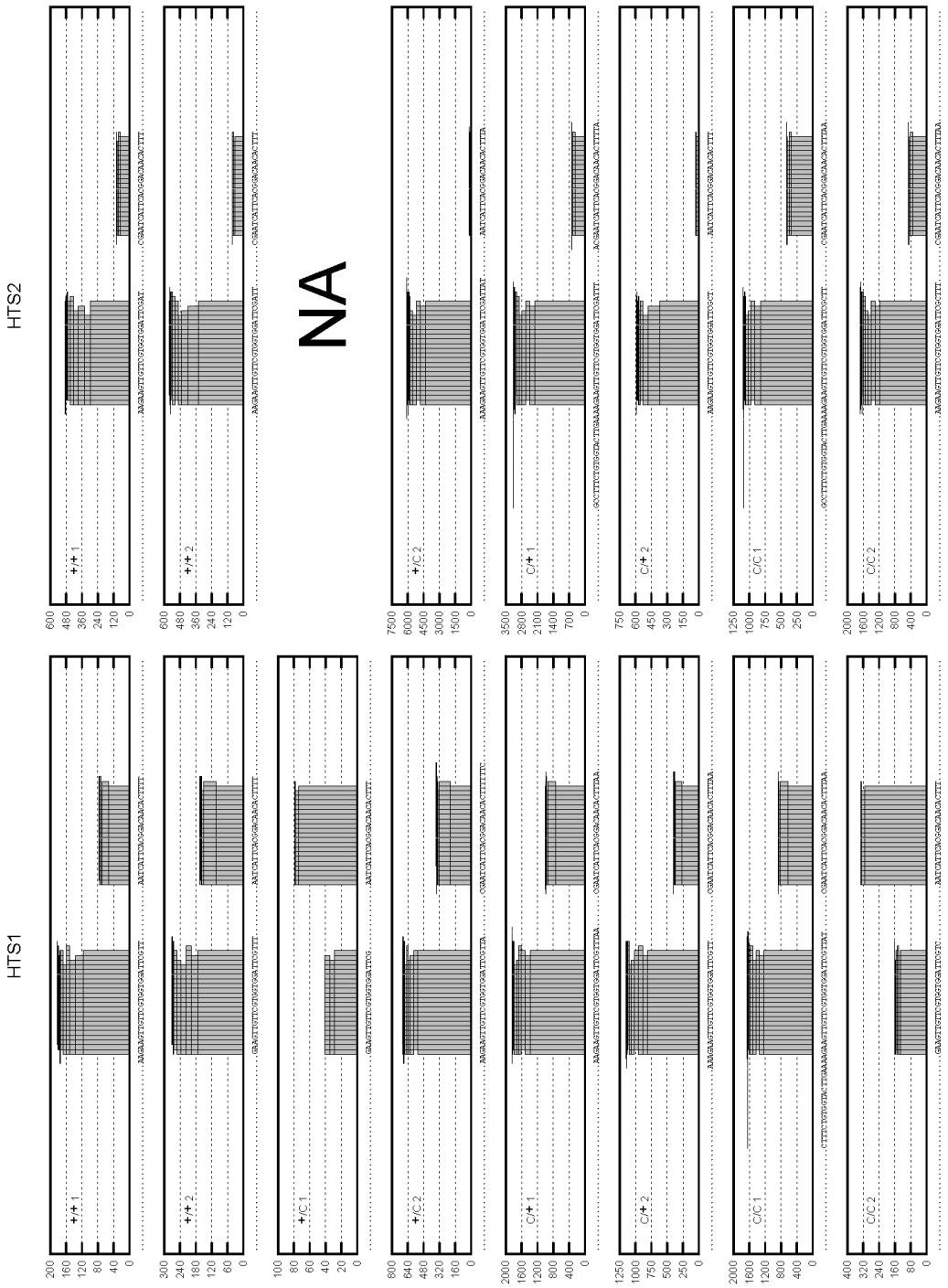


B

Supplemental Figure 4: Plot of average pairwise similarities of sliding 8-nt windows across 10 mammals (human, chimpanzee, orangutan, rhesus macaque, guinea pig, mouse, rat, cow, horse, dog) of the *MIRG* **(A)** and *MEG8* **(B)** genes. The position and identifier of the miRNAs **(A)** and C/D snoRNAs **(B)** are given as horizontal orange bars. The coincidence between the peaks of conservation and miRNAs but not C/D snoRNAs positions is striking.

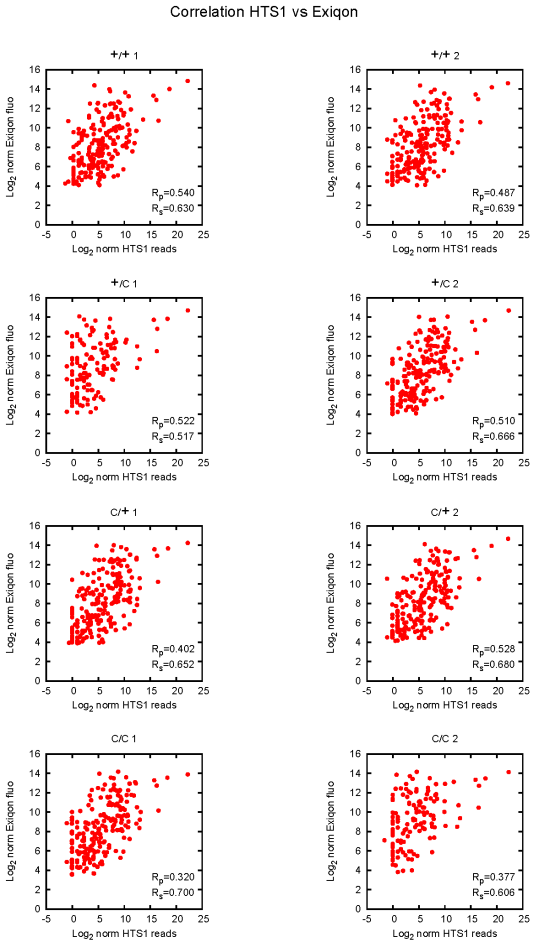


Supplemental Figure 5: Correlation between the expression levels of 826 miRNA species in skeletal muscle (longissimus dorsi) of eight-week old sheeps representing the four possible *CLPG* genotypes *(+/+; CLPG*^{Mat}*/+*^{Pat} *= C/+; +*^{Mat}*/CLPG*^{Pat} *= +/C; CLPG/CLPG = C/C*) estimated from the number of HTS reads obtained in two independent experiments (HTS1 and HTS2). Two animals were analyzed per genotype, except *+/C* for which only one animal was sequenced twice (HTS1 + HTS2). Raw read numbers were divided by the total number of reads for the corresponding animal and multiplied by the average total number of reads across animals. X- and Y-axis correspond to log-base-2 of the corresponding normalized number of reads for the first (HTS1) and second (HTS2) sequencing experiment, respectively. R_P and R_S are Pearson and Spearman correlation coefficients, respectively.

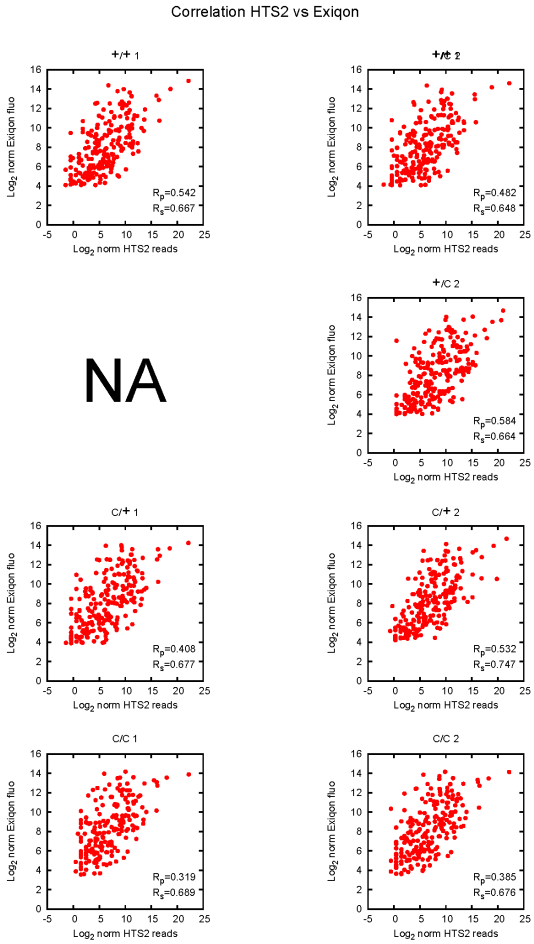


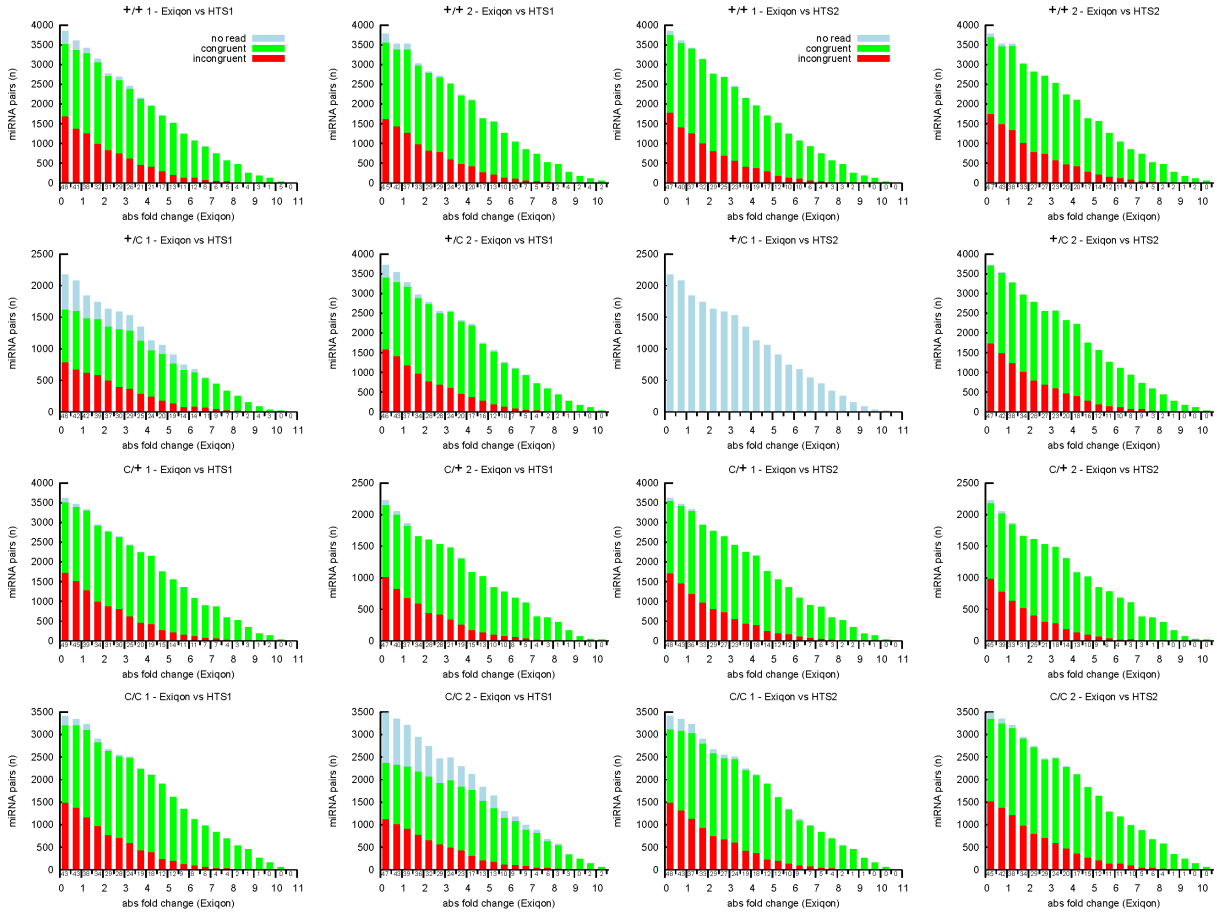
Supplemental Figure 6: Number of sequence reads mapping to the 5p and 3p arm of the miR-382 precursor in the first (HTS1) and second (HTS2) sequencing experiments for the eight animals representing the four possible *CLPG* genotypes. It can for instance be seen that in HTS1, two individuals *(+/C 1* and *C/C 2*) showed 5p/3p ratios that were in opposite direction as the other individuals. The HTS2 experiment did not confirm this finding for *C/C 2*.

A

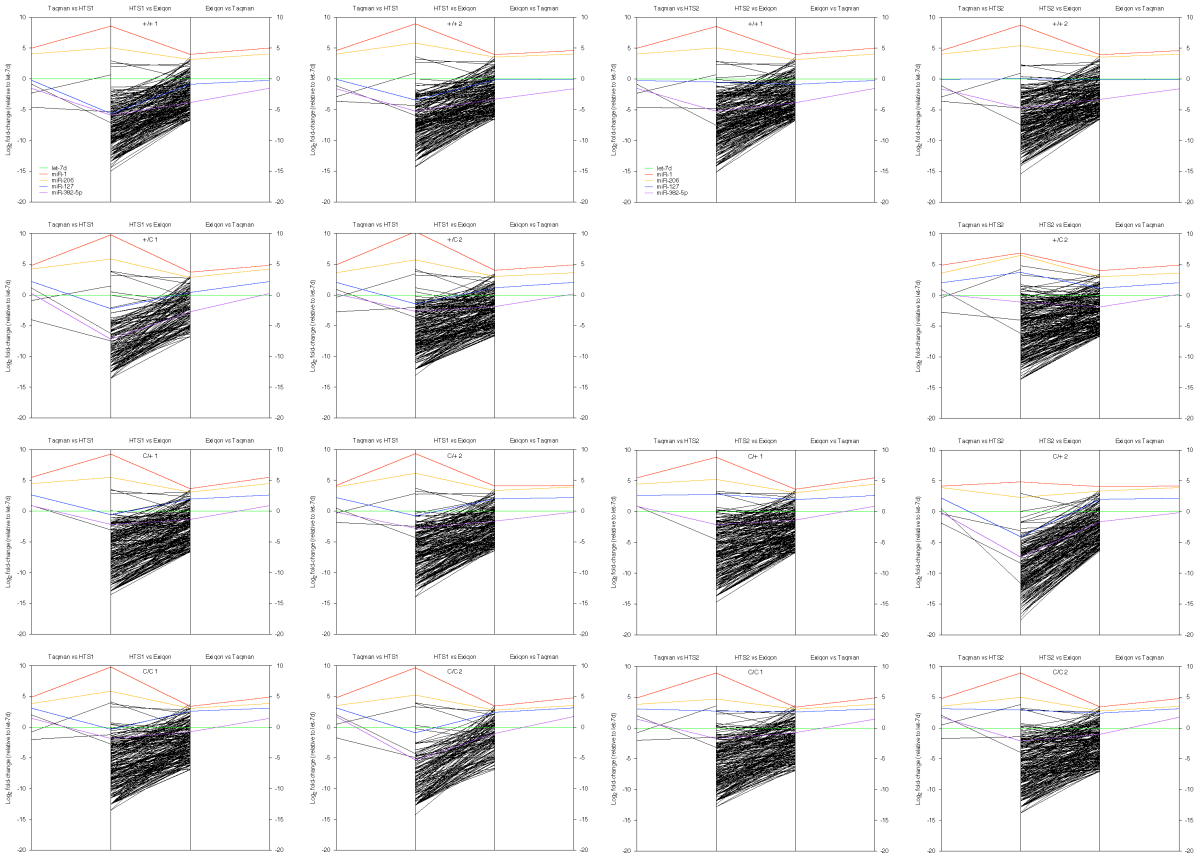

Correlation HTS1 vs Exiqon
+/+ 1
Rp=0.540
Rs=0.630
+/+ 2
Rp=0.487
Rs=0.639
+/C 1
Rp=0.522
Rs=0.517
+/C 2
Rp=0.510
Rs=0.666
C/+ 1
Rp=0.402
Rs=0.652
C/+ 2
Rp=0.528
Rs=0.680
C/C 1
Rp=0.320
Rs=0.700
C/C 2
Rp=0.377
Rs=0.606
Log2 norm Exiqon fluo
Log2 norm HTS1 reads

B

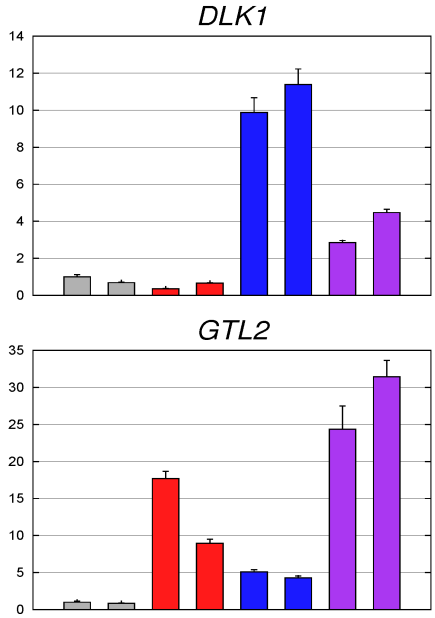

Correlation HTS2 vs Exiqon
+/+ 1
Rp=0.542
Rs=0.667
+/+ 2
Rp=0.482
Rs=0.648
NA
+/C 2
Rp=0.584
Rs=0.664
C/+ 1
Rp=0.408
Rs=0.677
C/+ 2
Rp=0.532
Rs=0.747
C/C 1
Rp=0.319
Rs=0.689
C/C 2
Rp=0.385
Rs=0.676
Log2 norm Exiqon fluo
Log2 norm HTS2 reads



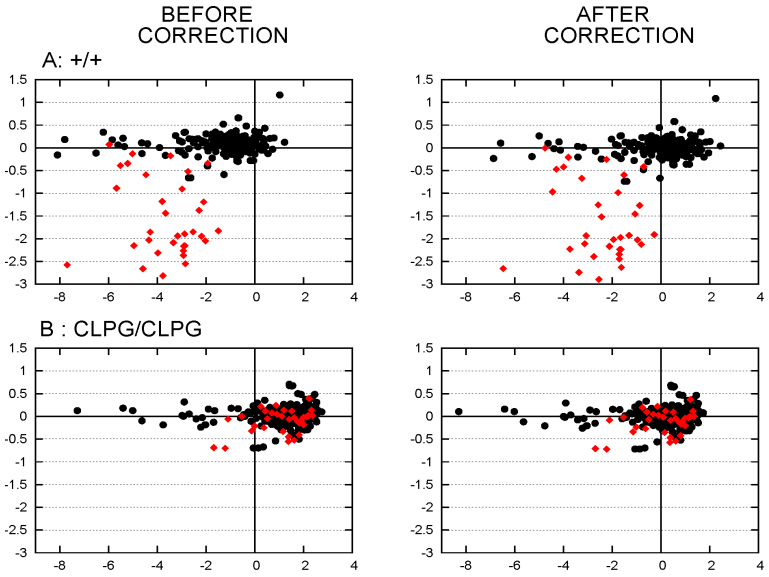
Supplemental Figure 7: (A&B) Correlation between the expression levels of 265 miRNA species in skeletal muscle (longissimus dorsi) of eight-week old sheep representing the four possible *CLPG* genotypes *(+/+; CLPG^{Mat}/+^{Pat} = C/+; +^{Mat}/CLPG^{Pat} = +/C; CLPG/CLPG = C/C*) estimated from (i) the number of HTS reads (A: HTS1 and B: HTS2) and (ii) fluorescent intensities obtained by hybridizing total Hy3-labelled RNA on Exiqon miRCURY™ LNA (Version 9.2 - updated to miRbase 11.0). Raw read numbers were divided by the total number of reads for the corresponding animal and multiplied by the average total number of reads across animals. Raw fluorescent intensities were divided by the total intensity of the corresponding animal and multiplied by the average total intensity across animals. X- and Y-axis correspond to log-base-2 of the corresponding normalized number of HTS reads and normalized intensities, respectively. R_P and R_S are Pearson and Spearman correlation coefficients, respectively. **(C&D)** Number of miRNA pairs ranked congruently (green) or incongruently (red) with regard to expression level as assessed by HTS (C: HTS1 and D: HTS2) versus Exiqon. miRNA pairs are categorized according to the fold difference in expression level assessed on the Exiqon arrays (X-axis). The numbers under the bars correspond to the % of incongruent pairs in the corresponding category.



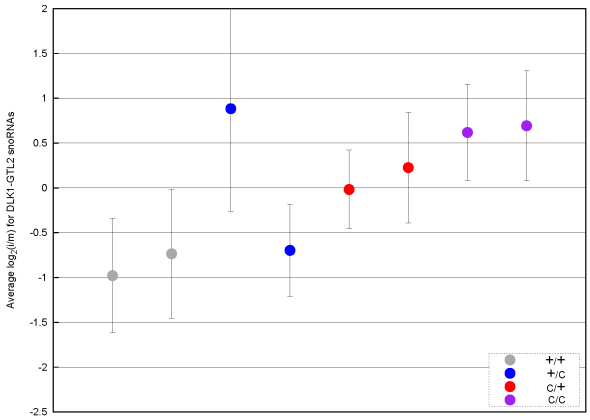
Supplemental Figure 8: Comparison of miRNA expression levels (relative to *let-7d*) in skeletal muscle of 8-week old animals representing the four possible *CLPG* genotypes *(+/+; CLPG^{Mat}/+^{Pat} = C/+; +^{Mat}/CLPG^{Pat} = +/C; CLPG/CLPG = C/C*), estimated by: (i) HTS (**A**: HTS1 and **B**:HTS2), (ii) Exiqon array-hybridization and (iii) QRT-PCR. Relative expression levels are measured on a log-base-2 scale. Bars corresponding to miRNAs also studied by QRT-PCR are colored as indicated in the legend.



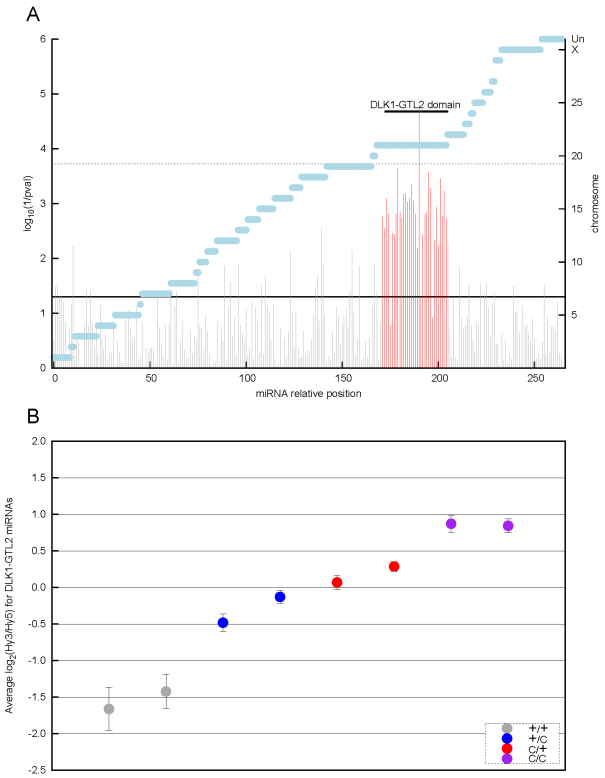
Supplemental Figure 9: Quantitative RT-PCR analysis of the expression level of *DLK1* and *GTL2* in longissimus dorsi of eight week old sheep representing the four possible genotypes at the *CLPG* locus: *+/+* (gray bars), $CLPG^{Mat}/+^{Pat}$ (red bars), $+^{Mat}/CLPG^{Pat}$ (blue bars), *CLPG/CLPG* (purple bars). Expression levels are relative to the *+/+* individual marked as "Ref". Error bars correspond to standard errors of the estimates over triplicates.



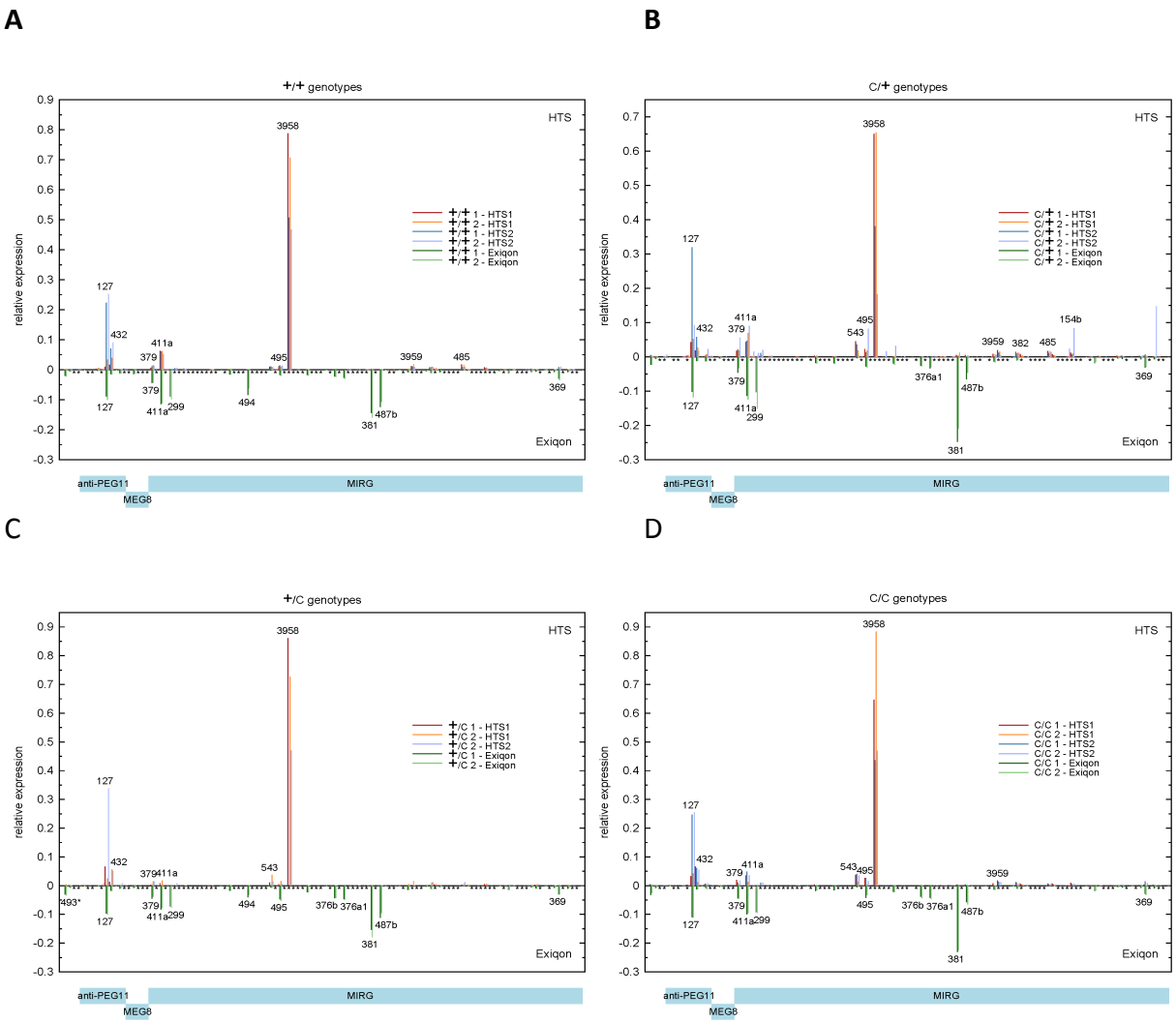
Supplemental Figure 10: Illustration of the need for a correction of $\log_2(i/m)$ to account for the deviations of the average (across miRNAs) of this parameter between individuals. Two individuals are shown, corresponding to a *+/+* (A) and a *CLPG/CLPG* animal (B). Each dot corresponds to a miRNA mapping either to the *DLK1-GTL2* domain (red diamonds) or not (gray circles). The X-axis corresponds to $\log_2$ of the ratio between the number of adjusted reads for the individual over the average number of adjusted reads across the eight samples. The Y-axis corresponds to $\log_2$ of the ratio between the normalized Hy3 fluorescence of the sample over the normalized Hy5 fluorescence of a pool of the eight samples. It can be seen that for most of the miRNAs, both individuals have low $\log_2(i/m)$ despite the fact that miRNA read numbers were adjusted for the total number of mappable reads for a given individual. For unexplained reasons, some animals indeed had zero or near zero reads for miRNAs that were characterized by low level expression, yet with considerably higher read numbers in most animals. The correction, based on the average $\log_2$ ratio over non *DLK1-GTL2* miRNAs (to avoid erasing the *CLPG* effect if it exists) shifts the cloud to the right.



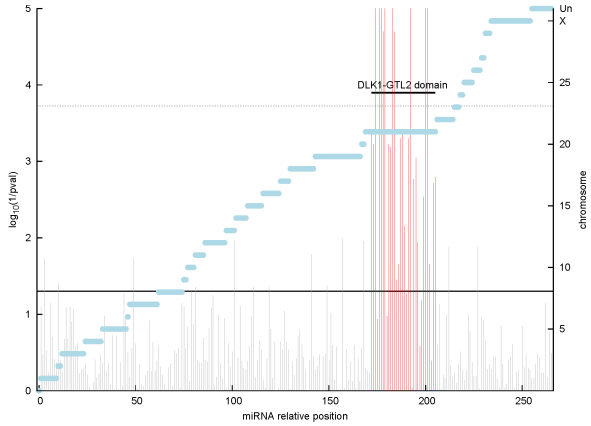
Supplemental Figure 11: Average expression level, relative to the mean expression level of seven individuals sequenced twice (HTS1 and HTS2), of 25 small RNA species derived from C/D snoRNAs within the *DLK1-GTL2* domain in skeletal muscle of eight sheep sorted by *CLPG* genotype *(*gray: *+/+;* blue: $+^{Mat}/CLPG^{Pat}$; red: $CLPG^{Mat}/+^{Pat}$*;* purple: *CLPG/CLPG*). Error bars correspond to 1.96 x the standard error of the estimate.



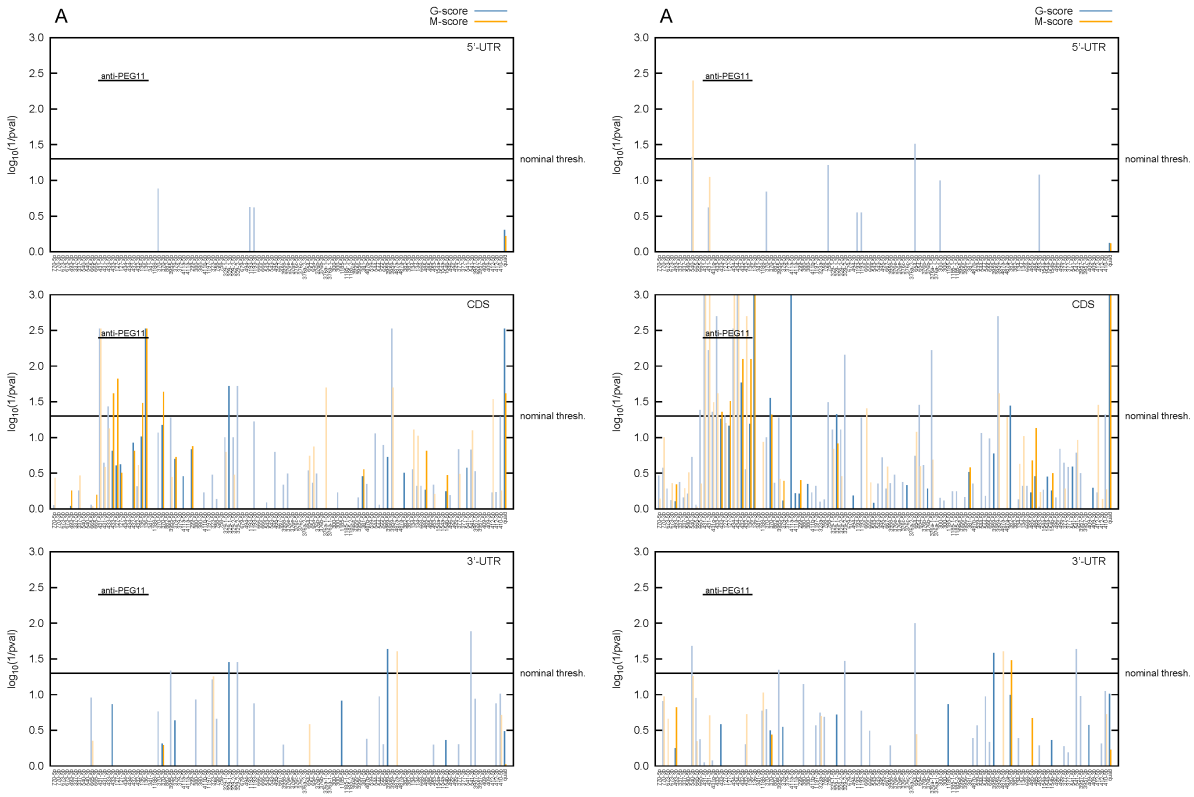
Supplemental Figure 12: (A) $Log_{10}(1/p)$ values of the effect of *CLPG* genotype on the expression level of 265 small RNAs in skeletal muscle of eight 8-week old sheep. Expression levels were estimated from fluorescence intensity on Exiqon arrays. The statistical significance of the *CLPG* effect was estimated by ANOVA. Gray vertical bars correspond to miRNAs outside of the *DLK1-GTL2* domain, red vertical bars to miRNAs from the *DLK1-GTL2* domain and orange vertical bars to small RNAs derived from C/D snoRNA precursors. Horizontal black lines correspond to the nominal (plain line) and Bonferroni-adjusted (dotted line) 5% significance thresholds. Horizontal blue bars mark the different chromosomes (right Y-axis). UN: correspond to unassigned sequence contigs. (B). Average expression level, relative to the mean expression level of seven individuals sequenced twice (HTS1 and HTS2), of 34 miRNAs from the *DLK1-GTL2* domain in skeletal muscle of eight sheep sorted by *CLPG* genotype *(*gray: *+/+;* blue: $+^{Mat}/CLPG^{Pat}$; red: $CLPG^{Mat}/+^{Pat}$*;* purple: *CLPG/CLPG*). Error bars correspond to 1.96 x the standard error of the estimate.



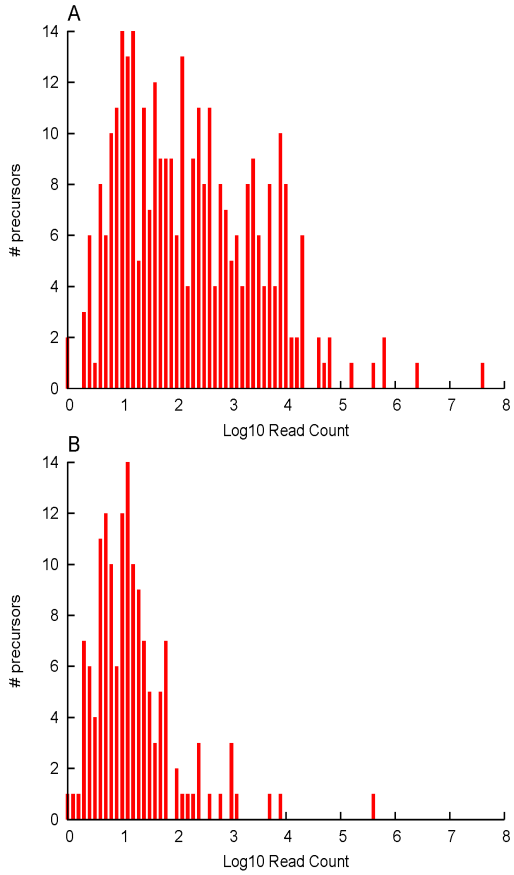
Supplemental Figure 13: Relative abundance of miRNA from the *DLK1-GTL2* domain in skeletal muscle of eight 8-week old sheep representing the four possible *CLPG* genotypes (A: *+/+;* B: *CLPG^{Mat}/+^{Pat} = C/+;* C: *+^{Mat}/CLPG^{Pat} = +/C;* D: *CLPG/CLPG = C/C*). Upper half: Reads derived from the corresponding miRNA expressed as a percentage of the total number of reads derived from all miRNAs in the *DLK1-GTL2* domain. Lower half: Fluorescence intensity for the corresponding miRNA expressed as a percentage of the total fluorescence for all miRNAs in the *DLK1-GTL2* domain. miRNAs that could not be interrogated on the Exiqon array are marked by asterisks. The colors of the vertical bars distinguish the individuals and sequencing experiments as indicated in the inlet. The limits of the long non-coding RNA precursors are marked by the blue bars underneath the graph.



Supplemental Figure 14: Experiment-wide log(1/p)-values (vertical red lines) of the effect of *CLPG* genotype on the expression level of 265 miRNAs in longissimus dorsi of eight 8-week old animals (2x *+/+;* 2x *CLPG^{Mat}/+^{Pat};* 2x *+^{Mat}/CLPG^{Pat};* 2X *CLPG/CLPG*). The analysis combines HTS and array hybridization data. In this analysis, the actual p-value of the miRNA-specific F-statistics was determined from permutation data rather than from tables. The horizontal blue bars mark the limits between the different chromosomes (right Y-axis). The horizontal black lines correspond to the nominal (full line) and Bonferroni-corrected (265 tests; dotted line) 5% significance thresholds.

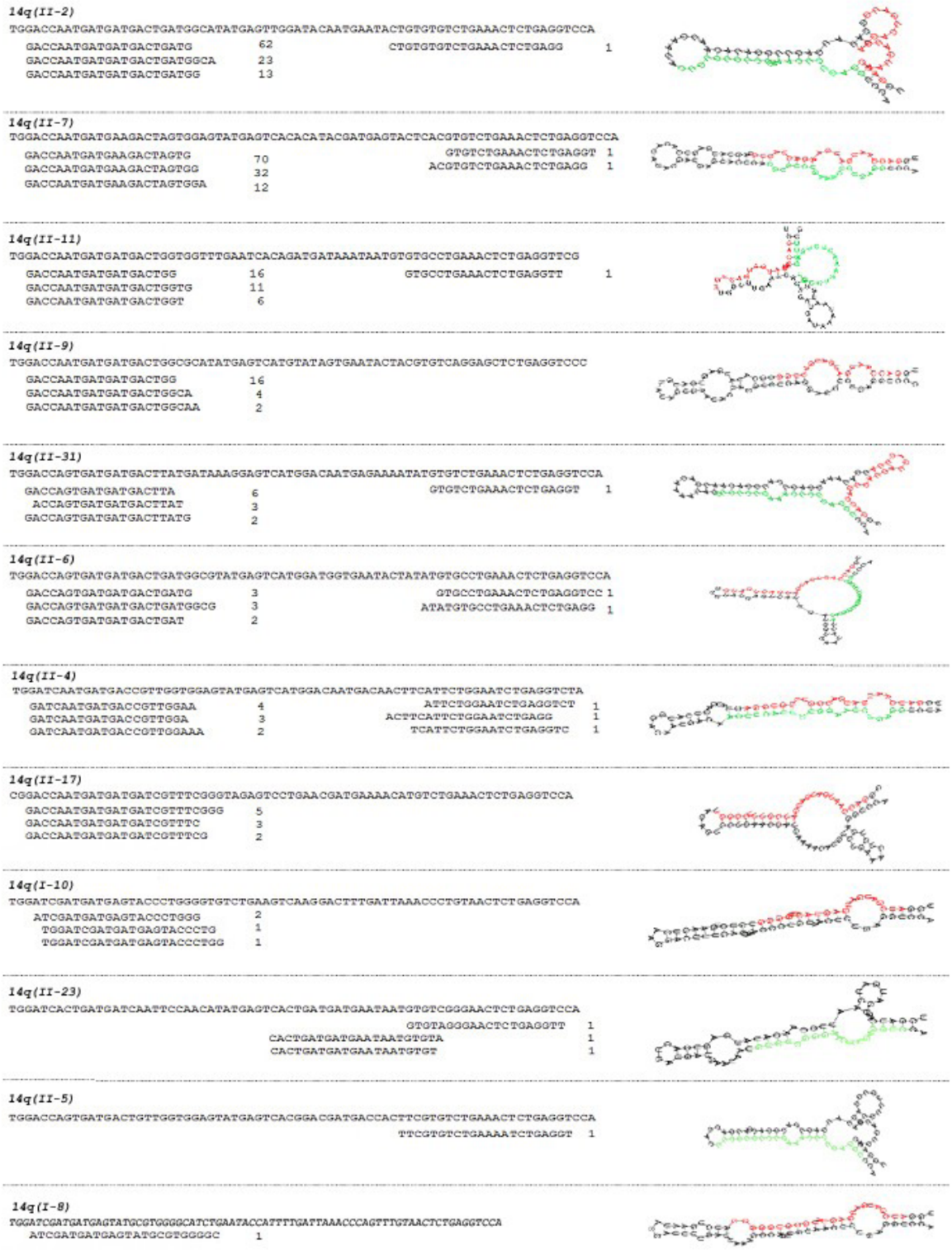


Supplemental Figure 15: (A) Statistical significance (log(1/p)) of the affinity of ovine miRNAs in the *DLK1-GTL2* domain for the 5'UTR, coding sequence (ORF) and 3'UTR of the ovine *PEG11*. The affinity was measured using either G- (blue) or M-scores (orange) as defined in the text. The last pair of bars ("quad") at the right of the graph correspond to the quadrille scores, the remaining bars to the species-scores and are labeled accordingly. p-values were determined using the sequence shuffling test described in the main text. Species-scores require a Bonferroni correction for 127 (or 140) independent tests. (B) Same as in (A) except that the scores are "multiorganism (MO) scores" combining information from sheep, human and mouse.

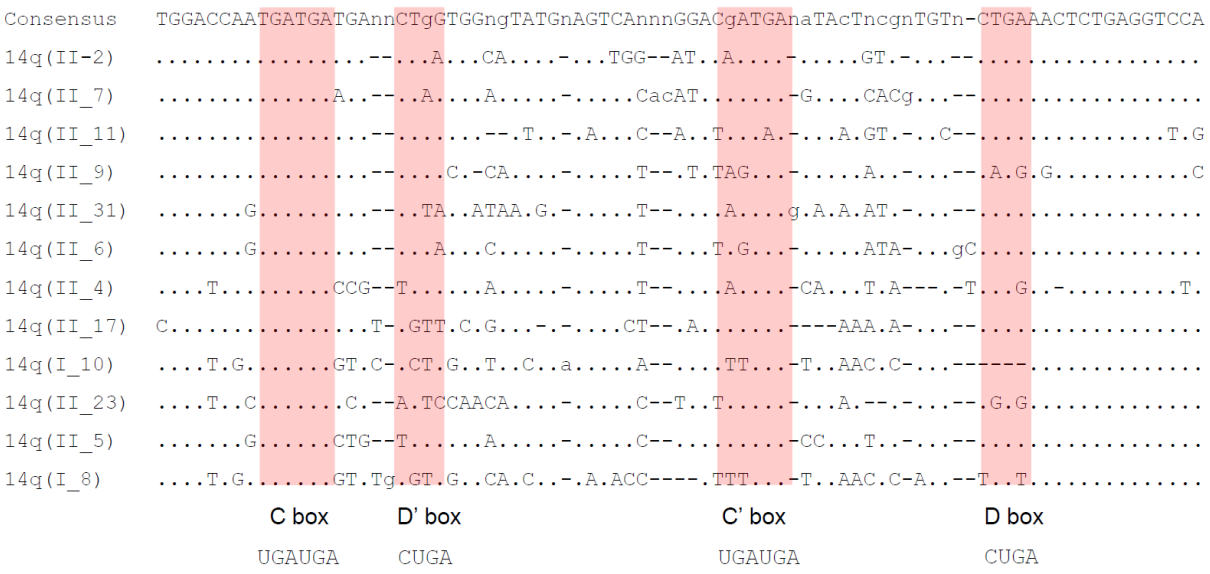


Supplemental Figure 16: (A) Frequency distribution of read numbers (log_{10}) mapping to previously known miRNA precursors. (B) Same for newly detected miRNA precursors.

A


14q(II-2)
TGGACCAATGATGATGACTGATGGCATATGAGTTGGATACAATGAATACTGTGTGTCTGAAACTCTGAGGTCCA
GACCAATGATGATGACTGATG 62
GACCAATGATGATGACTGATGGCA 23
GACCAATGATGATGACTGATGG 13
CTGTGTGTCTGAAACTCTGAGG 1
14q(II-7)
TGGACCAATGATGAAGACTAGTGGAGTATGAGTCACACATACGATGAGTACTCACGTGTCTGAAACTCTGAGGTCCA
GACCAATGATGAAGACTAGTG 70
GACCAATGATGAAGACTAGTGG 32
GACCAATGATGAAGACTAGTGGA 12
GTGTCTGAAACTCTGAGGT 1
ACGTGTCTGAAACTCTGAGG 1
14q(II-11)
TGGACCAATGATGATGACTGGTGGTTTGAATCACAGATGATAAATAATGTGTGCCTGAAACTCTGAGGTTCG
GACCAATGATGATGACTGG 16
GACCAATGATGATGACTGGTG 11
GACCAATGATGATGACTGGT 6
GTGCCTGAAACTCTGAGGTT 1
14q(II-9)
TGGACCAATGATGATGACTGGCGCATATGAGTCATGTATAGTGAATACTACGTGTCAGGAGCTCTGAGGTCCC
GACCAATGATGATGACTGG 16
GACCAATGATGATGACTGGCA 4
GACCAATGATGATGACTGGCAA 2
14q(II-31)
TGGACCAGTGATGATGACTTATGATAAAGGAGTCATGGACAATGAGAAAATATGTGTCTGAAACTCTGAGGTCCA
GACCAGTGATGATGACTTA 6
ACCAGTGATGATGACTTAT 3
GACCAGTGATGATGACTTATG 2
GTGTCTGAAACTCTGAGGT 1
14q(II-6)
TGGACCAGTGATGATGACTGATGGCGTATGAGTCATGGATGGTGAATACTATATGTGCCTGAAACTCTGAGGTCCA
GACCAGTGATGATGACTGATG 3
GACCAGTGATGATGACTGATGGCG 3
GACCAGTGATGATGACTGAT 2
GTGCCTGAAACTCTGAGGTCC 1
ATATGTGCCTGAAACTCTGAGG 1
14q(II-4)
TGGATCAATGATGACCGTTGGTGGAGTATGAGTCATGGACAATGACAACTTCATTCTGGAATCTGAGGTCTA
GATCAATGATGACCGTTGGAA 4
GATCAATGATGACCGTTGGA 3
GATCAATGATGACCGTTGGAAA 2
ATTCTGGAATCTGAGGTCT 1
ACTTCATTCTGGAATCTGAGG 1
TCATTCTGGAATCTGAGGTC 1
14q(II-17)
CGGACCAATGATGATGATCGTTTCGGGTAGAGTCCTGAACGATGAAAACATGTCTGAAACTCTGAGGTCCA
GACCAATGATGATGATCGTTTCGGG 5
GACCAATGATGATGATCGTTTC 3
GACCAATGATGATGATCGTTTCG 2
14q(I-10)
TGGATCGATGATGAGTACCCTGGGGTGTCTGAAGTCAAGGACTTTGATTAAACCCTGTAACTCTGAGGTCCA
ATCGATGATGAGTACCCTGGG 2
TGGATCGATGATGAGTACCCTG 1
TGGATCGATGATGAGTACCCTGG 1
14q(II-23)
TGGATCACTGATGATCAATTCCAACATATGAGTCACTGATGATGAATAATGTGTCGGGAACTCTGAGGTCCA
CACTGATGATGAATAATGTGTA 1
CACTGATGATGAATAATGTGT 1
GTGTAGGGAACTCTGAGGTT 1
14q(II-5)
TGGACCAGTGATGACTGTTGGTGGAGTATGAGTCACGGACGATGACCACTTCGTGTCTGAAACTCTGAGGTCCA
TTCGTGTCTGAAAATCTGAGGT 1
14q(I-8)
TGGATCGATGATGAGTATGCGTGGGGCATCTGAATACCATTTTGATTAAACCCAGTTTGTAACTCTGAGGTCCA
ATCGATGATGAGTATGCGTGGGGC 1

B



Supplemental Figure 17: (A) Small RNAs derived from C/D snoRNA precursors mapping to the *DLK1-GTL2* domain. Isomirs are aligned with their cognate precursors, and numbers of corresponding reads indicated. The position of the most abundant 5p (red) and 3p (green) derived small RNAs on the RNAFOLD-predicted hairpin is shown. **(B)** CLUSTALW alignment of the 12 C/D snoRNAs. Consensus sequence and localization of C/D and C'/D' boxes are displayed.

Available online at www.genome.org

Supplemental Table 1: Genome-wide MIRDEEP miRNA catalogue built from HTS data. Precursor names are taken from miRBase 14, based on sequence similarity of the corresponding miRNA species. Chromosomal locations are expressed in bovine coordinates, except for the *DLK1-GTL2* domain where coordinates are relative to the revised GenBank entry AF354168. For clarity, only non-singleton family numbers are shown. Species types (mature, star, 5p, 3p) were determined as in Landgraf et al. (2007). Note that species types, sequences and seeds are based on HTS1 data only, with the most abundant isomir from each arm selected as the representative of its species. Read numbers are computed on the merged data sets and correspond to the height of the stack for each miRNA species. See main text and Supplemental Methods for details.

GO	GO ID	GO TERM NAME	N° of genes	p_{NOM}	p_{BONF}
SLIM	GO:0005488	binding	2070	$5x10^{-4}$	0.027
	GO:0005515	protein binding	1530	$2x10^{-4}$	0.011
	GO:0003676	nucleic acid binding	640	$\leq 1x10^{-4}$	0.005
	GO:0030528	transcription regulator activity	387	$\leq 1x10^{-4}$	0.005
	GO:0050789	regulation of biological process	1320	$\leq 1x10^{-4}$	0.005
	GO:0005634	nucleus	964	$\leq 1x10^{-4}$	0.005
	GO:0007275	multicellular organismal development	631	$\leq 1x10^{-4}$	0.005
	GO:0030154	cell differentiation	363	$\leq 1x10^{-4}$	0.005
WHOLE	GO:0000932	cytoplasmic mRNA processing body	15	$9.5x10^{-5}$	0.363
	GO:0003723	RNA binding	126	$6.5x10^{-5}$	0.215
	GO:0006511	ubiquitin-dependent protein catabolic process	45	$3.5x10^{-5}$	0.129
	GO:0051246	regulation of protein metabolic process	21	$2x10^{-5}$	0.098
	GO:0007399	nervous system development	110	$1.5x10^{-5}$	0.098
	GO:0005634	nucleus	932	$\leq 5x10^{-6}$	0.034
	GO:0045449	regulation of transcription	242	$\leq 5x10^{-6}$	0.034
	GO:0043565	sequence-specific DNA binding	140	$\leq 5x10^{-6}$	0.034

Table B&C available online at www.genome.org

Supplemental Table 2: (A) Gene ontology terms enriched amongst mammalian genes with conserved target sites for miRNAs in the *DLK1-GTL2* domain. **(B&C)** List of bovine genes targeted by miRNAs encoded in the *DLK1-GTL2* domain for which the GO annotation was significantly enriched. Only 3'-UTR conserved target sites for conserved miRNA families identified in the human TargetScan were considered. To be included, a given gene has to be (1) targeted by at least one the miRNA families with representatives in the *DLK1-GTL2* domain and (2) associated to one or more of the eight most enriched GO terms. B & C correspond to two variants of the GO analysis (Slim and whole; see main text and Supplemental Methods for details).

PriFW-miR-376a-Ov	GGGTCTCTCGTCCATCCATCCAC
PriRV-miR-376a-Ov	ACCACCCCCTTCCTTCTTTGATG
PriFW-miR-1185-Ov	ACCGGCTCCTCCCCCTTGGTCTC
PriFW2-miR-1185-Ov	GGGGCCGGGTCTCAGTGCTCAG
PriFW-miR-381-Ov	GATTCTGCATCCCCCACCCCTTC
PriRV-miR-381-Ov	CCTCTGCCAAGTGGTGACCCTGA
PriFW-miR-478b-Ov	CTAATTTGTCCCCAGGACTGTGC
PriRV-miR-478b-Ov	GCACGGCACGATGAAATATCCTT
PriFW-miR-539-Ov	TTGGCTGAGTTGCAGTCCCTGTC
PriRV-miR-539-Ov	CCAAAGGGTTGACACACGCATCT
PriFW-miR-544-Ov	TGAACTGTCAGCTAGAGGGGTGT
PriRV-miR-544-Ov	GTGTGTTGGCCTGGGAAGTGTTC
PriFW-miR-655-Ov	ATCTGAGTTCCCAGCTCCCTGAG
PriRV-miR-655-Ov	TCACAGCTCACACCAGGCACACG
PriFW-miR-487a-Ov	CCAGTCCCCAATTCGTCCTTAGG
PriRV-miR-487a-Ov	CACTGTGATCCACCAGCCACGTA
PriFW-miR-382-Ov	GATGCGCGTCGTCTGTCTCACTA
PriRV-miR-382-Ov	CACCCAGTCGCGTCTTCCATGAA
PriFW-miR-134-Ov	TCGTCCGAGCTGTAAGTCCTCTG
PriRV-miR-134-Ov	ACGTCCACCGCTGTCCCCAAAGT
PriFW-miR-485-Ov	TCTCGGGTGAGCATGCAATGAAT

PriRV-miR-485-Ov	CGTCAATAAAGCAGGCCTCCTCA
PriFW-miR-453-Ov	TCCGCTCATCCTCATCCTCAGCA
PriRV-miR-453-Ov	GGCCAACTGCAGGTTCAAGTACA
PriFW-miR-154a-Ov	TCGGGGGTATGTTTCACTCATGA
PriRV-miR-154a-Ov	TGCGATCAACACTCCCTGCACTT
PriFW-miR-496-Ov	TGACAGTGTGTTGCCAGCATCTT
PriRV-miR-496-Ov	AGTGGCATGGGTCACACATCATC
PriFW-miR-377-Ov	CGCAGCTGGACCCTGGAGACAGT
PriRV-miR-377-Ov	CCTCATCCACGGGACAGGTGTGG
PriFW-miR-409-412-369-Ov	GTCTTCTGCAAGCTCAGCCACCT
PriRV-miR-369-412-369-Ov	ACCCCGTTCTCACCCCGACAAG
PriFW-miR-410-Ov	CCTGCCTCCTCTCACCTGTGATG
PriRV-miR-410-Ov	TCCAGGAAGCTCTTGCAGGATGA
PriFW-miR-656-Ov	GTTGTGGGCGGGCCTGTGAAGTC
PriRV-miR-656-Ov	ATGCAGCTGGAAAACGAACAAGC
PriFW-miR-1197-Ov	GGTCTTGGCCGCCCTTGCTAAT
PriRV-miR-1197-Ov	GGGCCAGTGGCTCTTCCTCAAA
PriFW-miR-1193-Ov	GAGAAGGGCCCAGCTTGCTGTG
PriRV-miR-1193-Ov	CTTTTCCAGAGTCGCCATGTGC
PriFW2-miR-134-Ov	ACAGCCGTGGTTTTGTCATCAG
PriRV2-miR-134-Ov	GCTGTCCCCAAAGTCGGCAGGA
PriFW2-miR-382-Ov	CTAACACTCTGTGTCCCATCTG
PriRV2-miR-382-Ov	CTGGGGTGGGCATCCTTAGAGG
PriFW2-miR-410-Ov	TGTGATGTCTGATCCACCCTCA
PriRV2-miR-410-Ov	AGCCCCATGCATTCACATACCC
PriFW2-miR-656-Ov	CAGCCTCTGTCTGCAGTCAACC
PriRV2-miR-656-Ov	CCAGTGGCTTGTCCTACATCAA
PriFW-miR-770-Ov	CCGCAGCTACAGACAGGGTACA
PriRV-miR-770-Ov	TGGGGACCCTTCTCTGAGTAGC
PriFW-miR-337-Ov	TGTCTTGAACCCTTCTCGCTCT
PriRV-miR-337-Ov	GTGGTCGACAAGCTTGAGGACA
PriFW-miR-665-Ov	CCGGGACAGACAGGTGGACATC
PriRV-miR-665-Ov	CCCCAGTGGACCCAGAGCTTAG
PriFW2-miR-495-Ov	CCCTGACGCTCAGTGTCCCTTC
PriRV2-miR-495-Ov	GGGCTCAGCCTCCACATACATA
PriFW-miR-299-OV	CAGACATTGCATAGTGGCCTCT
PriRV-miR-299-OV	ATACAGCGAAATGGTGCCGAAG
PriFW-miR-337-OV	TTGAACCCTTCTCGCTCTCATG
PriRV-miR-337-OV	GTCGACAAGCTTGAGGACATGG
PriFW-miR-665-OV	GAGAGACCGTGGTGCCCTTTGC
PriRV-miR-665-OV	CCCAGTGGACCCAGAGCTTAGG

Supplemental Table 3: Mouse and sheep primer sequences used in RNA editing analysis of miRNA precursors from *DLK1-GTL2* domain.

CHAPITRE V

Discussions et Perspectives

Discussion de l'article « The callipyge mutation enhances bidirectional long-range *DLK1-GTL2* intergenic transcription in cis »

Quel est le mécanisme impliqué dans l'effet en *cis* de la mutation callipyge ?

En dépit de l'identification d'une transition A => G (Freking et al. 2002 ; Smit et al. 2003) affectant un motif conservé de 12 pb comme mutation causale du phénotype callipyge (cf introduction), le mystère demeure quant aux mécanismes moléculaires aboutissant à la surdominance polaire. En effet, comment un simple SNP peut-il influer sur la régulation d'autant de gènes répartis sur plus d'un mégabase ? Si notre hypothèse principale repose sur la destruction d'un élément de contrôle à longue distance par le SNP^{CLPG}, peu de choses sont connues à ce jour sur son mode d'action.

Selon la littérature, les éléments de contrôle à longue distance sont très répandus dans les génomes complexes. Les deux grands modèles pour ce type d'interaction sont les loci des globines alpha et bêta. Tous les paralogues de la bêta globine sont régulés par un seul élément, nommé le LCR (pour Locus Control Region), localisé à 25 kb du gène le plus proche et jusqu'à 80 kb du plus éloigné. Des expériences de 3C (pour Chromosome Conformation Capture, cf Dekker et al., 2002) et de RNA-TRAP (Carter et al. 2002) ont démontré que le LCR interagissait directement avec la région promotrice du paralogue de globine en cours d'expression, et que la zone d'interaction changeait pendant le développement. Cette interaction physique à longue distance (la première jamais démontrée) requiert l'intervention de facteurs de transcription permettant la liaison des deux régions d'ADN. Ainsi, la protéine GATA1 est capable de se lier à la fois au LRC et aux régions promotrices des gènes (Vakoc et al. 2005), tandis que la protéine Ikaros peut changer la cible d'interaction du LCR (Keys et al. 2008). L'interaction à longue distance est donc le fait de nombreuses protéines permettant la jonction de secteurs chromosomiques éloignés et contrôlées finement en fonction du stade de développement cellulaire. Les études d'autres loci ont montré que les interactions à longue distance impliquaient souvent des remodelages conséquents de la chromatine.

Quel est l'élément *cis*-régulateur agissant sur le site de mutation callipyge ?

L'hypothèse principale pour expliquer l'effet en *cis* de la mutation *CLPG* postule que le SNP^{CLPG} détruit un élément de contrôle à longue distance dont la fonction est inhibitrice. La mutation étant située dans un motif conservé de 12 pb, on a d'abord envisagé la destruction d'un site de fixation d'un facteur de transcription. Afin de vérifier cette possibilité, des EMSAs (pour Electrophoretic Mobility Shift Assays) furent réalisées (Maria Smith, Thèse) sur des sondes contenant soit le motif conservé sauvage soit le motif muté *CLPG*. Une protéine (ou un complexe protéique) fut bien détectée sur l'allèle sauvage, dont la migration était retardée comparativement à celle de l'allèle *CLPG*, ce qui semblait valider le modèle d'une perte de liaison consécutive à la mutation *CLPG*. Cependant, ces résultats prometteurs furent pris avec précaution, car les expériences correspondantes avaient été réalisées sur des extraits cellulaires de lignées C2C12 (cellules musculaires murines) et Hela (cellules ovariennes humaines). Dès lors, une validation sur des extraits de cellules musculaires des 4 génotypes callipyges serait souhaitable. Si l'implication de la protéine spécifiquement musculaire MYOD1 a déjà pu être écartée (Maria Smith, Thèse), la (ou les) protéine(s) liée(s) au dodécamère conservé reste(nt) à découvrir.

Quels sont les différents mécanismes induits par la présence du SNP^{CLPG} ?

Bien qu'elle n'identifie pas l'(ou les) agent(s) responsable(s) de l'effet en *cis*, notre étude fournit trois types d'information sur les mécanismes déclenchés par la mutation *CLPG* pour modifier l'expression génique à son voisinage.

1. Le séquençage bisulfite des dinucléotides CpG démontre que si l'allèle sauvage est globalement méthylé (la méthylation de l'ADN est un marqueur d'inhibition de l'expression), l'allèle *CLPG* des animaux C^{mat}/C^{pat} ou $C^{mat}/+$ est pratiquement dépourvu de méthylation, à l'exception de deux CpG consécutifs. Ces deux CpG correspondant au pic de méthylation chez l'allèle sauvage, ils pourraient constituer le site de nucléation de la méthylation dans la région. En empêchant la méthylation de se propager aux alentours, l'allèle *CLPG*

influencerait l'expression génique en *cis*. Fait intéressant, l'allèle *CLPG* des animaux de génotypes *+/C^{pat}* présente un profil de méthylation intermédiaire, laissant donc présumer d'une autre interaction en *trans* impliquant l'allèle maternel. À ce sujet, une étude récente a montré que les miRNAs pouvaient recruter, via le RISC, les protéines CAF1 (Chromatin Assembly Factor 1 subunit) et PABP (Poly(A)-Binding Proteins) et déméthyler ainsi leur site d'interaction (Fabian et al. 2009). Au vu du nombre important de miRNAs maternellement exprimés depuis le locus *DLK1-GTL2,* il est raisonnable d'envisager un rôle pour ces derniers dans les différences de méthylation observées.

2. La découverte de sites de sensibilité à la DNase (DHS) à proximité de la mutation *CLPG* met pour la première fois en évidence l'existence d'un site de contrôle à longue distance (LRCE) dont le DHS constitue la signature (Li et al., 2002). A ce titre, l'hypothèse de l'inactivation par le SNP^{CLPG} d'un LRCE inhibiteur sort renforcée de l'identification de DHS spécifiques à l'allèle *CLPG*. En effet, les DHS témoignent d'une structure de la chromatine plus ouverte et donc plus accessible à la transcription.

3. La découverte que le SNP^{CLPG} active l'expression d'un transcrit couvrant à la fois le motif conservé de la mutation et l'entièreté de la région intergénique (IG) entre *DLK1* et *GTL2* constitue une avancée majeure dans la compréhension des mécanismes de l'effet en *cis* de la mutation. Sa transcription est biallélique sur seulement 13 kb (incluant le SNP^{CLPG}, qui peut par conséquent former des ARNs double brin), mais l'intégralité de la région entre *DLK1* et *GTL2* (soit 90 kb) est transcrite au moins sur l'un des deux brins. Le rôle de ces transcrits IG reste toutefois incertain à ce jour. Jouent-ils un rôle actif dans la genèse du phénotype callipyge ou ne sont-ils qu'une conséquence sans importance de l'effet en *cis* de la mutation ? Deux arguments semblent aller dans le sens de l'effet actif. D'abord, ces transcrits IG connectent physiquement la mutation *CLPG* à deux gènes majeurs du locus (*DLK1* et *GTL2*) dont ils pourraient directement affecter l'expression. Ensuite, l'observation que ces transcrits IG sont plus exprimés deux semaines avant la naissance qu'à l'âge de huit semaines (où le phénotype d'hypertrophie musculaire devient observable) suggère qu'ils ont un rôle majeur dans la mise en place du phénotype callipyge.

Discussion de l'article «RNAi-Mediated Allelic trans-Interaction at the imprinted RTL1/PEG11 Locus»

Confirmation de l'existence d'une *trans*-inhibition dans le domaine *DLK1-GTL2*

L'hypothèse de l'existence d'une *trans*-inhibition de l'allèle paternel par l'allèle maternel dans le domaine *DLK1-GTL2* (cf Introduction) avait été évoquée dès 1996 par Cockett et al. avant d'être strictement définie par Georges et al. (2003), sans pourtant avoir jamais été démontrée expérimentalement.

La détection par extension d'amorce de miRNAs issus du transcrit non codant *anti-PEG11*, exprimés dans les tissus musculaires et affectés par l'effet en *cis* de la mutation *CLPG* (plus abondants chez les animaux de génotype $+/C^{pat}$ et C^{mat}/C^{pat}) a fourni une liste de candidats idéaux pour évaluer l'hypothèse d'une *trans*-interaction. Des expériences de RLM 5' RACE (RNA Ligase-Mediated Rapid Amplification of cDNA Ends) sur un animal C^{mat}/C^{pat} (surexprimant à la fois *PEG11* et *anti-PEG11*) nous ont ainsi permis de démontrer l'existence d'une série de sites de clivage sur *PEG11* correspondant aux différents sites de coupure prédits pour les miRNAs d'*anti-PEG11*.

L'hypothèse d'une trans-interaction était donc vérifiée : un transcrit du domaine DLK1-GTL2, maternellement exprimé et affecté par l'effet en cis de la mutation CLPG, hébergeait quatre miRNAs dont les séquences étaient parfaitement complémentaires à celle du gène paternel PEG11.

Premières interactions de miRNAs présentant une complémentarité parfaite avec leurs cibles observées chez un mammifère.

Cette publication rapportait en outre le premier exemple d'interaction parfaite entre un miRNA et sa cible impliquant des gènes soumis à l'empreinte chez un mammifère. En effet, si le clivage du messager est un procédé très courant chez les plantes, les mammifères semblent ne recourir que très rarement à ce mécanisme, lui préférant généralement l'inhibition post-transcriptionnelle (cf Introduction sur les miRNAs). Seuls deux autres cas de

clivage de la cible d'un miRNA sont connus à ce jour chez les mammifères. Le premier, découvert par Yekta *et al.* (2004) avant d'être confirmé par Mansfields *et al.* (2004), met en jeu miR-196, un miRNA de l'homéodomaine *HOX*. Les complexes *HOX* sont des clusters de gènes (au nombre de quatre chez les mammifères, *HOX* A à D, regroupant un total de 39 gènes) au rôle crucial dans le développement embryonnaire. MiR-196a, transcrit à la fois à partir des clusters Hox B et Hox C, génère une interaction parfaitement complémentaire avec le gène *HOX8B*, à l'exception d'un seul appariement de type Wobble (G-U). Par l'intermédiaire d'un système rapporteur utilisant la luciférase, il a été démontré que le clivage de la cible avait bien lieu et que l'appariement Wobble n'avait pas d'influence par rapport à un contrôle à interaction parfaite. Fait intéressant, miR-196 inhibe également d'autres gènes HOX, mais cette fois par interaction imparfaite au niveau de la 3' UTR. Le second cas implique le virus EBV (Epstein-Barr virus). Ce dernier exprime plusieurs miRNAs dont l'un, miR-BART2, est issu du transcrit antisens de l'ARNm *BALF5,* un facteur de transcription encodé par le virus lui-même (Davis et al. 2004 ; Sullivan et al. 2005). Tout comme *PEG11* et les miRNAs d'a*nti-PEG11*, l'interaction entre miR-BART2 et *BALF5* est parfaitement complémentaire et induit le clivage de la cible.

Très récemment, Shin et al. (2010) ont mis en évidence une nouvelle classe d'interaction entre un miRNA et sa cible fondée sur des « sites centraux ». Ces auteurs ont ainsi montré qu'en l'absence de graine ou de compensation 3', un appariement Watson-Crick de 11 pb en région centrale du miRNA (région 4 à 14 ou 5 à 15 nt) était suffisant pour conduire au clivage du messager cible. Les premiers tests *in vitro* ont révélé que si ces « sites centraux » impliquaient effectivement la protéine Argonaute 2, ils requéraient des concentrations en ions Mg^{2+} bien supérieures à la concentration physiologique. L'étude du dégradome par séquençage haut débit a néanmoins permis de détecter *in vivo* cette forme d'interaction, dans le cerveau, pour 13 messagers différents. En conclusion, même s'il s'agit d'un cas particulier impliquant surtout des éléments répétés (LINE), cette étude ouvre la voie à un nouveau mécanisme de régulation par les miRNAs chez les mammifères.

Hypothèses évolutives au sujet du nombre de miRNAs ciblant *PEG11*

Aucune nouvelle interaction de ce genre n'a été mise au jour à l'heure actuelle, et le clivage du messager reste donc un phénomène très rare chez les mammifères. *Anti-PEG11* ne contient pas moins de quatre miRNAs capables de dégrader le transcrit *PEG11* et l'on peut se demander quelles forces évolutives ont conduit à maintenir autant de régulateurs négatifs. Si l'occurrence de plusieurs miRNAs suggère qu'un seul ne suffirait pas à réguler efficacement *PEG11* (dont la surexpression crée un phénotype létal ; voir Introduction sur *PEG11*), les mécanismes évolutifs ayant permis la conservation de la structure de quatre miRNAs se partageant la même cible reste un mystère. Cela dit, *anti-PEG11* étant l'antisens de *PEG11,* il est possible que la conservation soit pour partie imposée par la séquence codante de la protéine, manifestement soumise à une forte pression de sélection (le KO de *PEG11* étant létal lui aussi). Byrne et al. (2010) ont montré que c'était bien le cas pour au moins deux d'entre eux, miR-431 et miR-433, dont la séquence correspond respectivement à un domaine structurel spiralé et au domaine RNase H de la protéine PEG11. Enfin, ces quatre miRNAs peuvent aussi avoir été conservés en raison de leur action sur d'autres cibles du génome, tout aussi essentielles au développement de l'organisme. Dans cette hypothèse, tous seraient individuellement indispensables à la génération d'un individu sain. Seuls des KOs spécifiques à chaque miRNA d'anti-PEG11 (KOs qui maintiendraient toutefois la structure de la protéine PEG11) permettraient de répondre à cette question.

Les travaux de Song *et al.* (2008) fournissent une autre hypothèse explicative de la présence de quatre précurseurs de miRNAs dans *anti-PEG11*. En effet, ils démontrent que les précurseurs de miR-127 et miR-433 se chevauchent (fait rarissime chez les mammifères) et qu'une seule et même molécule d'*anti-PEG11* ne peut donner naissance aux formes matures des deux miRNAs en même temps. Les auteurs précisent d'ailleurs que ce phénomène pourrait également affecter d'autres miRNAs du locus. Il existerait donc un mécanisme de compétition pour la maturation entre miRNAs et, suivant le tissu ou les conditions physiologiques, l'expression de ces miRNAs pourrait grandement varier. Autrement dit, *PEG11* serait toujours régulé par au moins l'un des miRNAs issus d'anti-PEG11, mais suivant les cas, le miRNA majoritairement exprimé différerait. A ce sujet, nos

propres données confirment que dans le muscle squelettique différencié, miR-127 est très significativement surexprimé par rapport aux trois autres (Caiment *et al.*, 2010).

Le KO de *PEG11* et la théorie du conflit parental

La *trans*-inhibition du transcrit *PEG11* par des miRNAs provenant de l'allèle maternel cadre parfaitement avec la théorie généralement admise pour l'évolution des domaines soumis à l'empreinte parentale : la théorie du conflit (Trivers and Burt 1999; Haig 2000). Celle-ci prédit que lorsque l'expression d'un gène chez un individu commence à avoir des conséquences différentes sur la « fitness » des lignées germinales maternelles et paternelles, l'expression biallélique devient évolutivement instable. Dès lors, l'augmentation d'expression de l'un des allèles est compensée par la réduction de l'autre (pour maintenir un niveau d'expression globalement équivalent à l'expression biallélique), jusqu'à atteindre une expression purement monoallélique, dans laquelle le niveau d'expression idéal est assuré par un seul allèle tandis que l'autre demeure silencieux.

La théorie du conflit est aisément applicable aux gènes impliqués dans le développement. Ainsi, les gènes paternellement exprimés favorisent la croissance des descendants au détriment du bien-être maternel, notamment via des échanges plus importants au niveau du placenta (la majorité des gènes soumis à l'empreinte s'exprimant dans cet organe). Inversement, les gènes exprimés uniquement par l'allèle maternel s'efforcent de réduire la taille des fœtus, afin de préserver les ressources de la mère et d'allonger sa période de reproductibilité. Les premiers gènes soumis à l'empreinte décrits correspondaient totalement à ce modèle (par exemple les gènes *Igf2* et *Igf2r*).

Cette théorie a toutefois été mise à l'épreuve par la découverte de deux loci d'expression paternelle, *Peg1* et *Peg3*, qui influencent positivement chez la souris les soins portés par la mère à ses descendants (construction du nid, placentophagie, récupération des souriceaux). Wilkins & Haig (2003) ont réconcilié ces loci avec la théorie par un modèle d'asymétrie sélective favorisant un plus grand investissement de la mère sur la descendance paternelle. Si cette théorie est aujourd'hui largement acceptée, Wood & Oakey (2006) ont rapporté qu'il semble de plus en plus douteux que ce modèle puisse rendre compte de

l'origine de tous les gènes soumis à l'empreinte. En effet, alors qu'une centaine de gènes soumis à l'empreinte avaient été identifiés jusqu'ici, des analyses récentes sur génome entier par séquençage haut débit ont mis en évidence de nombreux nouveaux loci soumis à l'empreinte uniquement dans certains tissus (on parle de « microimprinted domains »). Ainsi, Gregg et al. (2010), ont identifié 1308 loci de ce type dans le cerveau (824 gènes encodant des protéines et 484 ARNs non codants), la plupart impliqués dans le développement cérébral et l'adhésion cellulaire (et donc sans rapport avec la croissance ou la nutrition). Ces observations, difficilement conciliables avec la théorie du conflit, suggèrent l'implication d'un avantage sélectif conférée par l'empreinte parentale encore inconnu.

En revanche, les transcrits *PEG11* et *anti-PEG11* s'inscrivent quant à eux parfaitement dans le modèle du conflit parental. Leur présence limitée aux génomes euthériens associée à leur puissante expression placentaire, ainsi que les phénotypes induits par la surexpression ou le KO de la protéine (Sekita et al., 2008, cf Introduction sur PEG11), indiquent que *PEG11* promeut l'apport de nutriments du placenta par la mère. *PEG11* est donc bien un transcrit paternel dont le rôle est d'augmenter la taille des descendants, et ce au détriment des ressources maternelles, avec lequel le transcrit maternel *anti-PEG11* entre en conflit par l'intermédiaire de ses miRNAs.

Discussion de l'article «Assessing the effect of the CLPG mutation on the microRNA catalogue of skeletal muscle using high throughput sequencing»

Quel est le mécanisme impliqué dans la *trans*-inhibition induit par la mutation callipyge sur *DLK1* ?

Si un effet en *trans* de la mutation *CLPG* a été démontré dans le locus pour *PEG11*, qu'en est-il du second messager paternel : *DLK1* ? Une *trans*-inhibition du messager *DLK1* impliquant des transcrits maternels est en effet envisagée afin d'expliquer le phénotype sauvage des animaux de génotype C^{mat}/C^{pat} (Georges *et al.*, 2004 ; cf Introduction sur l'effet en *trans* de la mutation *CLPG*). Pour rappel, les animaux C^{mat}/C^{pat} se caractérisent par une expression ectopique du messager *DLK1* associée à l'absence de la protéine correspondante

dans les fibres musculaires (Davis et al. 2004). Le candidat le plus évident pour expliquer cette inhibition de la traduction est l'un des nombreux miRNAs du domaine *DLK1-GTL2*.

Avant notre travail, les miRNAs identifiés dans le domaine ovin l'avaient pour la plupart été soit par comparaison avec leurs orthologues murins ou humains (repris dans miRBase), soit par prédiction *in silico* basée sur la conservation de la structure en épingle à cheveux caractéristique des précurseurs de miRNAs. Afin de disposer d'un catalogue exhaustif des miRNAs présents dans le tissu musculaire squelettique ovin, ainsi qu'une idée de leur expression huit bibliothèques de petits ARNs furent générées puis séquencées à haut débit sur un Genome Analyzer II (Illumina).

Une solution pour l'analyse des séquences de petits ARNs : miRDeep

Les 50 millions de séquences obtenues (environ 6 millions par bibliothèque) ont ensuite été analysées par miRDeep (Friedlander et al. 2008) en prenant le génome bovin comme référence. Au total, miRDeep a affilié 98,3 % des séquences sur 472 précurseurs de miRNAs potentiels, ne laissant donc que 1,7 % de séquences non assignées. La grande majorité des petits ARNs de nos bibliothèques correspondaient donc à des miRNAs, et ce, dans des proportions supérieures aux observations d'autres auteurs, dont Frielander et al. (2008) pour qui la majorité des séquences de bibliothèques de petits ARNs ne sont pas des miRNAs. De même, Wei et al. (2009) ont obtenu seulement 35 % de miRNAs, le reste des séquences étant constitué d'ARNr, ARNt, snoRNAs, endo-siRNAs, piRNAs et de petits ARNs non annotés (tels que des produits de dégradation de messagers). La proportion inhabituelle de miRNAs dans nos bibliothèques reste pour l'instant inexpliquée, mais pourrait être spécifique au muscle, dont certains miRNAs caractéristiques (notamment miR-1 et miR-206) paraissent très fortement exprimés.

Sur les 472 précurseurs potentiels prédits par miRDeep, 324 correspondent à des miRNAs dont au moins un orthologue est identifiable dans miRBase. On y retrouve les miRNAs communs à tous les tissus (comme la famille des let-7), ainsi que les miRNAs connus pour être spécifiquement musculaires. Nos analyses prédisent donc l'existence de 148

nouveaux précurseurs putatifs. La majorité d'entre eux ne représentent que quelques copies dans nos bibliothèques et pourraient donc correspondre à des miRNAs exprimés dans d'autres tissus (comme le sang, le tissu conjonctif ou les fibres nerveuses, tous présents en faible proportion dans les échantillons de fibres musculaires utilisés pour générer les bibliothèques). Frielander et al. (2008) ont évalué la proportion de faux positifs prédits par miRDeep à environ 1 sur 15. Sur base de cette estimation, nos bibliothèques compteraient plus de 135 nouveaux précurseurs. Il faut cependant noter qu'une même molécule (classée comme un miRNA mature) s'aligne sur plus d'une soixantaine de loci dans le génome. Même si ces loci sont tous capables d'adopter une structure en épingle à cheveux, il est impossible d'affirmer avec certitude qu'il s'agit bien là d'un précurseur de miRNA et non d'un simple élément répété sans lien avec l'interférence ARN. De plus, déterminer quels loci s'expriment véritablement parmi tous les candidats proposés par miRDeep relève de la gageure. Le nombre réel de nouveaux précurseurs est donc vraisemblablement surestimé.

Importance des snoRNAs dans le catalogue de petits ARNs.

Les snoRNAs sont de petits ARNs responsables de la modification des ARNr. Sans entrer dans les détails (pour une revue sur les snoRNAs, cf Bachellerie et al., 2002), il existe deux grandes classes de snoRNAs. Les C/D snoRNAs comportent deux motifs conservés, les boîtes C (UGAUGA) et D (CUGA), et contiennent une région (entre les nucléotides 10 à 21), capable de reconnaître un ARNr cible (par complémentarité de séquence) et de procéder à sa méthylation. La seconde classe, les H/ACA snoRNAs, aux motifs H (consensus ANANNA) et ACA (consensus ACA), forment une structure composée de deux épingles à cheveux et dirigent la pseudouridylation des ARNr dont ils sont complémentaires.

Ender et al. (2008) ont démontré que certains H/ACA snoRNAs pouvaient être clivés par Dicer en petits ARNs (d'environ 22 nucléotides), pour lequels ces auteurs ont en outre démontré une action identique à celle des miRNAs sur la traduction de gènes cibles. Notre découverte de séquences du cluster de C/D snoRNAs du domaine *DLK1-GTL2* compatibles avec la taille de miRNAs suggère que cette deuxième classe de snoRNAs pourrait également

avoir la double fonction de modifier certains ARNr tout en générant des miRNAs ciblant d'autres gènes. Pour valider cette hypothèse, il faudrait démontrer, tout comme Ender et al., que la génération des snoRNAs du cluster est indépendante du clivage par Drosha. Cependant, les C/D snoRNAs du domaine sont dits « orphelins » (leurs cibles ARNr sont inconnues) tandis qu'aucune véritable activité de méthylation n'y a jamais été associée. Il reste donc possible que ces C/D snoRNAs ne soient que des miRNAs mal classés, et ce d'autant plus que la conservation de leurs boîtes C et D est assez faible.

Enfin, il est à noter qu'aucun des petits ARNs issus des C/D snoRNAs du cluster de *DLK1-GTL2* ne présente de graine complémentaire à la 3' UTR (ni à la région codante) de DLK1, si bien que ces derniers ne semblent pas directement impliqués dans l'établissement de l'effet en *trans* de la mutation *CLPG*.

Expression relative et absolue des miRNAs du domaine

Bien qu'ils soient connus comme spécifiquement musculaires, miR-1 et miR-206 (responsables entre autres de l'hypertrophie musculaire du mouton Texel, cf Clop et al. en annexe), rien dans la littérature ne laissait présager qu'ils représenteraient à eux seuls plus de 90 % des miRNAs exprimés dans nos bibliothèques Illumina. Afin de confirmer les données de séquençage haut débit, nous avons réalisé, sur les mêmes échantillons musculaires des quatre génotypes CLPG, des études d'expression utilisant soit des puces LNA d'Exiqon (569 sondes humaines, dont 265 exprimées dans les bibliothèques Illumina), soit huit sondes de QRT-PCR en système Taqman (Applied Biosystems).

L'expression relative des différents miRNAs, c'est-à-dire le niveau d'expression (en nombre de séquences pour Illumina, en intensité de fluorescence pour Exiqon et en *Ct* pour la QRT-PCR) d'un miRNA donné au travers des différents génotypes callipyges s'est avérée reproductible avec les trois méthodes. De plus, les miRNAs issus du locus *DLK1-GTL2* sont affectés par l'effet en *cis* de la mutation *CLPG* et présentent un profil d'expression similaire aux transcrits maternels non codants du locus (les C^{mat}/C^{pat} étant les plus exprimés, suivis des $C^{mat}/+^{pat}$, des $+^{mat}/C^{pat}$ et enfin des $+^{mat}/+^{pat}$). On note cependant que l'amplitude de l'effet en

cis est supérieure sur les longs ncRNAs. Ainsi, si *GTL2* augmente d'environ 40 fois chez un animal C^{mat}/C^{pat} en comparaison d'un mouton de type sauvage, l'augmentation d'expression des miRNAs ne dépasse pas un facteur 16. Les raisons sous-tendant cette disparité sont pour l'instant inconnues, mais elle a été observée avec les trois méthodes utilisées ; il est donc peu probable que ce soit un artefact. Ainsi, il pourrait s'agir d'une réponse différentielle des promoteurs aux agents responsables de l'effet en *cis* de la mutation. Une autre possibilité serait d'évoquer une saturation de la machinerie de maturation des miRNAs (principalement *Drosha* et *Dicer*).

L'expression absolue, c'est-à-dire l'expression d'un miRNA donné par rapport aux autres au sein du même génotype, est infiniment plus délicate à analyser. En effet, nos résultats indiquent qu'il est impossible de corréler les trois méthodes entre elles. Un exemple flagrant est miR-1, dont le niveau d'expression en Illumina est 30 fois supérieure au deuxième de la liste (miR-206), mais qui le dépasse à peine en QRT-PCR et se voit carrément supplanté par miR-381 dans le génotype C^{mat}/C^{pat} en Exiqon (miR-381 étant lui-même très faiblement exprimé dans les bibliothèques Illumina).

Dans ces conditions, trancher quant à la méthode la plus proche de la réalité est très difficile. En effet, si le séquençage haut débit fournit un décompte précis du nombre de molécules présentes dans l'échantillon analysé, la génération des bibliothèques de petits ARNs comprend des étapes de ligation d'adaptateurs, de synthèse de cDNA et d'amplification par PCR (15 cycles), soit autant de sources de biais susceptibles d'influer sur le résultat final. En phase avec nos observations, Linsen et al. (2009) ont étudié trois méthodes de génération de bibliothèque et montré que chacune d'elles présentait un biais conséquent. Ainsi, l'expression de 48 % des miRNAs estimée en Illumina varie de plus d'un ordre de grandeur d'une méthode à l'autre. Cependant, ces variations sont reproductibles pour chaque méthode et permettent donc l'étude de l'expression relative. Les puces LNA d'Exiqon ont elles le désavantage, comme toutes les techniques basées sur des sondes, de pouvoir saturer. Ainsi l'excès de miR-1 pourrait être masqué par l'effet limitant de la

quantité de sondes. Quant aux sondes Taqman, à l'instar de tout système QRT-PCR, elles présentent des efficacités dépendant des amorces utilisées.

Il faut cependant mettre au crédit du séquençage haut débit l'avantage de se restreindre aux formes matures des miRNAs. Or, malgré l'hébergement par un seul transcrit (*MIRG*) de très nombreux précurseurs de miRNAs, nos résultats indiquent que les formes matures correspondantes présentent des profils d'expression très variés, ce qui suggère que l'efficacité de maturation et/ou la stabilité des précurseurs varient grandement selon les miRNAs. En utilisant de l'ARN total, les puces d'Exiqon et les sondes Taqman risquent dès lors de comptabiliser des molécules sans véritable effet physiologique. En conclusion, le séquençage haut débit reste à ce jour la meilleure technique pour l'étude des miRNAs, surtout s'il on considère que cette méthode permet en outre (i) une discrimination parfaite entre familles proches (au nucléotide près), (ii) la détection de tous les isomirs existants et (iii) l'exploration de génomes non séquencés.

Les miRNAs du domaine sont-ils édités ?

Kawahara et al. (2007) ont rapporté que certains miRNAs du domaine *DLK1-GTL2* étaient édités dans les cerveaux murin et humain et que les formes matures des miRNAs affectés pouvaient par conséquent cibler des messagers différents de leurs homologues non-édités (voir Encadré 2 sur l'édition dans l'Introduction). Dans ce travail, nous avons démontré que certains des miRNAs édités dans le cerveau, mais également d'autres précurseurs du locus, étaient édités dans le muscle squelettique de mouton (et non de souris pour une raison inexpliquée). Cependant, alors que Kawahara et al. décrivent un niveau d'édition équivalent entre précurseurs et miRNAs matures, nos propres données indiquent que si le précurseur présente un niveau d'édition élevé, la fraction de bases éditées diminue considérablement au niveau de la forme mature du miRNA (estimée à la fois par RFLP radioactive, par séquençage de miRNAs clonés et par séquençage haut débit). Tout porte donc à croire que, dans le muscle (contrairement au tissu cérébral), un mécanisme empêche la maturation des précurseurs édités, ou bien qu'une possible édition

de novo des miRNAs matures est inhibée. Fait notable, Lim et al. (2005) ont montré que l'une des deux protéines impliquées dans l'édition des ARNs (*ADAR1*) était la cible de miR-1 et miR-206. TargetScan confirme par ailleurs que cette cible est conservée chez les mammifères. Dès lors, l'absence d'édition des miRNAs matures dans le muscle pourrait s'expliquer par l'inhibition de la protéine *ADAR1* par ces miRNAs spécifiquement musculaires.

Une observation très intrigante est l'effet inhibiteur de la mutation CLPG (aussi bien chez les deux hétérozygotes qu'à l'état homozygote) sur le niveau d'édition des précurseurs comparativement au mouton sauvage. Les trois génotypes possédant la mutation *CLPG* présentent une expression plus forte des miRNAs du domaine par rapport au mouton de type sauvage, et pourraient donc interférer avec certains composants de la machinerie de l'édition ARN. Une analyse rapide montre que la 3'UTR bovine du messager d'*ADAR1* possède une graine solide (un octamère) avec miR-432, un miRNA fortement exprimé (plus de 90 000 copies en illumina) situé dans *anti-PEG11*, ainsi que plusieurs interactions plus faibles (trois heptamères, 10 hexamères) avec d'autres. Même si une confirmation au laboratoire serait nécessaire, cette analyse préliminaire laisse à penser que la machinerie de l'édition est inhibée par l'augmentation d'expression induite sur les miRNAs du locus *DLK1-GTL2*.

Quoi qu'il en soit, le niveau d'édition des précurseurs est tellement faible dans les muscles des animaux C^{mat}/C^{pat} qu'il est très probable que l'édition des miRNAs du domaine ne contribue pas à l'établissement de la *trans*-inhibition du messager *DLK1*.

La mutation *CLPG* a-t-elle un effet sur des miRNAs en dehors du locus *DLK1-GTL2* ?

Les analyses statistiques (ANOVA) des données d'expression obtenues par séquençage haut débit, hybridation sur puces Exiqon et QRT-PCR avec sondes Taqman démontrent que la mutation *CLPG* n'influence significativement aucun miRNA extérieur au domaine *DLK1-GTL2*. L'effet direct du SNP callipyge est donc localisé, tandis qu'aucune régulation indirecte de l'expression des miRNAs n'a pu être mise en évidence. En revanche, la modification du profil d'expression des nombreux miRNAs du locus doit avoir

d'importantes répercussions sur les différents gènes musculaires. L'étude ontologique des gènes ciblés par l'ensemble de ces miRNAs (TargetScan + GO terms) montre un biais envers les régulateurs transcriptionnels, traductionnels et post-traductionnels (principalement dans le système nerveux).

Le séquençage haut débit a-t-il permis de découvrir de nouveaux miRNAs ?

Si miRDeep prédit plus 148 nouveaux précurseurs potentiels, il n'est pas réaliste d'en vérifier individuellement l'existence. Cependant, une étude détaillée du domaine *DLK1-GTL2* nous a permis d'y dénombrer 10 nouveaux miRNAs. Étonnamment, l'un d'eux (miR-3958) représentait à lui seul ~ 75 % de toutes les molécules matures issues du domaine et se plaçait au sixième rang du classement général ; pourtant il n'était pas encore référencé dans miRBase. L'étude de la distribution taxonomique de miR-3958 (données non montrées) a révélé que sa présence se limitait aux génomes des Laurasiatheria, un clade des mammifères placentaires regroupant les carnivores (félins et canins), les ongulés (dont les ruminants), les cétacés et les chauves-souris. A priori, cette découverte contredit donc l'hypothèse de Chen & Rajewsky (2007) selon laquelle un miRNA apparu récemment présenterait un niveau d'expression inférieur aux miRNAs plus anciens.

Cependant, si les données Illumina suggèrent effectivement un niveau d'expression très important pour miR-3958, la QRT-PCR avec sondes Taqman ne confirme pas ce résultat. En effet, si miR-3958 est ~ 35 fois plus exprimé que miR-3959 en séquençage haut débit, la PCR en temps réel donne un niveau d'expression équivalent pour les deux miRNAs (niveau très inférieur à celui enregistré en Illumina). Il est donc difficile de conclure sur la véritable expression de miR-3958 dans le tissu musculaire. Sa parfaite conservation dans un clade important (données non montrées) appuie toutefois l'hypothèse d'une fonction *in vivo*.

Un miRNA est-il responsable de la *trans*-inhibition de *DLK1* ?

Même si l'expression des miRNAs du locus est difficile à quantifier précisément, le séquençage haut débit a néanmoins produit un catalogue exhaustif de miRNAs susceptible de mener à l'identification d'un potentiel inhibiteur de *DLK1*. De façon un peu décevante,

aucun de ces miRNAs ne peut former une graine canonique (score Grimson) avec la 3' UTR de *DLK1*, tandis que miRanda (qui admet des graines imparfaites) n'identifie que le seul miR-376c comme candidat. En autorisant la formation de graines avec la région codante du gène, de meilleurs candidats émergent (notamment des miRNAs issus d'a*nti-PEG11*), mais aucun ne passe le seuil de significativité corrigé pour tests multiples. Le candidat idéal, du moins si l'on s'en tient aux règles canoniques décrites pour les interactions miRNA–cibles, n'existe donc pas, et ce, que l'on applique ou non le critère de conservation entre espèces. À noter qu'il existe une forme mineure (« isomir ») d'un miRNA du locus, miR-3957, capable de former une graine canonique de sept nucléotides consécutifs avec la 3' UTR de *DLK1*, mais son expression est tellement infime (moins de 10 séquences sur les 6 millions de l'individu *C/C*) que nous considérons que son effet sera négligeable *in vivo*. Un autre miRNA, miR-15a, a été validé comme inhibiteur de *DKL1* par Andersen et al. (2010) dans les pré-adipocytes humains. Cependant, si ce miRNA est effectivement présent dans nos bibliothèques de petits ARNs (~ 1000 copies), son expression n'est pas affectée par l'effet en *cis* de la mutation *CLPG* et sa cible sur *DLK1* n'est pas conservée chez le mouton.

Malgré cela, les indices suggérant un effet inhibiteur chez l'animal *C/C* des miRNAs du domaine sur le transcrit *DLK1* existent bel et bien : (i) légère diminution du niveau du messager, (ii) importante réduction de l'abondance de la protéine, (iii) pléthore de miRNAs soumis à l'effet en *cis* de la mutation et exprimés maternellement. Dès lors, l'hypothèse initiale d'un *trans*-inhibiteur miRNA ne peut pas être exclue. Ainsi, nos analyses indiquent que l'effet cumulé de tous les miRNAs du locus en tant qu'équipe est significatif sur la région codante du gène *DLK1*. Hélas, les systèmes rapporteurs tels que la luciférase ne permettent pas de tester l'effet d'autant de miRNAs simultanément avec efficacité.

Une autre possibilité consisterait à envisager que les miRNAs du locus soient responsables de l'inhibition de *DLK1*, mais selon un mode d'interaction non canonique. En effet, plusieurs publications récentes ont rapporté des exemples d'inhibition par des miRNAs sans formation de graine parfaitement complémentaire à la 3' UTR. Ainsi, Tay et al. (2008) ont notamment décrit une inhibition de ~ 60 % de la protéine NANOG par miR-470, et ce en

dépit d'une interaction ciblant la région codante du gène et de deux appariements de type Wobble (G : U) au milieu de la « graine ». En relaxant de la sorte la définition de l'interaction miRNA – cible, il nous serait aisé de trouver de très nombreux « candidats » du domaine *DLK1-GTL2* capables d'interagir avec *DLK1*. Tay et al. n'exploitent qu'un sous-ensemble des interactions prédites (via le logiciel RNAhybrid) et utilisent des western blots pour les confirmer. De notre côté, le trop grand nombre de candidats et l'absence d'anticorps contre la forme ovine de DLK1 nous empêchent de poursuivre dans cette voie.

Enfin, un nouveau type d'interaction a été décrit par Lee et al. (2009), dans lequel la région 3' du miRNA peut s'apparier avec des régions complémentaires situées dans la 5' UTR du messager. Ces auteurs ont ainsi démontré expérimentalement que miR-34a inhibait le transcrit d'*AXIN2* en s'appariant à la fois à sa 3' UTR (via la graine classique en 5' du miRNA) et à sa 5' UTR (via la région 3' du miRNA), augmentant du même coup l'effet inhibiteur sur la protéine. La définition des éléments clés fondant la « graine 3' » du miRNA étant encore inconnus, il est difficile d'évaluer l'impact potentiel de cette découverte sur l'évaluation des candidats miRNAs du locus *DLK1-GTL2*. Toutefois, ce nouveau type d'interaction miRNA – cible pourrait à l'avenir s'avérer intéressant à explorer.

Et si ce n'était pas seulement un miRNA ?

Même si l'interaction directe entre un miRNA du domaine et *DLK1* reste l'hypothèse la plus parcimonieuse, d'autres modes d'action sont envisageables. On pourrait ainsi postuler l'existence d'une interaction indirecte entre un (ou plusieurs) miRNA(s) du locus et un activateur de *DLK1* (ou une chaine de transduction aboutissant à l'activation de *DLK1*), dont le rôle serait d'augmenter la traduction du messager et/ou de stabiliser la protéine. Dans ce contexte, la surexpression des miRNAs maternels chez l'individu C^{mat}/C^{pat} aurait pour effet de faire disparaître cet activateur, et donc aussi la protéine DLK1 en dépit de la présence continue de son messager. Cependant, le grand nombre de candidats (plus de 90 miRNAs) issus du locus et l'absence d'un catalogue complet des 3' UTR ovines rendent cette analyse impraticable à l'heure actuelle.

Hormis les nombreux miRNAs (et autres snoRNAs) du locus, les longs transcrits non-codants (ncRNAs) *GTL2*, *MEG8* et *MIRG* (cf Introduction) sont eux aussi maternellement exprimés et affectés par l'effet en *cis* de la mutation. Puisqu'aucune fonction ne leur est pour l'instant attribuée, ils pourraient en principe être impliqués dans la surdominance polaire. La preuve que ce genre d'interaction existe est fournie par le plus étudié des domaines soumis à l'empreinte, *H19* et *IGF2* (Insulin-like Growth Factor 2). Runge et al. (2000) y ont rapporté que le transcrit maternel non codant *H19* pouvait se lier à la protéine IMP-2 *(IGF2* mRNA-binding Protein 2), qui comme son nom l'indique, interagit elle-même avec *IGF2*. En se liant aux deux transcrits réciproquement soumis à l'empreinte, IMP-II pourrait ainsi inhiber la traduction d'*IGF2*. Appliqué à la surdominance polaire, ce modèle reviendrait à imaginer que l'un des trois longs transcrits non codants maternels fonctionne en tant que *trans*-inhibiteur par liaison avec une protéine interagissant avec *DLK1*.

Récemment, Elcheva et al. (2009) ont fourni encore un autre mode d'interaction possible impliquant à la fois les miRNAs et des facteurs protéiques régulateurs. Ils ont en effet rapporté que la traduction de l'ARNm *BTRCP1* était inhibée par miR-183, via appariement à la région codante (en l'absence de graine parfaite) puis recrutement du RISC (dont Argonaute 2). Au delà de la découverte d'un type supplémentaire d'interaction non canonique entre un miRNA et sa cible, cette étude démontre que cette inhibition de la traduction peut être annulée par interaction avec la protéine CRD-BP (c-myc mRNA coding Region Determinant-Binding Protein), dont le rôle est ici d'empêcher la liaison efficace d'Argonaute 2. Dans le cas du locus *CLPG*, la surexpression des miRNAs chez les animaux C^{mat}/C^{pat} contrerait l'effet protecteur d'une protéine stabilisatrice du transcrit *DLK1*. Cependant, la vérification de l'existence de ce type de mécanisme est très délicate. Bien peu de données sont disponibles pour caractériser les interactions entre miRNAs et régions codantes, tandis que l'identification d'un facteur protéique ciblant *DLK1* représenterait un lourd investissement.

Enfin, si DLK1 reste notre principal candidat pour expliquer l'hypertrophie musculaire du mouton callipyge, son implication n'est pas incontestable d'après la littérature. Ainsi, des modèles murins exprimant DLK1 à un niveau deux ou trois fois supérieur à la normale dans tous les tissus montrent des effets sur le squelette, mais aucun sur le phénotype musculaire embryonnaire (da Rocha et al., 2009). Même si cette observation est difficilement transposable à la surexpression musculaire post-natale de DLK1 à l'œuvre dans le modèle ovin, nous continuons à envisager la possibilité qu'un autre transcrit soit, du moins partiellement, impliqué dans l'hypertrophie musculaire.

Perspectives

Existe-t-il un modèle murin récapitulant la surdominance polaire observée chez le mouton callipyge ?

Le mouton est loin d'être un modèle idéal pour des études génétiques, que ce soit par son besoin de structures animalières importantes, sa lenteur de génération ou l'absence d'outils moléculaires adaptés. A l'inverse, un modèle murin récapitulant la surdominance polaire observée chez les moutons callipyges faciliterait l'accès fréquent aux tissus, accélérerait les croisements et permettrait l'utilisation d'outils d'analyse plus variés (tels que les anticorps). C'est pourquoi des souris transgéniques porteuses de la même mutation que le mouton callipyge (lignée A->G) ont été générées par recombinaison homologue, sélection des clones de cellules ES, injection dans des blastocystes et implantation dans des femelles porteuses. De plus, afin de tester si la mutation *CLPG* correspondait à une perte ou à un gain de fonction, une seconde lignée caractérisée par la délétion complète du dodécamère conservé a également été générée (lignée Δ12). L'idée était que si la lignée Δ12 récapitulait l'effet de la lignée A->G, le SNP^{CLPG} correspondrait alors à une perte de fonction d'un motif régulateur (comme le prévoit notre hypothèse de travail).

Les premières analyses ont tout d'abord laissé présager que ces deux lignées récapitulaient le phénotype callipyge. En effet, les transcrits du domaine *DLK1-GTL2* y conservaient leur statut d'empreinte génétique, tandis que des expériences de QRT-PCR

démontraient que les transcrits maternels (*GTL2*, *anti-PEG11* et *MIRG*) étaient affectés par l'effet en *cis* de la mutation, sans différence notable entre les deux lignées. Cependant, il est ensuite apparu que l'expression des transcrits paternels *DLK1* et *PEG11* n'était affectée chez aucun des différents génotypes des deux mutations. Or, en l'absence d'expression ectopique de la protéine DLK1 dans les muscles squelettiques après la naissance, une hypertrophie musculaire avait peu de chance de se manifester, ce qui a finalement été confirmé par l'analyse phénotypique.

Même si ces deux lignées n'ont pas encore livré tous leurs secrets, il semble que le modèle murin ne récapitule pas tous les critères du phénotype callipyge observé chez le mouton. Loin d'être l'outil espéré, ces souris transgéniques ont au contraire suscité de nouvelles questions : (i) Quels mécanismes permettent l'effet de la mutation sur les transcrits maternels sans affecter les transcrits paternels ? (ii) Quelles différences peuvent expliquer la différence d'impact de la mutation callipyge chez la souris et le mouton ? L'étude détaillée des lignées transgéniques devra être envisagée si l'on veut un jour y répondre.

La protéine PEG11 joue-t-elle un rôle dans la genèse du phénotype callipyge ?

Dans le domaine *DLK1-GTL2*, seuls deux gènes paternels sont soumis à l'effet en *cis* de la mutation (et sont donc surexprimés dans le muscle après la naissance dans le génotype $+/C^{pat}$). Si un effet de l'expression ectopique de DLK1 sur le développement musculaire a été démontré dans le modèle murin (Davis et al. 2004), l'implication de PEG11 ne peut être exclue. En effet, si l'importance de cette protéine n'est plus à démontrer (létalité tant de sa délétion que de sa surexpression durant le stade embryonnaire), son rôle reste exact mystérieux et ses effets potentiels sur la cellule musculaire squelettique inconnus.

Afin d'étudier la question, une souris transgénique porteuse du cadre ouvert de lecture du *PEG11* ovin sous le contrôle du promoteur de la chaîne légère de la myosine et de l'enhancer 2E a été générée. Ce système est identique à celui de la souris *DLK1* et permet une expression ectopique dans les cellules musculaires squelettiques. Sur les 23 fondateurs issus de micro-injections pronucléaires, l'expression du transgène a pu être vérifiée par RT-PCR dans la génération F1 pour 13 lignées différentes. Parmi celles-ci, neuf portaient un tag

histidine C-terminal (aucun anticorps n'étant disponible pour détecter la protéine PEG11 ovine) et quatre n'en contenaient pas. La génération F2 a conduit à l'obtention d'homozygotes transgéniques. Jusqu'ici, aucun phénotype flagrant n'a été observé : ni létalité, ni trouble majeur, ni hypertrophie musculaire importante. Toutefois, une étude phénotypique rigoureuse (incluant des coupes histologiques) est en cours pour conclure quant au rôle potentiel de PEG11 dans la genèse du phénotype callipyge.

Pour aller plus loin sur l'hypothèse des miRNAs acteurs de l'effet en *trans*.

L'hypothèse des miRNAs restant la plus parcimonieuse à nos yeux, deux expériences importantes sont actuellement menées pour tenter de valider leur implication dans l'établissement de la surdominance polaire.

Premièrement, plusieurs auteurs (Nelson et al. 2004 ; Vasudevan et al. 2007 ; Takeda et al. 2010) ont démontré qu'il était possible d'utiliser un anticorps dirigé contre l'une des protéines du RISC (souvent une protéine Argonaute) afin d'immunoprécipiter les miRNAs et leurs cibles. Après transfection d'un miRNA candidat, on peut ainsi étudier l'enrichissement en un messager donné par RT-PCR sur le produit d'immunoprecipitation. En supposant que DLK1 est bien la cible d'un (ou plusieurs) miRNA(s), son messager devrait se révéler davantage représenté chez les animaux de génotype C^{mat}/C^{pat} que chez les animaux $+^{mat}/C^{pat}$. Cette approche pourrait donc nous fournir une preuve indirecte du fait que les miRNAs ciblent DLK1, et ce, indépendamment de leur mode d'appariement. Aujourd'hui, il est en outre envisageable d'associer cette immunoprécipitation aux technologies de séquençage haut débit. Appelée « HITS-CLIP » (pour HIgh-Throughput Sequencing of RNAs from in vivo Cross-Linking and Immuno-Precipitation, Darnel et al., 2010), cette variante de la méthode permet d'identifier simultanément tous les sites d'interaction miRNAs – cibles sur tous les transcrits. Actuellement testée par Haruko Takeda, cette technique devrait à la fois trancher la question de savoir si DLK1 est bien la cible préférentielle de miRNAs, tout en donnant des indications sur la régulation de tous les autres transcrits dont l'expression est modifiée par l'expression ectopique des nombreux miRNAs du locus induite par la mutation CLPG.

La seconde expérience en cours, plus directe, est basée sur des transfections en modèle cellulaire. Partant de l'hypothèse selon laquelle des miRNAs du locus pourraient cibler *DLK1* selon un mode d'appariement non canonique, le recours à des miRNAs de synthèse devrait permettre de reproduire un possible effet inhibiteur affectant son transcrit. Puisque les données d'expression de nos bibliothèques de petits ARNs sont difficilement exploitables d'un point de vue quantitatif, Hui Jun Cheng prévoit de transfecter tous les miRNAs du domaine *DLK1-GTL2* (soit plus de 50 précurseurs potentiels) en présence du transcrit *DLK1* complet. L'effet inhibiteur sera alors évalué par western blot et test ELISA sur la protéine DLK1 native (via des anticorps commerciaux) et sur une version taggée.

Et si ce n'est pas les miRNAs...

Si les expériences de HITS-CLIP et les transfections en modèle cellulaire devaient conclure au rejet des miRNAs comme *trans*-inhibiteurs de *DLK1*, l'étude d'une implication directe des longs ARNs non codants (ncRNAs) du locus sera envisagée. Pour ce faire, des expériences d'hybridation *in situ* (FISH) seront réalisées sur *DLK1* et sur les trois ncRNAs (*GTL2, MEG8* et *MIRG)* en coupe musculaire de mouton C^{mat}/C^{pat} afin de vérifier la colocalisation des deux transcrits. En effet, pour être de bons candidats à la *trans*-inhibition, les transcrits non-codants du domaine devraient présenter un profil d'expression compatible avec celui de *DLK1*. Ensuite, il sera nécessaire d'identifier un facteur protéique liant *DLK1* et un des ncRNAs par des expériences de cross-linking UV (cf Pashev et al., 1991 pour revue) sur des extraits protéiques cytoplasmiques de tissu musculaire squelettique ovin. Si une telle protéine est identifiée, il faudra la purifier puis obtenir sa séquence peptidique par spectrométrie de masse. Dans l'hypothèse d'une issue positive, le gène correspondant serait alors un bon candidat pour l'effet en *trans* de la mutation *CLPG*.

CHAPITRE VI

Annexes :

A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep

Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibe B, Bouix J, Caiment F, Elsen JM, Eychenne F et al. 2006. Nature Genetics 38(7): 813-818.

A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep

Alex Clop[1, 6], Fabienne Marcq[1, 6], Haruko Takeda[1, 6], Dimitri Pirottin[1, 6], Xavier Tordoir[1], Bernard Bibé[2], Jacques Bouix[2], Florian Caiment[1], Jean-Michel Elsen[2], Francis Eychenne[2], Catherine Larzul[2], Elisabeth Laville[3], Françoise Meish[1], Dragan Milenkovic[4], James Tobin[5], Carole Charlier[1]& Michel Georges[1]

[1] Unit of Animal Genomics, Department of Animal Production, Faculty of Veterinary Medicine & Centre for Biomedical Integrative Genoproteomics, University of Liège (B43), 20 Boulevard de Colonster, 4000 Liège, Belgium.

[2] Institut National de la Recherche Agronomique–Station d'Amélioration Génétique des Animaux (INRA-SAGA), BP 52627, 31326 Castanet-Tolosan CEDEX, France.

[3] Station de Recherches sur la Viande, INRA, Theix, 63122 Saint-genès-Champanelle, France.

[4] INRA/Université de Limoges, Faculté des Sciences, 87060 Limoges Cedex, France.

[5] Cardiovascular and Metabolic Diseases, Wyeth Research, 87 Cambridge Park Drive, Cambridge, Massachusetts 02140, USA.

[6] *These authors contributed equally to this work.*

SUMMARY

Texel sheep are renowned for their exceptional meatiness. To identify the genes underlying this economically important feature, we performed a whole-genome scan in a Romanov × Texel F2 population. We mapped a quantitative trait locus with a major effect on muscle mass to chromosome 2 and subsequently fine-mapped it to a chromosome interval encompassing the myostatin (*GDF8*) gene. We herein demonstrate that the *GDF8* allele of Texel sheep is characterized by a G to A transition in the 3' UTR that creates a target site for miR-1 and miR-206, microRNAs (miRNAs) that are highly expressed in skeletal muscle. This causes translational inhibition of the myostatin gene and hence contributes to the muscular hypertrophy of Texel sheep. Analysis of SNP databases for humans and mice demonstrates that mutations creating or destroying putative miRNA target sites are abundant and might be important effectors of phenotypic variation.

RESULTS AND DISCUSSION

To locate quantitative trait loci (QTL) underlying the muscular hypertrophy of Texels, we generated a Romanov Texel F2 with 258 offspring. We chose a hypermuscled Belgian strain of Texel for these crosses. We examined 37 phenotypes measuring body composition, muscularity and fat deposition on the F2 animals1, 2. We genotyped 153 microsatellites and scanned the genome by linear regression assuming one QTL per chromosome and fixation of alternate (T and R) QTL alleles in the parental breeds3. A QTL with major effect on muscularity was identified on chromosome 2 (OAR2). For traits yielding genome-wide $P < 0.05$, the QTL accounted for 5–25% of the variance, the difference between alternate homozygotes (2a) ranged from 0.68 to 1.66 residual standard deviations, and the dominance deviation ranged from -0.70a to 0.50a. The QTL explained one-fifth to one-third of the difference between parental breeds (Supplementary Table 1 online). The confidence interval (c.i.) for the QTL spanned 10 cM (Fig. 1a). 'Within-family' analyses showed that the three F1 rams were heterozygous TR (Supplementary Fig. 1 online). A QTL with similar effects and location has been detected in other Texel-based pedigrees4.

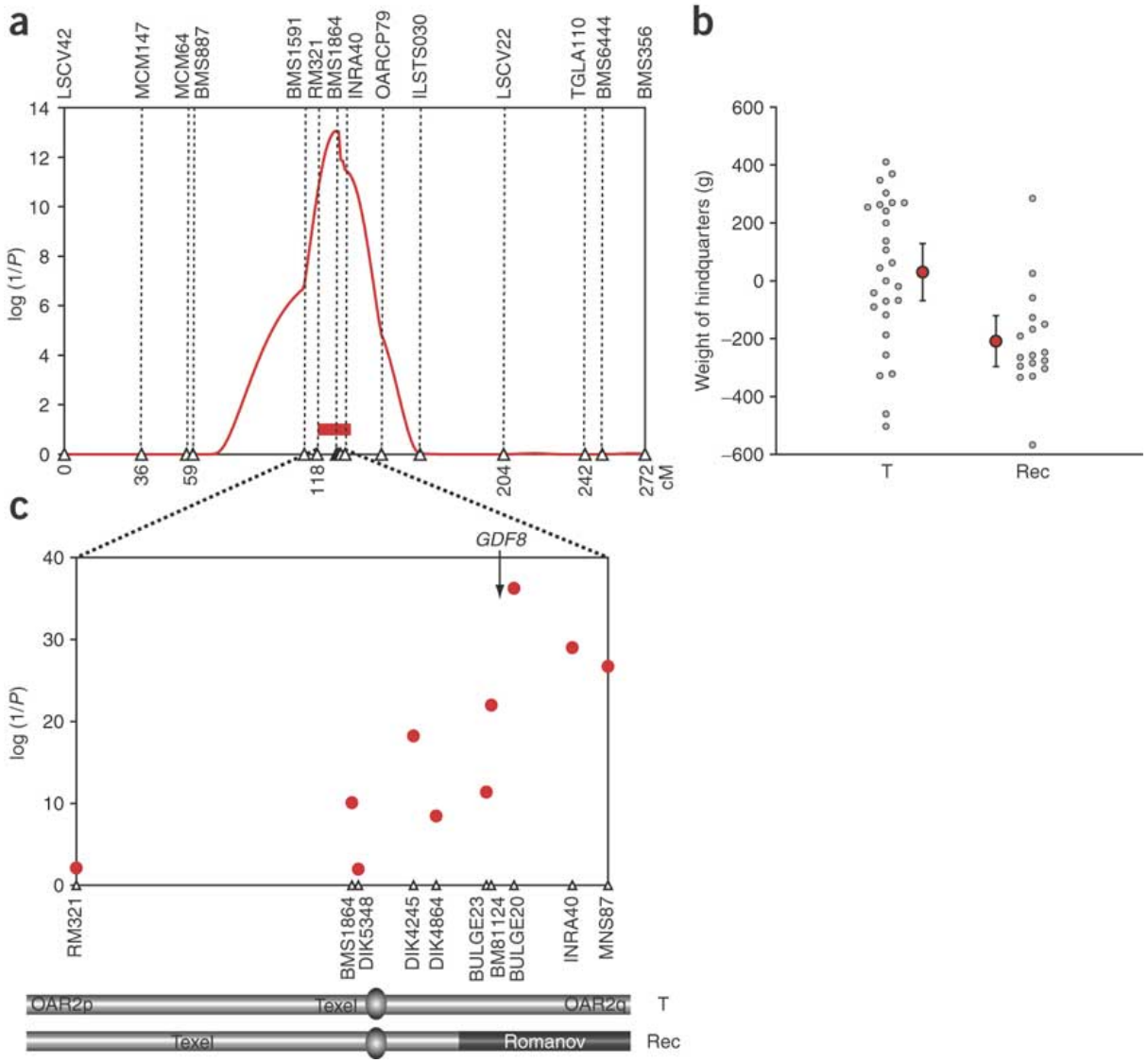


Figure 1. Mapping and fine-mapping of a QTL influencing muscularity on sheep chromosome 2.(**a**) QTL location scores (expressed as log(1/*P*), where *P* corresponds to the genome-wide *P* value of the data under the null hypothesis of no QTL) obtained for 'weight of the hindquarters (in g)' along the marker map of sheep chromosome 2 (OAR2). The red horizontal bar corresponds to the 95% c.i. for the QTL location obtained by bootstrapping. (**b**) Marker-assisted segregation analysis. Gray circles: 'weight of the hindquarters' (in g) (residuals after correction for fixed effects) of offspring, sorted according to the chromosome 2 inherited from their sire: 'T' (full-length Texel chromosome) or 'Rec' (recombinant Texel-Romanov chromosome; the position of the recombination breakpoint is shown in **c**). Red circles represent the group phenotypic mean (± 1.96 s.e.m. of the estimate). The higher variance in the T group is in agreement with the recessive effect of the T allele on weight of the hindquarters (see Supplementary Table 1). (**c**) Results of the DISLAMB[5] linkage disequilibrium analysis in the 95% c.i. for the QTL to identify the effects of a putative selective sweep. Log(1/*P*) corresponds to the logarithm (base 10) of 1/*P*, where *P* is the likelihood of the data under the null hypothesis of no enrichment of a specific marker allele in Texel when compared with controls. The position of the *GDF8* gene is shown by the arrow. The structures of 'T' and 'Rec' chromosomes of the ram used in **b** are schematically represented.

To refine the map position of this QTL we generated a higher-density map of the c.i. In a first approach (marker-assisted segregation analysis (MASA)), we produced 43 offspring

from an F2 ram (20254) that inherited an intact Texel chromosome and a chromosome recombining within the DIK4864–BM81124 interval. The recombinant chromosome was of Romanov descent distal from DIK4864 and of Texel descent proximal from BM81124. The 'weight of the hindquarters' was 238 g heavier in the offspring that inherited the Texel chromosome than in those that inherited the recombinant chromosome (P = 0.0013) (Fig. 1b). This difference was similar to the R to T substitution effect estimated in the F2 animals (a = 203 g). The effects on all other measured traits were very similar to those observed in the F2 animals (Supplementary Fig. 2). This indicated that the ram was TR and the QTL located distal from DIK4864 (Fig. 1b).

In the second approach, we hypothesized that selection for meatiness in Belgian Texel animals might have caused near-fixation of a favorable QTL allele T ('selective sweep') and that most hypermuscled Texels would be homozygous for a segment encompassing the QTL. To test this, we genotyped 42 hypermuscled Texels as well as 108 controls (16 breeds) for ten microsatellites spanning the c.i. We measured the increase in frequency of a given marker allele in the Texels with respect to controls using DISLAMB5. The likelihood ratio test maximized on the OAR2q side of the c.i., in agreement with the MASA (Fig. 1c). The signal peaked at marker BULGE20: one of its 15 alleles had a frequency of 94% in Texels but only 12% in controls; heterozygosity was 11% in Texels versus 85% in controls.

In cattle, the myostatin (*GDF8*) gene is located in the BM81124–BULGE20 interval. *GDF8* loss-of-function mutations cause double-muscling in mice, cattle and humans6, 7, making it an obvious candidate. We sequenced the coding regions of the *GDF8* gene from DNA of three F0 Texel rams and seven controls (five F0 Romanov ewes, one Dorset and one Tarasconnais), but we did not find any polymorphism. RNA blots showed a band of the expected size with comparable intensity in both Texels and controls (Fig. 2a). Real-time quantitative RT-PCR did not demonstrate any significant breed effect on *GDF8* RNA levels (Supplementary Fig. 3 online). We amplified the *GDF8* ORF by RT-PCR on RNA derived from muscle, sequenced the corresponding PCR products and confirmed the identity of the Texel

and control mRNA sequences. At first glance, these results suggested that Texel animals produce normal levels of functional *GDF8* mRNA.

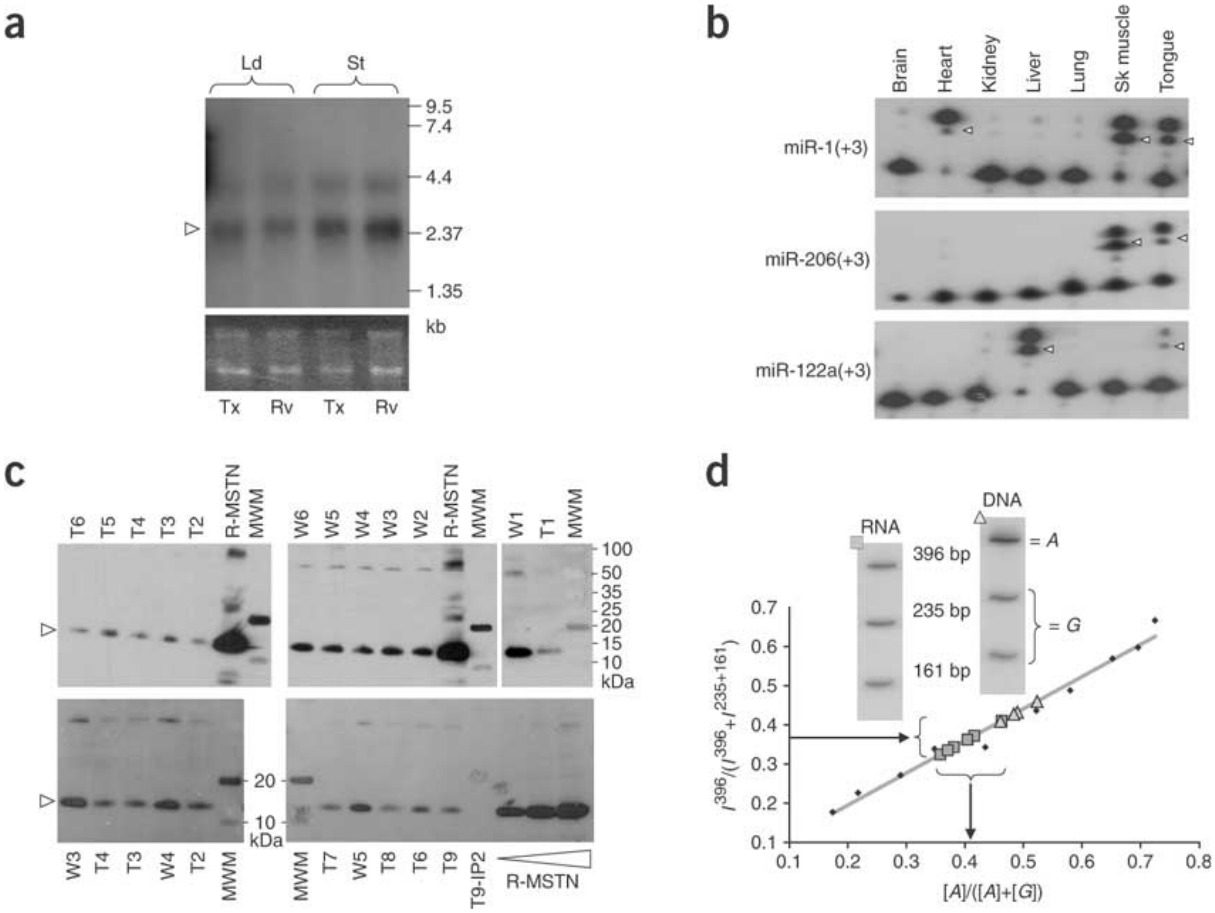


Figure 2. Expression analysis of *GDF8* and potentially interacting miRNAs.(a) RNA blot analysis of *GDF8* expression in longissimus dorsi (Ld) and semitendinosus (St) of Texels (Tx) and wild-type Romanovs (Rv). Arrow: *GDF8* transcript. The ethidium bromide–stained gel is shown. (b) Expression analysis of miR-1, miR-122a and miR-206 in adult sheep tissues. Arrowheads: products of the primer extension. Numbers in parentheses: expected number of nucleotides added by reverse transcriptase, given the sequence of primers and mature miRNA. (c) Detection of circulating myostatin by immunoprecipitation (IP) and protein blotting in Texels (T1–T9) and controls (W1–W6) (matched for sex (male) and age (4 months)). Arrowheads: 12.5-kDa band corresponding to mature myostatin monomer. Except for T1 and W1, all samples were analyzed blindly. T9-IP2: second IP performed on sample T9 showing that the IP depletes myostatin from the sample. R-MSTN: recombinant myostatin11. The three R-MSTN lanes in the lower panel correspond to recombinant myostatin at 25, 62.5 and 312.5 ng/ml. MWM: molecular weight markers. (d) Estimating the relative amounts of A (Texel) versus G (wild-type) transcripts in AG heterozygotes. The y axis shows relative intensity of the 396-bp fragment (A allele) compared with the sum of the intensities of the 396-bp and 235 + 161 bp (G allele) fragments; the x axis predicts the proportion of A molecules. Diamonds: data points obtained by mixing varying amounts of genomic DNA from AA and GG animals, yielding the gray calibration curve. Triangles: results obtained with AG genomic DNA. Squares: results obtained with skeletal muscle RNA from AG heterozygotes.

However, in light of our fine-mapping results, we decided to further examine *GDF8*. We sequenced 10.5 kb spanning the sheep *GDF8* gene from the same three Texel animals

and seven controls. This identified 20 SNPs (Supplementary Fig. 4 online). None of these reside in particularly conserved sequence elements. We genotyped all the SNPs on 42 Texels, 90 controls (11 breeds) and the four TR rams (three F1 animals and F2 20254) (Supplementary Table 2 online).

The first notable observation was the virtual monomorphism of Texels, contrasting with the considerable variation of the 11 other breeds. This would be expected if a *GDF8* mutation had undergone a selective sweep in Texels.

We were able to exclude 18 SNPs because at least one of the four TR rams was homozygous for the SNP. For these SNPs, the Texel allele was at high frequency in the other breeds (0.62; Supplementary Table 2). Two SNPs could not be excluded on the same basis: g-2449C-G, located 2.5 kb upstream from the transcription start site, and g+6723G-A, located in the 3' UTR. For g-2449C-G, the frequency of the C allele was 98% in Texels versus 11.5% in controls. For g+6723G-A, the frequency of the A allele was 99% in Texels versus 1% in controls; thus, the g+6723G-A A allele seemed to be virtually Texel-specific.

G-2449C-G is located in a region devoid of conserved sequence elements. It is difficult to imagine how it might affect myostatin function, especially given the comparable amounts of intact *GDF8* mRNA found in Texels and controls. Moreover, one of the Texel animals was heterozygous only for g-2449C-G (Supplementary Table 2) and homozygous for the Texel allele for all other SNPs. The easiest explanation is that this animal is TT but that it inherited one T haplotype with a recombination just upstream of the *GDF8* gene. This would exclude g-2449C-G.

Closer examination of the sequences flanking g+6723G-A showed that the A allele creates one of the 3' UTR octamer motifs (ACATTCCA) discovered earlier8 and assumed to correspond to miRNA targets. The probability that a random mutation in the sheep *GDF8* 3' UTR creates one of these 540 octamer motifs is 0.045; the probability to create an octamer motif with an equally high conservation score (18.2; ref. 8) is 0.0088. This suggested that

g+6723G-A might contribute to the muscular hypertrophy by causing miRNA-mediated translational inhibition of the Texel *GDF8* allele.

Three known miRNAs target the ACATTCCA octamer described above: miR-1, miR-206 and miR-122a (ref. 8). Notably, miR-1 has previously been shown to be strongly expressed in skeletal muscle and heart in the mouse9. We designed primer pairs based on interspecies alignments to amplify the sheep miRNA orthologs. Note that in man and mice, mature miR-1 is processed from two paralogous genes: miR-1.1 and miR-1.2. We obtained PCR products for the four miRNA genes and sequenced them. This confirmed the presence of the corresponding genes in the sheep and the perfect conservation of the mature miRNAs, including their seed (Supplementary Fig. 5 online). We examined their expression profiles in sheep by primer extension (Fig. 2b). Although miR-122 was expressed in liver and tongue but not in skeletal muscle, miR-1 and miR-206 were strongly expressed in skeletal muscle and tongue. As in the mouse, we also detected miR-1 in the heart. The observation of a strong expression of miR-1 and miR-206 in skeletal muscle, the primary site of *GDF8* expression, lent further strength to our hypothesis.

If g+6723G-A causes miRNA-mediated translational repression of *GDF8* transcripts, Texel sheep should have reduced levels of circulating myostatin. Indeed, myostatin protein is detected in serum of wild-type mice and humans but is absent in serum of mice and humans homozygous for *GDF8* loss-of-function mutations10, 11. Using a monoclonal antibody to myostatin, we immunoprecipitated myostatin from serum of nine Texel animals and six controls and detected its presence using a polyclonal antibody specific for the mature protein by protein blotting. We identified a 12.5-kDa band corresponding to mature myostatin in all 15 sera (Fig. 2c). Most importantly, the intensity of the 12.5-kDa band in Texel animals was approximately one third of the intensity in wild-type animals.

The model of miRNA-mediated translational inhibition predicts reduced stability for the mutant *GDF8* transcript owing to accelerated degradation in P-bodies (see, for example, ref. 12). To test this, we compared the relative abundance of A versus G transcripts in

skeletal muscle of AG heterozygotes. This approach is more sensitive, as it is less affected by variation between individuals and between samples, as is observed when comparing *GDF8* levels between homozygotes. The wild-type G transcripts were indeed 1.5 times more abundant than the A transcripts (Fig. 2d).

We then aimed to test the interaction between mutant *GDF8* transcripts and miR-1 and miR-206 directly. We cloned four tandem repeats9 of an 80-bp sequence centered around g+6723G-A into the 3' UTR of luciferase reporter vectors (creating constructs pRL-TK-4 A, containing Texel sequences, and pRL-TK-4 G, containing control sequences (Fig. 3a)). We cotransfected COS1 cells with these reporter constructs and with pcDNA3 vectors expressing either miR-1 and miR-206 or the control miR-136 and miR-377. In agreement with our prediction, when we cotransfected the reporter constructs with pcDNA3 vectors expressing either miR-1 or miR-206, we observed a highly significant ($P = 0.0002$ and 0.0011, respectively) reduction of pRL-TK-4 A signal to 30% of the signal obtained with pRL-TK-4 G or the unmodified pRL-TK (Fig. 3b). On the other hand, when we cotransfected the same reporter constructs with an empty pcDNA3 vector or with pcDNA3 vectors expressing miR-136 or miR-377, there was no significant difference between luminescence obtained with pRL-TK-4 A, pRL-TK-4 G or unmodified pRL-TK.

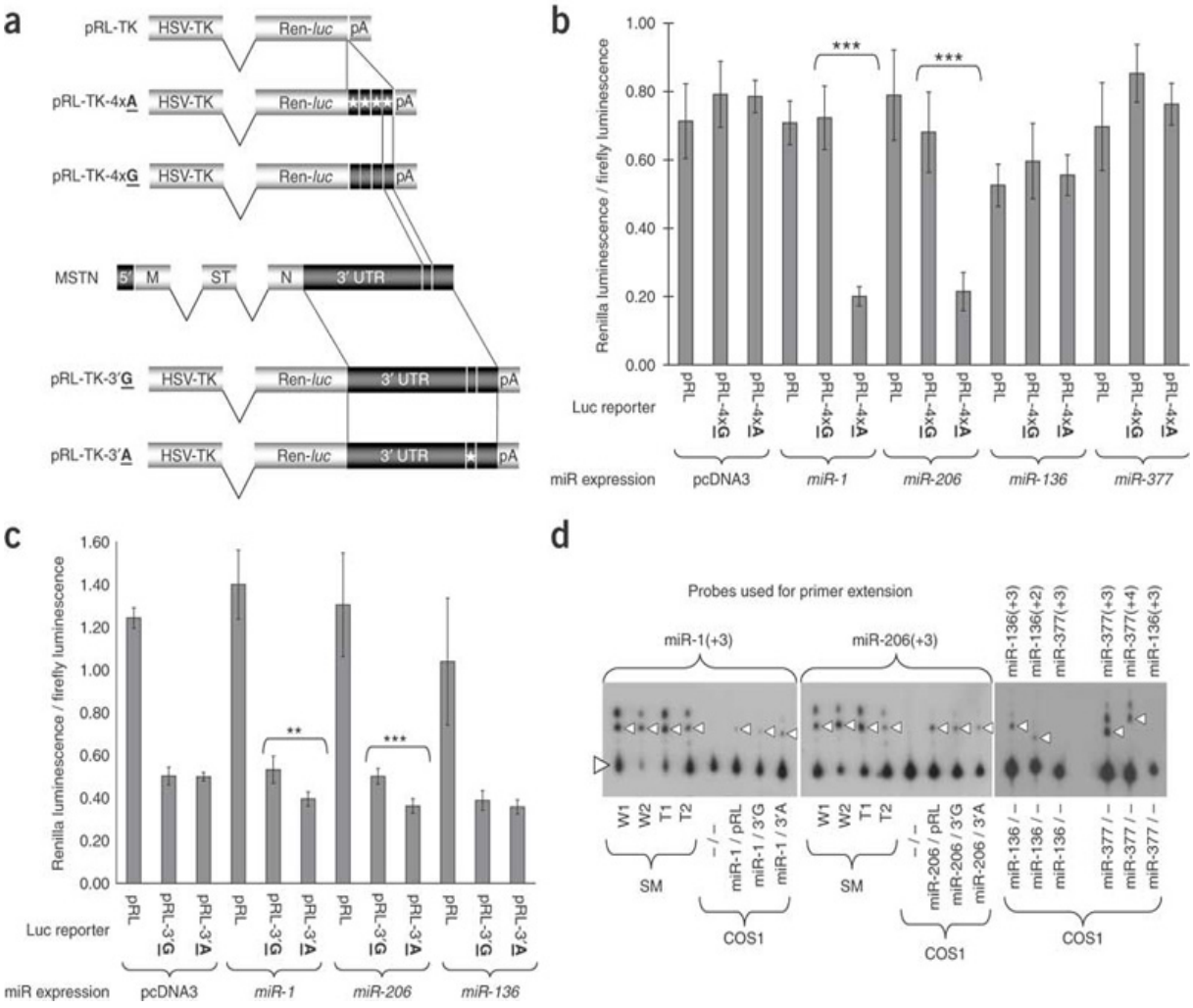


Figure 3. Reporter assay testing the interaction between miRNA-*GDF8* interaction. (**a**) Schematic representation of the Renilla luciferase expression vectors and the *GDF8* gene (MSTN). (**b**,**c**) Renilla-to-firefly luminescence ratios observed when cotransfecting COS1 cells with the indicated luciferase reporter and either an empty pcDNA3 vector or a pcDNA3 vector expressing the miRNAs as indicated (miR-1, miR-206, miR-136 and miR-377). Error bars: ±1.96 s.e.m. obtained from three (**b**) or five (**c**) replicates. ***: $P < 0.001$; **: $P < 0.01$. (**d**) Detection of the mature miRNAs in sheep skeletal muscle (SM) and in transfected COS1 cells. Probes are given above the autoradiograms and the origin of the RNA samples underneath. W1–W2 correspond to SM RNA from two controls; T1–T2 to SM RNA from two Texels. In the case of the *COS1* RNA, we mention the corresponding pcDNA3 and pRL-TK vectors used for transfection. Small arrowheads mark the positions of the expected primer extension products; the large arrowhead marks the position of the unextended probes. Equal amounts (5 μg) of total RNA were used in all cases.

We repeated these experiments using luciferase vectors in the 3' UTR of which we cloned the complete 1.5-kb mutant or wild-type 3' UTR, creating constructs pRL-TK-3'A and pRL-TK-3'G, respectively (Fig. 3a). Inserting the *GDF8* 3' UTR in the vector reduced the luminescence to 40% of the signal obtained with the unmodified vector. More notably, the signal obtained from pRL-TK-3'A and pRL-TK-3'G did not differ when we cotransfected the

COS1 cells with an empty pcDNA vector or with one expressing miR-136, whereas we observed a significant reduction of the pRL-TK-3'A signal to 70% of the pRL-TK-3'G signal when expressing either miR-1 or miR-206 (P = 0.0029 and 0.0003, respectively) (Fig. 3c).

In these experiments, we were able to detect mature miRNAs only after transfection with the cognate pcDNA3 vectors. The expression levels of miR-1 and miR-206 in COS1 cells were, however, lower than in sheep muscle (Fig. 3d).

Our results support a model in which a point mutation in the *GDF8* 3' UTR creates an illegitimate target site for at least two miRNAs that are strongly expressed in skeletal muscle. This results in miRNA-mediated translational downregulation and reduction in myostatin concentrations contributing to muscular hypertrophy. It is tempting to speculate that such hypomorphic alleles are not a sheep *GDF8* idiosyncrasy but that they make a substantial contribution to phenotypic variation in other organisms, including man. To evaluate the frequency of polymorphic miRNA-target interactions, we searched among 73,497 SNPs mapping to the 3' UTR of 13,621 human genes for those creating or destroying one of the octamers from ref. 8. The ancestral allele was determined from the orthologous chimpanzee sequence. We identified 2,490 putative Texel-like SNPs, creating an illegitimate miRNA target site. In addition, we identified 2,597 SNPs destroying at least one motif. Of these 2,597 SNPs, 483 affect an octamer perfectly conserved across the four analyzed mammalian species. These 483 have a fairly high likelihood of affecting gene regulation and hence phenotypic variation. Among SNPs uncovering target sites, those promoting miRNA-antitarget interactions13 are of most interest. They could be identified by comparing the expression profile of target and miRNA. That polymorphic miRNA-target interactions may contribute to disease is illustrated by the recent identification of a SNP associated with Tourette syndrome that affects the interaction between miR-189 and the 3' UTR of *SLITRK1* (ref. 14). We have performed the same analyses for the mouse, using the orthologous rat sequences to infer ancestral state for 77,283 SNPs in the 3' UTR of 10,200 genes. We have identified 1,182 SNPs creating and 1,321 SNPs destroying putative miRNA target sites, 234 of

which are evolutionary conserved. We have generated a website (Patrocles) that lists SNPs that have the potential to affect miRNA-target interactions.

The nature of the mutations that underlie genetic variation for complex traits remains a matter of debate: do quantitative trait nucleotides (QTN) primarily affect gene structure or gene regulation? Are epistatic interactions between QTN the rule or the exception? This work identifies a new class of regulatory mutations that might make an important contribution to the heritability of complex traits. It also points toward possible epistatic interactions between polymorphisms in miRNA genes and their targets.

METHODS

Map construction.

Primers for the amplification of microsatellite markers were obtained from public-domain cattle and sheep maps. Microsatellite genotyping was performed using standard procedures. Linkage maps were constructed using CRIMAP15.

QTL mapping.

QTL mapping was performed using QTL Express16. The nominal P values of the F statistics generated by QTL Express were Bonferroni corrected for 17 independent tests (as deduced from the permutations performed by QTL Express) to obtain chromosome-wide P values and then were Bonferroni corrected for 27 additional tests (corresponding to the number of sheep chromosomes) to obtain genome-wide P values. Confidence intervals for the QTL location were obtained by bootstrapping implemented in QTL Express.

Marker-assisted segregation analysis.

The likelihoods of the pedigree data were computed as

$$L = \prod_i^{n_{rec}} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(Ph_i-(M+\frac{a}{2}))^2}{2\sigma^2}} \prod_j^{n_T} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(Ph_j-(M-\frac{a}{2}))^2}{2\sigma^2}}$$

where nrec represents the number of offspring inheriting the recombinant Texel-Romanov chromosome from the sire (see Fig. 1b), and nT represents the number of offspring inheriting the nonrecombinant Texel chromosome from their sire. Phi and Phj correspond to the phenotypic values of the ith and jth offspring, respectively; M represents the midpoint between the means of the rec and T offspring; a refers to the R T QTL allele substitution effect and 2 corresponds to the residual variance. To compute the likelihood of the data assuming that the sire is homozygous TT for the QTL (LTT), we set a at 0 and estimated the values of M and 2 that maximized the likelihood. To compute the likelihood of the data assuming that the sire is heterozygous TR for the QTL (LTR), we jointly estimated the values of a, M and 2 that yielded the highest likelihood of the data (LML). 2LN(LML/LTT) was assumed to have a 2 distribution with one degree of freedom under the null TT hypothesis.

Selective sweep detection.

To detect the effects of a putative selective sweep on the allelic frequency spectrum in hypermuscled Texel animals compared with control animals, we analyzed the microsatellite genotypes of 41 hypermuscled Texel animals and 108 wild-type controls representing 16 different breeds (Blackbelly: 2; Booroola: 4; Ile de France: 12; Lacaune: 5; Merinos: 2; Rambouillet: 10; Romanov: 17; Southdown: 3; Suffolk: 10; Tarasconnaise: 9; Targee: 3; Berrichon du Cher: 6; Blanc du Massif Central: 5; Charmoise: 4; Charollais: 8; Manech: 8) using DISLAMB5.

Resequencing of the myostatin gene.

To resequence the *GDF8* gene from Texel and control animals we (i) amplified the coding parts of the *GDF8* gene from genomic DNA in three PCR products of 372, 375 and 381 bp, respectively; (ii) amplified the entire *GDF8* ORF by RT-PCR from skeletal muscle mRNA in two overlapping PCR products of 805 and 625 bp, respectively and (iii) amplified 10.5 kb spanning the *GDF8* gene from genomic DNA in 13 overlapping segments. The corresponding

primer pairs are listed in Supplementary Table 3 online. For RT-PCR, RNA was extracted from skeletal muscle using Trizol (Invitrogen), and cDNA was synthesized using the SuperScript III First Strand Synthesis System for RT-PCR (Invitrogen). All PCR products were gel purified using Geneclean (Qbiogene) and sequenced on both strands using the same primers and BigDye Terminator v3.1 Cycle Sequencing kits (Applied Biosystems) and a 3730 DNA Analyzer (Applied Biosystems).

RNA blot analysis.

Total RNA was extracted from skeletal muscle using RNA Insta-Pure (Eurogentec), and mRNA was isolated using the Oligotex Direct mRNA mini kit (Qiagen). The mRNA was size fractionated using the Reliant RNA Gel System (BMA) and was blotted on an Ambion membrane in 5 SSC, 10 mM NaOH. The membrane was hybridized in Ultrahyb (Ambion) buffer at 42 °C to a sheep *GDF8* cDNA probe labeled with the Random-Primed DNA Labeling Kit (Boehringer Mannheim). The membrane was washed at 42 °C in 2 SSC, 0.1% SDS (wt/vol), washed in 0.1 SSC, 0.1% SDS and subjected to autoradiography.

Genotyping of the g+6723G-A and other myostatin SNPs.

Genotyping of the g+6723G-A SNP was done by PCR–restriction fragment length polymorphism analysis. A 1,003-bp fragment encompassing the SNP was amplified by PCR from genomic DNA using primers g+6723G-A.UP1 and g+6723G-A.DN1 (Supplementary Table 3), digested using HpyCH4IV and size fractionated by agarose gel electrophoresis. The g+6723G-A SNP destroys a restriction site that cleaves the G allele (but not the A allele) into a 270-bp and a 733-bp fragment. To genotype the 19 other *GDF8* SNPs, we amplified six amplicons from genomic DNA using standard PCR conditions and primers reported in Supplementary Table 3, purified them using the Multiscreen PCR 96 Filter Plate (Millipore) and sequenced them using a BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and a 3730 DNA Analyzer (Applied Biosystems).

PCR amplification and sequencing of the sheep miR-1.1, miR-1.2, miR-122 and miR-206 genes.

The human, mouse, rat and cattle miR-1.1, miR-1.2, miR-122 and miR-206 genes were aligned using ClustalW. Primer pairs (Supplementary Table 3) were designed in conserved segments of the gene and used to amplify the orthologous sheep genes by PCR from genomic DNA. The PCR products were gel purified using Geneclean (Qbiogene) and sequenced on both strands using the same primers and a BigDye Terminator v3.1 Cycle Sequencing Kit and a 3730 DNA Analyzer.

Primer extension assay to detect mature miRNAs.

miRNA expression was evaluated by primer extension as previously described17 using primers reported in Supplementary Table 3.

Detection of myostatin protein by immunoprecipitation and protein blotting.

We prepared beads coupled to a monoclonal antibody (JA-16) to myostatin (directed against a C-terminal peptide of myostatin), and we immunoprecipitated myostatin by incubating 60 l packed beads with 0.4 ml of serum18. Serum volumes were adjusted for total protein concentration, which was determined using the Quick Start Bradford Protein Assay (Biorad). After washing, bound myostatin was eluted with Laemmli buffer. Samples were separated by SDS-PAGE, blotted on a nitrocellulose membrane and probed with rabbit polyclonal antibody to myostatin (L8014).

Measuring g+6723G-A allelic imbalance at the mRNA level using hot-stop PCR.

Total RNA was extracted from skeletal muscle of heterozygous AG animals (three 70-d-old fetuses and four 4-month-old animals) using Trizol (Invitrogen). The RNA was treated with TurboDNase (Ambion). cDNA was synthesized using SuperScriptIII First Strand Synthesis System for RT-PCR (Invitrogen). Hot-stop PCR was performed according to ref. 19, using the g+6723G-A.UP2 and g+6723G-A.DN2 primers (Supplementary Table 3), amplifying a 396-bp fragment of the *GDF8* 3' UTR. The labeled PCR products were digested with HpyCH4IV

(cleaving the G allele into a 235-bp and a 161-bp fragment) and were size fractionated by denaturing PAGE. The intensity of the respective restriction fragments were quantified using a Phosphorimager (Molecular Dynamics). The proportion of A allele was estimated from the ratio I396 / (I396 + I 235+161) (where Ix corresponds to the intensity of the corresponding fragment) and from a calibration curve established using template DNA with known A-to-G ratios (see Fig. 2d).

Testing the interaction between the Texel myostatin 3' UTR and miR-1 and miR-206 using a dual-luciferase reporter assay in COS1 cells.

To construct the pRL-TK-4 A (Texel) and pRL-TK-4 G (wild-type) vectors, we amplified an 80-bp fragment of the *GDF8* 3' UTR encompassing the g+6723G-A SNP from genomic DNA of a Texel animal and a Romanov animal using primers Xba-ovmyo1211-f (with an XbaI tail) and Spe-ovmyo1290-r (with an SpeI tail) (Supplementary Table 3). The primers were chosen to avoid the occurrence in the final construct of secondary RNA structures that might occlude the miRNA target site as assessed using RNAfold20. XbaI- and SpeI-digested (New England Biolabs) PCR products were self-ligated (LigaFast Rapid DNA Ligation System, Promega). XbaI/SpeI-resistant tetramers were gel purified and ligated in the XbaI site of the pRL-TK vector (Promega). To construct the pRL-TK-3'A (Texel) and pRL-TK-3'G (wild-type) vectors, we amplified the entire *GDF8* 3' UTR from genomic DNA of a Texel animal and a Romanov animal using primers Xba-ovmyo3'UTR-f (with an XbaI tail) and Spe-ovmyo3'UTR-r (with an SpeI tail) (Supplementary Table 3). The XbaI- and SpeI-digested PCR products were cloned in the XbaI site of the pRL-TK vector. Plasmid DNA was purified using the EndoFree plasmid maxi kit (Qiagen), and the inserts of all constructs were completely sequenced. The sheep miR-1.1, miR-206, miR-136 and miR377 genes were amplified from genomic DNA using primer pairs with HindIII and NheI tails (Supplementary Table 3). The HindIII/NheI-digested PCR products were directionally cloned in the HindIII/NheI site of the pcDNA3.1(+) vector (Invitrogen). COS1 cells (European Collection of Cell Cultures (ECACC) no. 88031701) were maintained in DMEM supplemented with 10% fetal bovine serum, 2 mM glutamine, 0.1 mM non-essential amino acids, penicillin (100 units/ml) and streptomycin (100 g/ml). Using Lipofectamine 2000 (Invitrogen) following the manufacturer's recommendations, we

transfected the 0.8 105 cells per well in 24-well plates with a mixture comprising 400 ng of pRL-TK Renilla luciferase construct, 400 ng of pcDNA3.1 construct and 10 ng of pGL3 firefly luciferase control vector (Promega). The luciferase assays were performed 24 h after transfection using the dual-luciferase reporter assay system (Promega) and a Centro LB960 luminometer (Berthold Technologies).

REFERENCES

1. Marcq, F. et al. Preliminary results of a whole-genome scan targeting QTL for carcass traits in a Texel-Romanov intercross. Proc. 7th World Cong. On Genetics Appl. Livest. Prod., Montpellier, France 323–326 (2002).

2. Laville, E. et al. Effects of a QTL for muscle hypertrophy from Belgian texel sheep on carcass conformation and muscularity. J. Anim. Sci. 82, 3128–3137 (2004).

3. Haley, C.S. , Knott, S.A. & Elsen, J.-M. Mapping quantitative trait loci in crosses between outbred lines using least squares. Genetics 136, 1195–1207 (1994).

4. Johnson, P.L. , McEwan, J.C. , Dodds, K.G. , Purchas, R.W. & Blair, H.T. A directed search in the region of GDF8 for quantitative trait loci affecting carcass traits in Texel sheep. J. Anim. Sci. 83, 1988–2000 (2005).

5. Terwilliger, J.D. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic loci. Am. J. Hum. Genet. 56, 777–787 (1995).

6. Lee, S.J. & McPherron, A.C. Myostatin and the control of skeletal muscle mass. Curr. Opin. Genet. Dev. 9, 604–607 (1999).

7. Tobin, J.F. & Celeste, A.J. Myostatin, a negative regulator of muscle mass: implications for muscle degenerative diseases. Curr. Opin. Pharmacol. 5, 328–332 (2005).

8. Xie, X. et al. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. Nature 434, 338–345 (2005).

9. Zhao, Y. , Samal, E. & Srivastava, D. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. Nature 436, 214–220 (2005).

10. Zimmers, T.A. et al. Induction of cachexia in mice by systematically administered myostatin. Science 296, 1486–1488 (2002).

11. Schuelke, M. et al. Myostatin mutation associated with gross muscle hypertrophy in a child. N. Engl. J. Med. 350, 2682–2688 (2004).

12. Zamore, P.D. & Haley, B. Ribo-gnome: the big world of small RNAs. Science 309, 1519–1524 (2005).

13. Bartel, D.P. & Chen, C.-Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. Nat. Rev. Genet. 5, 396–400 (2004).

14. Abelson, J.F. et al. Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science 310, 317–320 (2005).

15. Lander, E. & Green, P. Construction of multilocus genetic linkage maps in humans. Proc. Natl. Acad. Sci. USA 84, 2363–2367 (1987).

16. Seaton, G. , Haley, C.S. , Knott, S.A. , Kearsey, M. & Visscher, P.M. QTL Express: mapping quantitative trait loci in simple and complex pedigrees. Bioinformatics 18, 339–340 (2002).

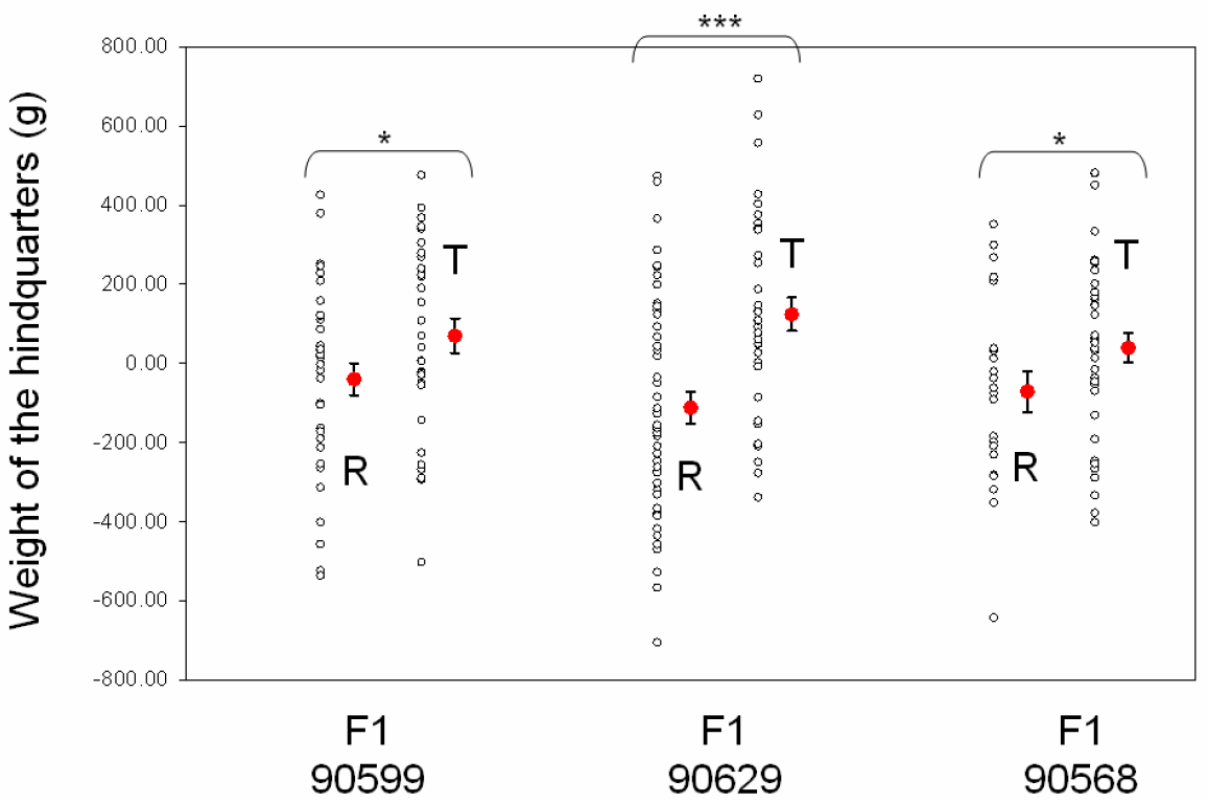
17. Davis, E. et al. RNAi-mediated allelic trans-interaction at the imprinted callipyge locus. Curr. Biol. 15, 743–749 (2005).

18. Hill, J.J. et al. The myostatin propeptide and the follistatin-related gene are inhibitory binding proteins of myostatin in normal serum. J. Biol. Chem. 277, 40735–40741 (2002).

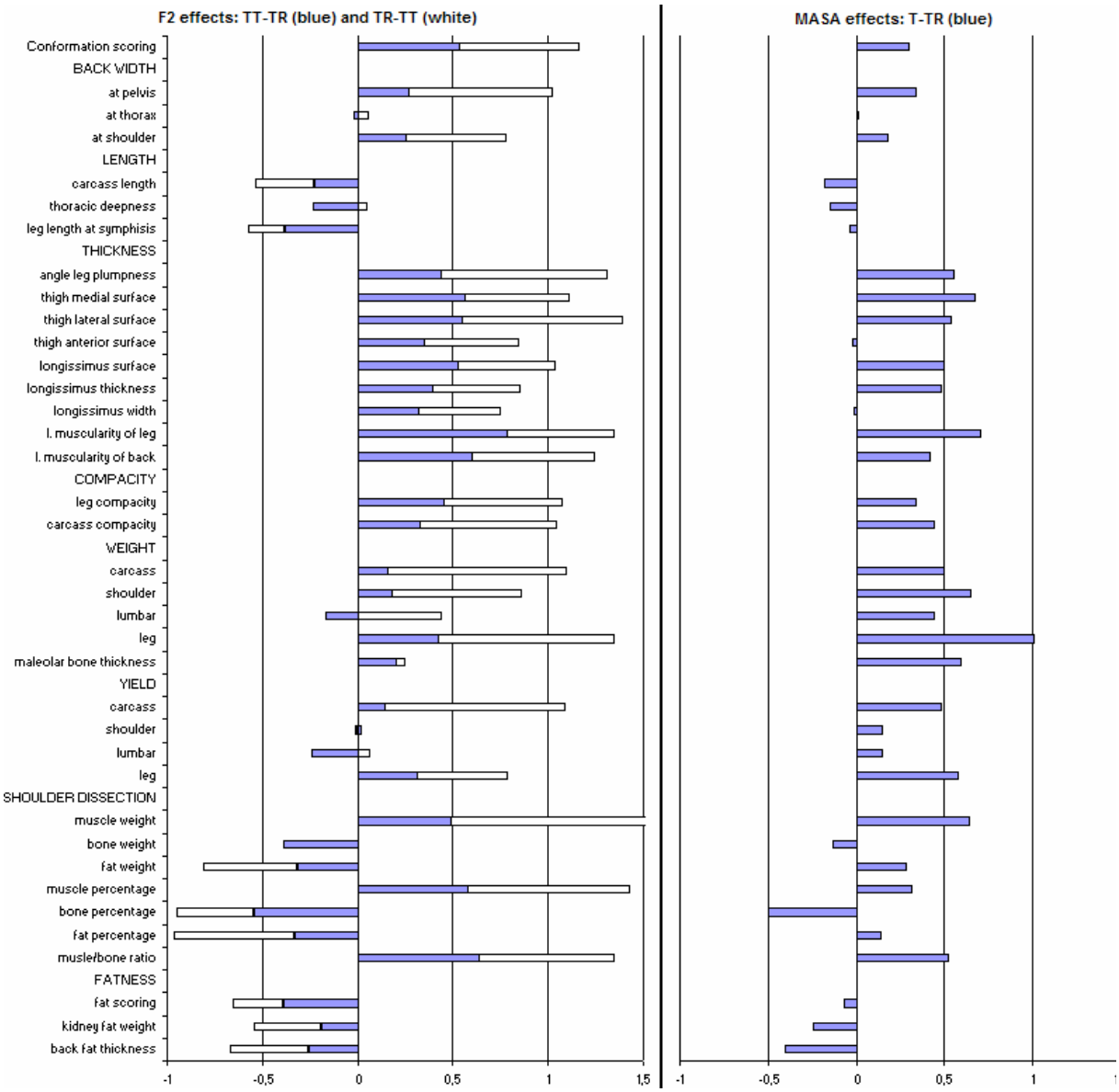
19. Uejima, H. , Lee, M.P. , Cui, H. & Feinberg, A.P. Hot-stop PCR: a simple and general assay for linear quantitation of allele ratios. Nat. Genet. 25, 375–376 (2000).

20. Hofacker, I.L. et al. Fast folding and comparison of RNA secondary structures. Monatsh. Chem. 125, 167–188 (1994).

SUPPLEMENTAL DATA



Supplementary Figure 1: "Within sire-family" QTL analysis demonstrating the TR genotype of the three F1 rams. *The offspring of the three F1 rams (90599, 90629, 90568) were sorted according to the chromosome inherited from their father: T (Texel) or R (Romanov). The phenotypic distribution (weight of the hindquarters) is shown for each offspring-group, as well as the mean ± standard error of the estimate (red circle ± error bars). The significance of the contrast evaluated using a one sided T-test is given: *: < 0.05; ***: < 0.001. The phenotypes were pre-corrected for sex, sire, year of birth, slaughter weight and genotype of the maternal chromosome (T or R).*

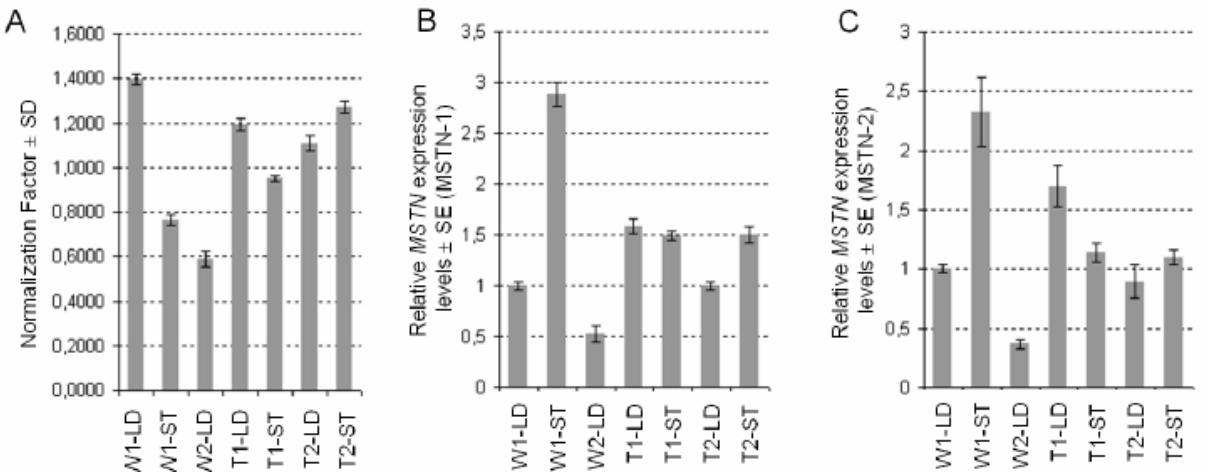


<u>Supplementary Figure 2</u>: Marker Assisted Segregation Analysis.

The effect of the non-recombinant "T" versus recombinant "rec" sire chromosome (cfr. Fig.1B) on the 37 phenotypes (measuring body composition, muscularity and fat deposition) measured on 43 to 64 offspring, was analyzed by ANOVA using the following model:

$$Ph = \mu + PCi + BDi + Si + \Delta i + \varepsilon i$$

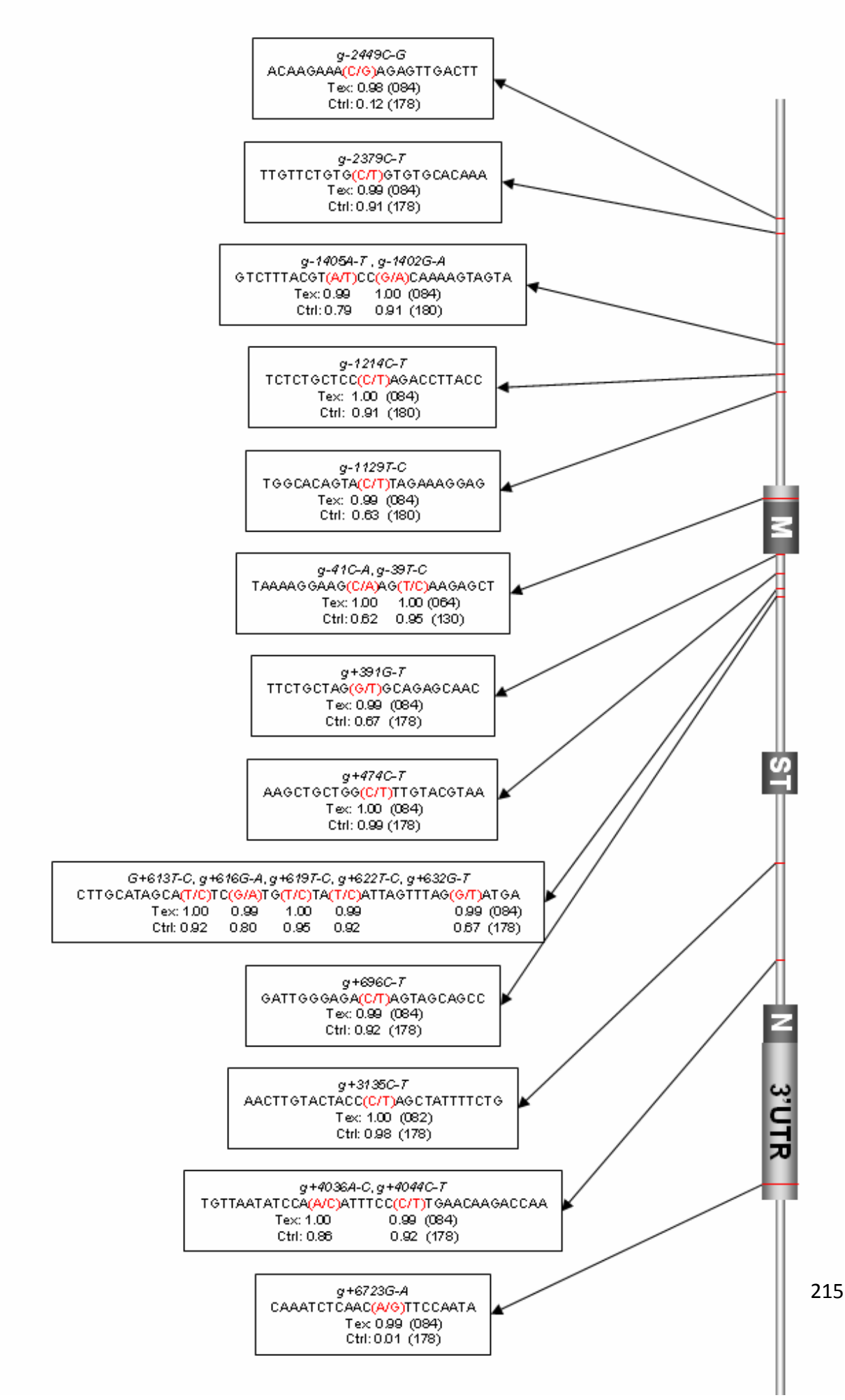
Where Phi is the phenotypic value of offspring i, μ the overall mean, PCi the paternal chromosome inherited by offspring i (T or rec), BDi the breed of the dam (Lacaune, Romanov, Romanov x Texel F1 or Romanov x Texel F2), Si the sex of offspring i (male or female), Δi a covariate corresponding to the difference between the actual weight at slaughter and the target weight at slaughter (39 Kg for the males and 33 Kg for the females), and εi to the residual error. The estimated R to T QTL allele substitution effects (T-rec) are shown in the following figure and compared with the corresponding effects (TT-TR and TR-RR) as estimated in the F2 intercross2. Effects are expressed in residual standard deviation units. *: significant at the 0.05 level; **: significant at the 0.01 level.



<u>Supplementary Figure 3</u>:

Comparing the amounts of *MSTN* mRNA in skeletal muscle of Texel and wild-type sheep using real-time quantitative RT-PCR.

To compare amounts of *MSTN* mRNA in skeletal muscle of Texel versus wild-type sheep we extracted total RNA from *longissimus dorsi* (LD) and *semitendinosus* (ST) of two 3-month old male Texel individuals (T1-T2) and two 3-month old male controls (W1-W2) using Trizol (Invitrogen). After DNase-treatment (Turbo DNA-free, Ambion), 500 ng total RNA was reverse transcribed in a final volume of 20 µl using the iScript cDNA Synthesis Kit (Bio-Rad). PCR reactions were performed in a final volume of 15 µl containing 2 µl of 2.5-fold diluted cDNA (corresponding to 20 ng of starting total RNA), 7.5 µl of 2X master mix prepared from the qPCR Core Kit for SYBR green I (Eurogentec), 0.45 µl of 1/2000 SYBR green I working solution prepared from the qPCR Core Kit for SYBR green I (Eurogentec), forward and reverse primers (250 nM each) and nuclease free water. PCRs were performed on a 7900HT Fast Real-Time PCR System (Applied Biosystems) under the following cycling conditions: 10 min at 95°C followed by 40 cycles at 95°C for 15 sec and 60°C for 1 min. Two primer sets were used to test *MSTN* expression (MSTN-1 and -2) and three genes were used as endogenous controls: ribosomal protein large P0 (RPLPO), succinate dehydrogenase complex subunit A flavoprotein (SDHC) and Tyr-3- & Trp-5-monooxygenase activation protein beta pp (YWHAB). The latter were selected as optimal reference set using geNorm 21 from a total of six candidate reference genes including in addition beta-2-microglobulin (B2M), glyceraldehyde-3-phosphate dehydrogenase (GAPD), ribosomal protein S18 (RPS18). The corresponding primer sequences are given in Supplementary Table 3. All sample / gene combinations were analyzed in triplicate. Relative *MSTN* expression levels in the different samples were computed using the qBase software package (http://medgen.ugent.be/qbase/). Figure A shows the normalization factors (±SD) applied to each of the 7 samples. Figures B&C show the relative *MSTN* levels (±SE) in the 7 samples, estimated using primer sets MSTN-1 and MSTN-2 respectively. From these results, one can conclude that (i) for a given muscle type (LD or ST), considerable differences in *MSTN* expression levels exist between individuals, (ii) for a given individual, considerable differences in *MSTN* expression levels exist between muscle type, and (iii) there is no obvious effect of breed (Texel versus wildtype) on *MSTN* expression levels in this test.


g-2449C-G
ACAAGAAA(C/G)AGAGTTGACTT
Tex: 0.98 (084)
Ctrl: 0.12 (178)
g-2379C-T
TTGTTCTGTG(C/T)GTGTGCACAAA
Tex: 0.99 (084)
Ctrl: 0.91 (178)
g-1405A-T , g-1402G-A
GTCTTTACGT(A/T)CC(G/A)CAAAAGTAGTA
Tex: 0.99 1.00 (084)
Ctrl: 0.79 0.91 (180)
g-1214C-T
TCTCTGCTCC(C/T)AGACCTTACC
Tex: 1.00 (084)
Ctrl: 0.91 (180)
g-1129T-C
TGGCACAGTA(C/T)TAGAAAGGAG
Tex: 0.99 (084)
Ctrl: 0.63 (180)
M
g-41C-A, g-39T-C
TAAAAGGAAG(C/A)AG(T/C)AAGAGCT
Tex: 1.00 1.00 (064)
Ctrl: 0.62 0.95 (130)
g+391G-T
TTCTGCTAG(G/T)GCAGAGCAAC
Tex: 0.99 (084)
Ctrl: 0.67 (178)
ST
g+474C-T
AAGCTGCTGG(C/T)TTGTACGTAA
Tex: 1.00 (084)
Ctrl: 0.99 (178)
G+613T-C, g+616G-A, g+619T-C, g+622T-C, g+632G-T
CTTGCATAGCA(T/C)TC(G/A)TG(T/C)TA(T/C)ATTAGTTTAG(G/T)ATGA
Tex: 1.00 0.99 1.00 0.99 0.99 (084)
Ctrl: 0.92 0.80 0.95 0.92 0.67 (178)
g+696C-T
GATTGGGAGA(C/T)AGTAGCAGCC
Tex: 0.99 (084)
Ctrl: 0.92 (178)
N
g+3135C-T
AACTTGTACTACC(C/T)AGCTATTTTCTG
Tex: 1.00 (082)
Ctrl: 0.98 (178)
3'UTR
g+4036A-C, g+4044C-T
TGTTAATATCCA(A/C)ATTTCC(C/T)TGAACAAGACCAA
Tex: 1.00 0.99 (084)
Ctrl: 0.86 0.92 (178)
g+6723G-A
CAAATCTCAAC(A/G)TTCCAATA
Tex 0.99 (084)
Ctrl: 0.01 (178)

Supplementary Figure 4: SNPs discovered in the ovine *MSTN* gene and allelic frequencies in hypermuscled Texels and wild-type controls.

The three exons are shown as larger boxes with the 5' and 3' UTR filled in light gray and the coding sequences in dark gray. The positions of the 20 discovered SNPs are marked by red horizontal lines. Their sequences as well as allelic frequencies in Texels (Tex) and controls (Ctrl) are detailed in the corresponding boxes; the frequency is shown for the allele listed first in the sequence; the numbers in parentheses correspond to the number of analyzed chromosomes.

A)

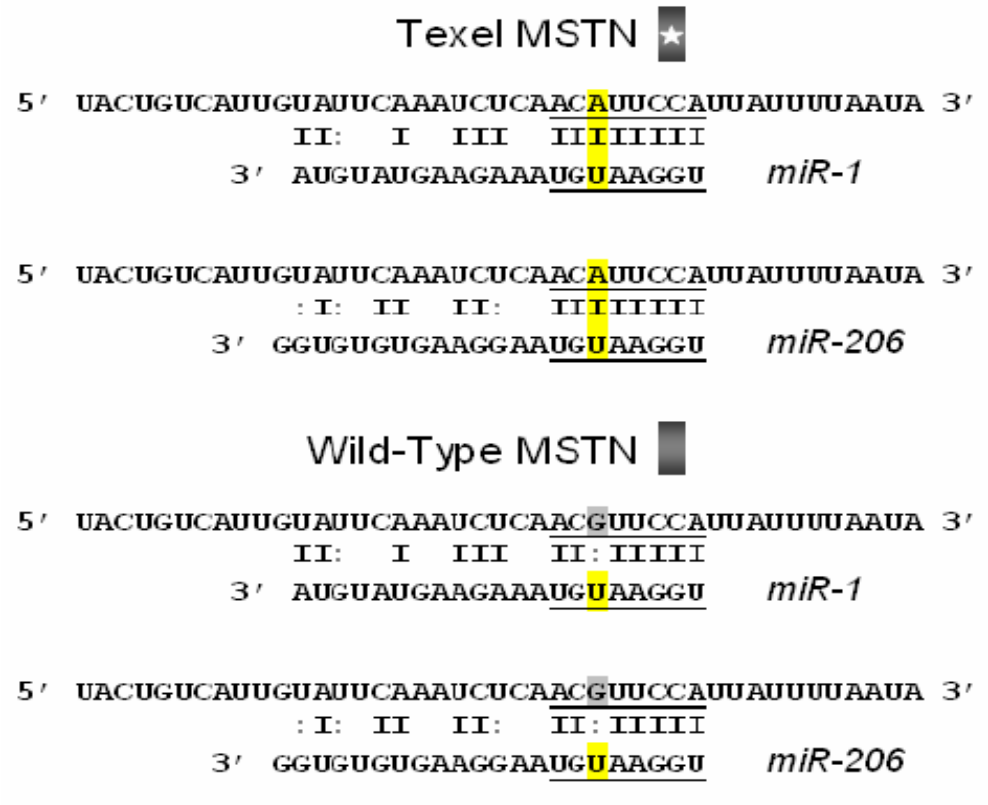
```
miR-1-1
Ovis    CTACCTGCTTGGGAAACATACTTCTTTATGTGCCCATATGGACCTGCTAAGCTATGGAATGTAAAGAAGTATGTATTTCAGGTCGGGA
Bos     CTACCTGCTTGGGAAACATACTTCTTTATGTGCCCATATGGACCTGCTAAGCTATGGAATGTAAAGAAGTATGTATTTCAGGTCGGGA
Canis   CTACCGCTTGGGGAACATACTTCTTTATATGCCCATATGGACCTGCTAAGCTATGGAATGTAAAGAAGTATGTATTTCAGGTGGGGA
Homo    CTGCCTGCTTGGGAAACATACTTCTTTATATGCCCATATGGACCTGCTAAGCTATGGAATGTAAAGAAGTATGTATCTCAGGCCGGGA
Mus     CTACCTGCTTGGGACACATACTTCTTTATATGCCCATATGAACCTGCTAAGCTATGGAATGTAAAGAAGTATGTATTTCAGGCTAGGA
        ** ** *******  ************** ********** ************************ *****   ***

miR-1-2
Homo    TACCTACTCAGAGTACATACTTCTTTATGTACCCATATGAACATACAATGCTATGGAATGTAAAGAAGTATGTATTTTTGGTAGGC
Canis   AACCTACTCAGAGTACATACTTCTTTATGTACCCATATGAACATACAATGCTATGGAATGTAAAGAAGTATGTATTTTTGGTAGGC
Bos     TACCTACTCAGAGTACATACTTCTTTATGTACCCATATGAACATACAATGCTATGGAATGTAAAGAAGTATGTATTTTTGGTAGGC
Ovis    TACCTACTCAGAGTACATACTTCTTTATGTACCCATATGAACATACAATGCTATGGAATGTAAAGAAGTATGTATTTTTGGTAGGC
Mus     TGCCTACTCAGAGCACATACTTCTTTATGTACCCATATGAACATTCAGTGCTATGGAATGTAAAGAAGTATGTATTTTGGGTAGGT
          *********** ******************************** ** ***********************************  ******

miR-206
Ovis    CTTCCCAAGGCCACATGCTTCTTTATATCCCCATACGGATTACTTTGCTATGGAATGTAAGGAAGTGTGTGGTTT
Bos     CTTCCCAAGGCCACATGCTTCTTTATATCCCCATACGGATTACTTTGCTATGGAATGTAAGGAAGTGTGTGGTTT
Canis   CTTCCCGAGGCCACATGCTTCTTTATATCCCCATACGGATTACGTTGCTATGGAATGTAAGGAAGTGTGTGGTTT
Homo    CTTCCCGAGGCCACATGCTTCTTTATATCCCCATATGGATTACTTTGCTATGGAATGTAAGGAAGTGTGTGGTTT
Mus     CTTCCCCAGGCCACATGCTTCTTTATATCCTCATAGATATCTCAGCACTATGGAATGTAAGGAAGTGTGTGGTTT
        ****** ************************ ****   **  *    ***************************

miR-122a
Ovis    CCTTAGCAGAGCTGTGGAGTGTGACAATGGTGTTTGTGTCCAAACTATCAAACGCCATTATCACACTAAATAGCTACTGTTAGGCAATCCTTCC-
Bos     CCTTAGCAGAGCTGTGGAGTGTGACAATGGTGTTTGTGTCCAAACTATCAAACGCCATTATCACACTAAATAGCTACTGTTAGGCAATCCTTCCG
Canis   CCTTAGCAGAGCTGTGGAGTGTGACAATGGTGTTTGTGTCCAAACTATCAAACGCCATTATCACACTAAATAGCTACTGTTAGGCAATCCTTCTG
Homo    CCTTAGCAGAGCTGTGGAGTGTGACAATGGTGTTTGTGTCTAAACTATCAAACGCCATTATCACACTAAATAGCTACTGCTAGGCAATCCTTCC-
Mus     CCTTAGCAGAGCTGTGGAGTGTGACAATGGTGTTTGTGTCCAAACCATCAAACGCCATTATCACACTAAATAGCTACTGCTAGGCAATCCGTCCA
        **************************************** **** ********************************** ********** **
```

B)



C)

Supplementary Figure 5:
(A) Multispecies alignment of *miR-1.1*, *miR-1.2*, *miR-122* and *miR-206*. The sequences of the mature miRNAs are highlighted in yellow.
(B) Effect of the *g+6723G-A* SNP in the *MSTN* 3'UTR on the base pairing with *miR-1* and *miR-206*. **(C)** Multispecies alignment of a segment of the ovine *MSTN* 3'UTR encompassing the *g+6723G-A* SNP.

Phenotype[#]	Genome-wide p-values[&]	Map position (cM)	95 % CI [$] (cM)	r^{2} %	a (a/σ_R)[£]	d (d/a) [@]	2a/(F2-BC) [##]
Weight hindquarters (g)	8.8E-14	127	9	0.26	203 (0.83)	-68 (-0.33)	
Weight of muscles in shoulder (g)	1.4E-12	127	16	0.24	52 (0.78)	-21 (-0.41)	0.39
Conformation score	4.8E-10	129	14	0.20	0.67 (0.67)	0.33 (0.49)	0.21
Carcass width (cm)	8.4E-09	128	14	0.18	0.35 (0.66)	0.00 (0.01)	
Angle leg plumpness (°)	3.5E-08	128	21	0.16	1.65 (0.63)	-0.03 (-0.02)	0.25
Percentage muscle in shoulder (%)	2.0E-08	128	20	0.16	1.61 (0.55)	-0.39 (-0.25)	0.78
Yield of the carcass (%)	3.5E-07	129	12	0.14	0.98 (0.55)	-0.64 (-0.65)	0.33
Percentage of bone (%)	1.7E-06	125	52	0.15	-0.63 (-0.60)	-0.04 (0.07)	0.16
Leg compactness	2.4E-06	128	46	0.15	1.43 (0.55)	0.34 (0.24)	0.19
Yield of the hindquarters (%)	9.6E-06	123	16	0.12	0.55 (0.56)	0.04 (0.08)	
Carcass compactness	1.4E-05	134	51	0.13	0.91 (0.58)	0.04 (0.05)	0.26
Weight of the shoulder (g)	1.4E-05	126	37	0.13	35.6 (0.50)	-25 (-0.70)	0.24
Percentage fat in shoulder (%)	3.8E-03	128	100	0.08	-1.06 (-0.40)	0.41 (-0.39)	-0.99
Weight of fat in shoulder (g)	4.0E-02	128	151	0.05	-14.6(-0.34)	3.8(-0.26)	-0.46

Supplementary Table 1: Effects of the OAR2 QTL on muscularity, fat deposition and body composition significant at the genome-wide 5% level.

Phenotypes are as defined in Laville et al. 2

& Genome-wide p-values were computed by applying a Bonferoni correction corresponding to the realization of 44 independent tests to the nominal p-values of the *F*statistics computed by QTL Express 15 . The 44 independent tests correspond to the equivalent of 17 independent tests performed on chromosome 2 (as deduced from the permutation analysis performed by QTL Express) plus 27 for the number of sheep chromosomes.

$ CI corresponds to the confidence interval for the QTL location computed by bootstrapping using QTL Express.

% r2 corresponds to the part of the phenotypic variance in the F2 population that is explained by the QTL.

d TR TT RR ; thus if a > 0 and d < 0, *T* is partially recessive, if a > 0 and d > 0, *T* is partially dominant, if a < 0 and d > 0, *T* is partially recessive, if a < 0 and d < 0, *T* is partially dominant.

##2a/(F2-BC) estimates the proportion of the phenotypic difference between Texel and Romanov (F2-BC) that is due to the effect of the QTL (2a). In this "F2" is the phenotypic mean of a Romanov x Texel F2 population, while "BC" is the phenotypic mean of a Romanov x (Romanov x Texel) back-cross population. (F2-BC) was obtained from Laville et al. 2 For "percentage fat in shoulder" and "weight of fat in shoulder" the QTL effect is opposite to the breed effect, explaining the negative values of 2a/(F2-BC) .

	g-2449 c-g	g-2379 c-t	g-1405 a-t	g-1402 g-a	g-1214 c-t	g-1129 t-c	g-41 c-a	g-39 t-c	g+391 g-t	g+474 c-t g-t	g+613 t-c	g+616 g-a	g+619 t-c	g+622 t-c	g+632 g-t	g+696 c-t	g+3135 c-t	g+4036 a-c	g+4044 c-t	g+6723 g-a
NAME	C/G	C/T	A/T	G/A	C/T	T/C	C/A	T/C	G/T	C/T	T/C	G/A	T/C	T/C	G/T	C/T	C/T	A/C	C/T	G/A
TEXEL-Utrecht	C	C	A	G	C	T			G	C	T	G	T	T	G	C	C	A	C	A
TEXEL-Tony	C	C	A	G	C	T			G	C	T	G	T	T	G	C	C	A	C	A
TEXEL-Tenor	C	C	A	G	C	T			G	C	T	G	T	T	G	C	C	A	C	A
TEXEL-Tarzan	C	C	A	G	C	T			G	C	T	G	T	T	G	C	C	A	C	A
Romanov-970014	G	C/T	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Romanov-960432	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C/T	A	C	G
Romanov-20339	G	C/T	A/T	G/A	C/T	C	C	T	T	C	T	G/A	T	C/T	T	C/T	C	A	C/T	G
Romanov-11631	G	C	A	G	C	T			G	C	T	G	T	T	G	C	C/T	A	C	G
Romanov-11609	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Romanov-11567	G	C	A	G/A	C/T	T/C	C/A	T	G/T	C	T	G	T	T	G/T	C	C	A	C	G
Romanov-11405	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Romanov-10820	G	C	A	G	C	T			G	C	T	G	T	T	G	C	C	A	C	G
Tarasconnais-990165	C/G	C/T	A	G/A	C/T	T/C	C/A	T	G/T	C	T	G	T	T	G/T	C	C	A	C	G
Tarasconnais-990021	G	C	A/T	G	C	T/C	C	T	G	C/T	T	G	T	T	G	C	C	A	C	G
Tarasconnais-20103	G	C	A	G	C	T			G	C	T	G	T	T	G	C	C/T	A	C	G
Tarasconnais-10140	G	C	T	G	C	C	A	C	T	C	T	G	C	T	T	C	C	C	C	G
Tarasconnais-00021	G	C	A	G	C	T	C/A	T	G/T	C	T	G	T	T	G/T	C	C	A	C	G
			A	G	C	T														G
Suffolk-Balto	G	C	A/T	G	C	T/C	C/A	C/T	G/T	C	T	G/A	C/T	T	G/T	C	C	A/C	C	G
Suffolk-9	G	C	A/T	G	C	T/C	C/A	T	G/T	C	T/C	G/A	T	T	G/T	C	C	A/C	C	G
Suffolk-8	G	C	A/T	G	C	T/C	C/A	C/T	G/T	C	T	G/A	C/T	T	G/T	C	C	A/C	C	G
Suffolk-7	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Suffolk-6	G	C	A	G	C	T			G/T	C	T/C	G/A	T	T	G/T	C	C	A/C	C	G
Suffolk-5	G	C	T	G	C	C			T	C	C	A	T	T	T	C	C	C	C	G
Suffolk-4	C/G	C	A	G	C	T			G	C	T	G	T	T	G	C	C	A	C	G
Suffolk-3	G	C	A/T	G	C	T/C	C/A	T	G/T	C	T/C	G/A	T	T	G/T	C	C	A/C	C	G
Suffolk-2	G	C	A	G	C	T			G	C	T	G	T	T	G	C	C	A	C	G
Suffolk-1	G	C	A/T	G	C	T/C			G/T	C	T	G/A	C/T	T	G/T	C	C	A/C	C	G
Merinos_Arles-992749	G	C/T	A/T	G	C	T/C			G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Merinos_Arles-992738	G	C	A	G	C	T/C			G	C	T	G	T	T	G	C	C	A	C	G
Merinos_Arles-992719	G	C	A	G	C	T			G	C	T	G	T	T	G	C	C	A	C	G
Merinos_Arles-992717	G	C/T	A/T	G	C	C			G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Merinos_Arles-992715	G	C	A/T	G	C	T/C			G/T	C	T/C	G/A	T	T	G/T	C	C	A/C	C	G
Manech_tete_rousse-9	G	C	A	G	C	T	C/A	T	G	C	T	G	T	T	G	C	C	A	C	G
Manech_tete_rousse-8	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Manech_tete_rousse-7	C/G	C	A	G/A	C/T	T/C	C/A	T	G/T	C	T	G	T	T	G/T	C	C	A	C	G
Manech_tete_rousse-6	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Manech_tete_rousse-5	C/G	C	A	G	C/T	T/C	C/A	T	G/T	C	T	G	T	T	G/T	C	C	A	C	G
Manech_tete_rousse-4	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Manech_tete_rousse-3	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Manech_tete_rousse-2	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Manech_tete_rousse-10	G	C	A	G/A	C/T	T/C	C/A	T	G/T	C	T	G	T	T	G/T	C	C	A	C	G
Manech_tete_rousse-1	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Lacaune-982129	C/G	C	A	G/A	C/T	T/C	C/A	T	G/T	C	T	G	T	T	G/T	C	C	A	C	G
Lacaune-982046	G	C	A/T	G	C	T/C	C/A	T	G/T	C	T/C	G/A	T	T	G/T	C	C	A/C	C	G
Lacaune-982043	G	C	A/T	G	C	C	A	C/T	G/T	C	T	G/A	C/T	T	G/T	C	C	A/C	C	G
Lacaune-982030	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Lacaune-982027	G	C	A	G	C	T/C	C/A	T	G/T	C	T	G	T	T	G/T	C	C	A	C	G
																				G
Ile_de_france-9	G	C/T	A/T	G	C	T/C			G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Ile_de_france-8	C/G	C	A/T	G	C	C	A	C/T	G/T	C	T	G/A	C/T	T	G/T	C	C	A/C	C	G
Ile_de_france-7	G	C	A/T	G	C	T/C	C/A	T	G/T	C	T/C	G/A	T	T	G/T	C	C	A/C	C	G
Ile_de_france-4	G	C/T	A/T	G	C	T/C	C/A	T	G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Ile_de_france-3	G	C/T	A/T	G	C	T/C	C/A	T	G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Ile_de_france-12	G	C	A/T	G	C	T/C	C/A	C/T	G/T	C	T	G/A	C/T	T	G/T	C	C	A/C	C	G
Ile_de_france-11	G	C/T	A/T	G	C	T/C			G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Ile_de_france-10	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G

Charollais-9	G	C	A/T	G	C	C	A	T	G/T	C	T/C	G/A	T	T	G/T	C	C	A/C	C	G
Charollais-8	G	C	A	G	C	T/C	C/A	T	G	C	T	G	T	T	G	C	C	A	C	G
Charollais-7	G	C	T	G	C	C	A		T	C	C	A	T	T	T	C	C	C	C	G
Charollais-6	C/G	C	A	G/A	C/T	C			G/T	C	T	G	T	T	G/T	C	C	A	C	G
Charollais-5	C/G	C	A/T	G	C	T/C			G/T	C	T/C	G/A	T	T	G/T	C	C	A/C	C	A/G
Charollais-4	G	C	A	G	C	T/C	C/A	T	G	C	T	G	T	T	G	C	C	A	C	G
Charollais-3	G	C	T	G	C	C	A	T	T	C	C	A	T	T	T	C	C	C	C	G
Charollais-10	C/G	C	A/T	G	C	T/C	C/A	T	G/T	C	T/C	G/A	T	T	G/T	C	C	A/C	C	A/G
Charollais-1	G	C	A	G	C	T	C/A	T	G	C	T	G	T	T	G	C	C	A	C	G
Charmoise-9	G	C/T	A/T	G	C	T/C	C/A	T	G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Charmoise-8	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Charmoise-7	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Charmoise-6	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Charmoise-5	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Charmoise-4	G	C	A	G/A	C/T	T/C	C/A	T	G/T	C	T	G	T	T	G/T	C	C	A	C	G
Charmoise-3	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Charmoise-2	G	C	A	G	C	T/C	C/A	T	G	C	T	G	T	T	G	C	C	A	C	G
Charmoise-10	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Charmoise-1	C/G	C	A	G	C	T/C	C/A		G/T	C	T	G	T	T	G/T	C	C	A	C	G
Blanc_du_M._Central-9	C/G	C	A	G/A	C/T	T/C	C/A	T	G/T	C	T	G	T	T	G/T	C	C	A	C	G
Blanc_du_M._Central-8	C/G	C	A	G/A	C/T	T/C	C/A	T	G/T	C	T	G	T	T	G/T	C	C	A	C	G
Blanc_du_M._Central-7	C/G	C	A	A	T	C	A	T	T	C	T	G	T	T	T	C	C	A	C	G
Blanc_du_M._Central-6	G	C	A	G	C	T			G	C	T	G	T	T	G	C	C	A	C	G
Blanc_du_M._Central-5	G	C	A	G	C	T			G	C	T	G	T	T	G	C	C	A	C	G
Blanc_du_M._Central-4	C	C	A	A	T	C	A	T	T	C	T	G	T	T	T	C	C	A	C	G
Blanc_du_M._Central-3	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Blanc_du_M._Central-2	C/G	C	A	G/A	C/T	T/C			G/T	C	T	G	T	T	G/T	C	C	A	C	G
Blanc_du_M._Central-10	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Blanc_du_M._Central-1	C/G	C	A	G/A	C/T	T/C			G/T	C	T	G	T	T	G/T	C	C	A	C	G
Berrichon_du_Cher-8	C/G	C/T	A/T	G	C	T/C	C/A	T	G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Berrichon_du_Cher-7	G	C/T	A/T	G	C	T/C			G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Berrichon_du_Cher-6	G	C	A	G	C	T	C	T	G	C	T	G	T	T	G	C	C	A	C	G
Berrichon_du_Cher-5	G	C/T	A/T	G	C	T/C	C/A	T	G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Berrichon_du_Cher-4	C/G	C/T	A/T	G	C	T/C	C/A	T	G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Berrichon_du_Cher-3	G	C/T	A/T	G	C	T/C	C/A	T	G/T	C	T	G	T	C/T	G/T	C/T	C	A	C/T	G
Berrichon_du_Cher-2	G	C/T	A/T	G	C	T/C	C/A	T	G/T	C	T	G/A	T	C/T	G/T	C/T	C	A	C/T	G
Berrichon_du_Cher-10	C	C	A	G	C	T/C	C/A	T	G	C	T	G	T	T	G	C	C	A	C	G
Berrichon_du_Cher-1	G	C	A/T	G	C	T/C	C/A	T	G/T	C	T/C	G/A	T	T	G/T	C	C	A/C	C	G

Supplementary Table 2:
Genotypes of 42 Texel, 90 controls and four TR rams (three F1, one F2) for the 20 SNPs discovered in the *MSTN* gene.

Primer name	Primer used	Primer sequence (5' – 3')
Resequencing of the MSTN gene		
	Exon 1 - UP	ATTCACTGGTGTGGCAAGTTGTCTC
	Exon 1 - DN	ACCTGTTATTATAAATGCATATTTC
	Exon 2 - UP	GTTCATAGATTGATATGGAGGTGTTCG
	Exon 2 - DN	ATAAGCACAGGAAACTGGTAGTTATT
	Exon 3 - UP	AATTCATGAAAAGATTGGTGCAG
	Exon 3 - DN	ATACTCTAGGCTTATAGCCTGTGGT
	Segment 1 - UP	AATATGCTTTATACTCTCAATGTTG
	Segment 1 – DN	TTCACATTATGTATATTTTGCCCAC
	Segment 2 - UP	TGCTAATTCTTCAACATTGGTTCACAACC
	Segment 2 - INT	TTATTTATGGATCCCGTGCTTGCAC
	Segment 2 -DN	TAGGTCACTACTTTCGCACATTG
	Segment 3 - UP	TTTTCTCAAGAGATGTAGATACCTC
	Segment 3 - DN	TTAAATGAAAAGACTGTGATACTAG
	Segment 4 - UP	ATATTAAAGTAGGATTTTCATTATG
	Segment 4 – DN	TCCTTCTGCGCTGTTCTCATTCAGATCCA
	Segment 5 – UP	ATTCACTGGTGTGGCAAGTTGTCTCTC
	Segment 5 - DN	ACCTGTTATTATAAATGCATATTTC
	Segment 6 - UP	GCCATAAAAATCCAAATCCTCAG
	Segment 6 - DN	AGTATAACAATTATTACAGCTTGTC
	Segment 7 – UP	TATCAGATAATCCTGGAATAG
	Segment 7 – DN	CAATGCCTAAGTTGGATTCAGGTTG
	Segment 8 - UP	ACAGTGTTTGTGCAAATCCTGAGAC
	Segment 8 - INT	TAAATAAGGCAAATCTATTCCAG
	Segment 8 - DN	TTGACTTTATATTAAATTAGTTGAC
	Segment 9 - UP	TGCTTTTACTTATAGAAATTAACTAG
	Segment 9 – DN	TGCTTTTACTTATAGAAATTAACTAG
	Segment 10 - UP	AATTCATGAAAAGATTGGTGCAG
	Segment 10 - DN	CTCTAGGCTTATAGCCTGTGGT
	Segment 11 – UP	CCAGGCATGGTAGTAGATCGCTGTGGG
	Segment 11 - DN	GCAAAAACTGTTAATGTACTGTATC
	Segment 12 - UP	ATTCCACAAAATAGGGATGGTACG
	Segment 12 - INT	TTTGGTATATTTTTACAGTAAGGAC
	Segment 12 - DN	TAAATAGTGTTGCACGTAAGGAT
Real time quantitative RT-PCR		
MSTN1.UP1	MSTN1	TTTGGGCTTGATTGTGATGA
MSTN1.DN1		CCAGAGCAGTAATTGGCCTTA
MSTN2.UP1	MSTN2	CACAGAAGGTCTTCCCCTCA
MSTN2.DN1		GGTTAAATGCCAACCATTGC
B2M.UP1	B2M	TTCTGTCCCACGCTGAGTTCA
B2M.DN1		CAACCCAAATGAGGCATCGT
GAPD.UP1	GAPD	TGACCCCTTCATTGACCTTCA
GAPD.DN1		GATGGTGATGGCCTTTCCATT
RPLPO.UP1	RPLPO	TGGGCAAGAACACGATGATG
RPLPO.DN1		TGAGGTCCTCCTTGGTGAACA
RPS18.UP1	RPS18	GCAGAATCCACGCCAATACAA
RPS18.DN1		TCTTCAGACGCTCCAGGTCTTC
SDHC.UP1	SDHC	CAGCAGAAGAAGCCGTTTGAG
SDHC.DN1		CACAGTCGGTCTCGTTCAAAG
YWHAB.UP1	YWHAB	GCGTGTCATCTCCAGCATTGA

YWHAB.DN1		AACAGCTCCAGCACGTCATTG
Genotyping of the MSTN polymorphism		
G+6723G-A.UP1		TTTGGTATATTTTTACAGTAAGGAC
G+6723G-A.DN1		TAAATAGTGTTGCACTTAAGGATTC
FRAG1.UP1 213	bp fragment 1	AGCATGGTGGTATACTGATTGCA
FRAG1.DN1		TTGCACATTTATGTCTTGCTGAG
FRAG2.UP1	463 bp fragment 2	GGCAGTAAATTAGTGTACAATGTG
FRAG2.DN1		AAGGTGTTCTTTATGTTCCATTT
FRAG3.UP1	321 bp fragment 3	TGCAAGACTTCATGAGAAATATG
FRAG3.DN1		GGCTTCAACATCTAGAGATTTCTT
FRAG4.UP1	159 bp fragment 4	TGGATTACTTTGTGAATTACTCCT
FRAG4.DN1		AGGCATTATTTTCTGTTTGACTTT
FRAG5.UP1	237 bp fragment 5	GGAGTTCGTCTTTCCAACCCTAT
FRAG5.DN1		GCCTAGCTTATGTCTTCCTGCTT
Analysis of miRNA expression by primer extension		
miR-1-1 (+3)		TACATACTTCTTTACATT
miR-1-1 (+2)		TACATACTTCTTTACATTC
miR-122a (+3)		CAAACACCATTGTCACACT
miR-122a (+2)		ACAAACACCATTGTCACAC
miR-136 (+3)		TCCATCATCAAAACAAATGG
MiR-136 (+2)		CCATCATCAAAACAAATGGA
miR-206 (+3)		CCACACACTTCCTTACATT
miR-206 (+4)		CCACACACTTCCTTACAT
Evaluating MSTN allelic imbalance in mRNA by hot-stop PCR		
G+6723G-A.UP2		TAAATAGTGGTCTTAAAACTCCAT
G+6723G-A.DN2		TCTACACATTAGATGTAAGAAATAA
Luciferase reporter assay in COS1 cells		
Xba-ovmyo1211-f		CAGTCTAGATGTGAAATTTCAATGGTTTACTGT
Spe-ovmyo1290-r		CAGACTAGTTATGCTTAATATTTATAAGTATTA
Xba-ovmyo3'UTR-f		TTACTAGTGGTCTATATTTGGTTCATACCT
Spe-ovmyo3'UTR-r		CATCTAGATTGTCTTTCAAAAAAGGTGAAA
pri-miR1_Nhe_f		TTTGCTAGCCTCGGGGGGCCCTTCCCACA
pri-miR1_Hind_r		TTTAAGCTTCGGCCCCGTGGGCGCGCTGC
pri-miR206_Nhe_f		TTTGCTAGCAGATGTGGGGCTGCATCCG
pri-miR206_Hind_r		TTTAAGCTTGGGAAGAAGGGGCCAGGTG
pri-miR136_Nhe_f		TTTGCTAGCGAGGTCAATGGACTGCTCTC
pri-miR136_Hind_r		TTTAAGCTTACGACCTTACCAGTCTTCAG
pri-miR377_Nhe_f		TTTGCTAGCCAGTGCCACCTCTCCTGAAG
pri-miR377_Hind_r		TTTAAGCTTGACCCTGAGCTCACCACTGG

Supplemental Table 3: Primer sequences

CHAPITRE VI

REFERENCES

Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, Mathews CA, Pauls DL, Rasin MR, Gunel M et al. 2005. Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science* 310(5746): 317-320.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25(17): 3389-3402.

Andersen DC, Jensen CH, Schneider M, Nossent AY, Eskildsen T, Hansen JL, Teisner B, Sheikh SP. 2010. MicroRNA-15a fine-tunes the level of Delta-like 1 homolog (DLK1) in proliferating 3T3-L1 preadipocytes. *Exp Cell Res* 316(10): 1681-1691.

Aravin AA, Hannon GJ, Brennecke J. 2007. The Piwi-piRNA pathway provides an adaptive defense in the transposon arms race. *Science* 318(5851): 761-764.

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 25(1): 25-29.

Bachellerie JP, Cavaille J, Huttenhofer A. 2002. The expanding snoRNA world. *Biochimie* 84(8): 775-790.

Baek D, Villen J, Shin C, Camargo FD, Gygi SP, Bartel DP. 2008. The impact of microRNAs on protein output. *Nature* 455(7209): 64-71.

Baladron V, Ruiz-Hidalgo MJ, Nueda ML, Diaz-Guerra MJ, Garcia-Ramirez JJ, Bonvini E, Gubina E, Laborda J. 2005. dlk acts as a negative regulator of Notch1 activation through interactions with specific EGF-like repeats. *Exp Cell Res* 303(2): 343-359.

Barrell D, Dimmer E, Huntley RP, Binns D, O'Donovan C, Apweiler R. 2009. The GOA database in 2009--an integrated Gene Ontology Annotation resource. *Nucleic Acids Res* 37(Database issue): D396-403.

Bartel DP. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116(2): 281-297.

Bartel DP, Chen CZ. 2004. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet* 5(5): 396-400.

Barton SC, Surani MA, Norris ML. 1984. Role of paternal and maternal genomes in mouse development. *Nature* 311(5984): 374-376.

Bass BL, Weintraub H. 1988. An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* 55(6): 1089-1098.

Bell TA, de la Casa-Esperon E, Doherty HE, Ideraabdullah F, Kim K, Wang Y, Lange LA, Wilhemsen K, Lange EM, Sapienza C et al. 2006. The paternal gene of the DDK syndrome maps to the Schlafen gene cluster on mouse chromosome 11. *Genetics* 172(1): 411-423.

Benne R, Van den Burg J, Brakenhoff JP, Sloof P, Van Boom JH, Tromp MC. 1986. Major transcript of the frameshifted coxII gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 46(6): 819-826.

Berghmans S, Segers K, Shay T, Georges M, Cockett N, Charlier C. 2001. Breakpoint mapping positions the callipyge gene within a 450-kilobase chromosome segment containing the DLK1 and GTL2 genes. *Mamm Genome* 12(2): 183-185.

Betel D, Wilson M, Gabow A, Marks DS, Sander C. 2008. The microRNA.org resource: targets and expression. *Nucleic Acids Res* 36(Database issue): D149-153.

Bonnet E, Wuyts J, Rouze P, Van de Peer Y. 2004. Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* 20(17): 2911-2917.

Bray SJ, Takada S, Harrison E, Shen SC, Ferguson-Smith AC. 2008. The atypical mammalian ligand Delta-like homologue 1 (Dlk1) can regulate Notch signalling in Drosophila. *BMC Dev Biol* 8: 11.

Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. 2008. An epigenetic role for maternally inherited piRNAs in transposon silencing. *Science* 322(5906): 1387-1392.

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. *Genome Res* 13(4): 721-731.

Busboom JR, Wahl TI, Snowder GD. 1999. Economics of callipyge lamb production. *J Anim Sci* 77 Suppl 2: 243-248.

Byrne K, Colgrave ML, Vuocolo T, Pearson R, Bidwell CA, Cockett NE, Lynn DJ, Fleming-Waddell JN, Tellam RL. 2010. The imprinted retrotransposon-like gene PEG11 (RTL1) is expressed as a full-length protein in skeletal muscle from Callipyge sheep. *PLoS One* 5(1): e8638.

Caiment F, Charlier C, Hadfield T, Cockett N, Georges M, Baurain D. 2010. Assessing the effect of the CLPG mutation on the microRNA catalogue of skeletal muscle using high throughput sequencing. *Genome Res*.

Carninci P Kasukawa T Katayama S Gough J Frith MC Maeda N Oyama R Ravasi T Lenhard B Wells C et al. 2005. The transcriptional landscape of the mammalian genome. *Science* 309(5740): 1559-1563.

Carpenter CE, Rice OD, Cockett NE, Snowder GD. 1996. Histology and composition of muscles from normal and callipyge lambs. *J Anim Sci* 74(2): 388-393.

Carter D, Chakalova L, Osborne CS, Dai YF, Fraser P. 2002. Long-range chromatin regulatory interactions in vivo. *Nat Genet* 32(4): 623-626.

Cavaille J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Huttenhofer A. 2000. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci U S A* 97(26): 14311-14316.

Cavaille J, Seitz H, Paulsen M, Ferguson-Smith AC, Bachellerie JP. 2002. Identification of tandemly-repeated C/D snoRNA genes at the imprinted human 14q32 domain reminiscent of those at the Prader-Willi/Angelman syndrome region. *Hum Mol Genet* 11(13): 1527-1538.

Cavaille J, Vitali P, Basyuk E, Huttenhofer A, Bachellerie JP. 2001. A novel brain-specific box C/D small nucleolar RNA processed from tandemly repeated introns of a noncoding RNA gene in rats. *J Biol Chem* 276(28): 26374-26383.

Chai JH, Locke DP, Ohta T, Greally JM, Nicholls RD. 2001. Retrotransposed genes such as Frat3 in the mouse Chromosome 7C Prader-Willi syndrome region acquire the imprinted status of their insertion site. *Mamm Genome* 12(11): 813-821.

Charlier C, Segers K, Karim L, Shay T, Gyapay G, Cockett N, Georges M. 2001. The callipyge mutation enhances the expression of coregulated imprinted genes in cis without affecting their imprinting status. *Nat Genet* 27(4): 367-369.

Charlier C, Segers K, Wagenaar D, Karim L, Berghmans S, Jaillon O, Shay T, Weissenbach J, Cockett N, Gyapay G et al. 2001. 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. *Genome Res* 11(5): 850-862.

Cheloufi S, Dos Santos CO, Chong MM, Hannon GJ. 2010. A dicer-independent miRNA biogenesis pathway that requires Ago catalysis. *Nature* 465(7298): 584-589.

Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR et al. 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res* 33(20): e179.

Chen K, Rajewsky N. 2007. The evolution of gene regulation by transcription factors and microRNAs. *Nat Rev Genet* 8(2): 93-103.

Cheng J, Kapranov P, Drenkow J, Dike S, Brubaker S, Patel S, Long J, Stern D, Tammana H, Helt G et al. 2005. Transcriptional maps of 10 human chromosomes at 5-nucleotide resolution. *Science* 308(5725): 1149-1154.

Chi SW, Zang JB, Mele A, Darnell RB. 2009. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 460(7254): 479-486.

Clare TL, Jackson SP, Miller MF, Elliott CT, Ramsey CB. 1997. Improving tenderness of normal and callipyge lambs with calcium chloride. *J Anim Sci* 75(2): 377-385.

Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibe B, Bouix J, Caiment F, Elsen JM, Eychenne F et al. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nat Genet* 38(7): 813-818.

Cockett NE, Jackson SP, Shay TL, Farnir F, Berghmans S, Snowder GD, Nielsen DM, Georges M. 1996. Polar overdominance at the ovine callipyge locus. *Science* 273(5272): 236-238.

Cockett NE, Jackson SP, Shay TL, Nielsen D, Moore SS, Steele MR, Barendse W, Green RD, Georges M. 1994. Chromosomal localization of the callipyge gene in sheep (Ovis aries) using bovine DNA markers. *Proc Natl Acad Sci U S A* 91(8): 3019-3023.

Conboy IM, Conboy MJ, Smythe GM, Rando TA. 2003. Notch-mediated restoration of regenerative potential to aged muscle. *Science* 302(5650): 1575-1577.

Croteau S, Charron MC, Latham KE, Naumova AK. 2003. Alternative splicing and imprinting control of the Meg3/Gtl2-Dlk1 locus in mouse embryos. *Mamm Genome* 14(4): 231-241.

da Rocha ST, Charalambous M, Lin SP, Gutteridge I, Ito Y, Gray D, Dean W, Ferguson-Smith AC. 2009. Gene dosage effects of the imprinted delta-like homologue 1 (dlk1/pref1) in development: implications for the evolution of imprinting. *PLoS Genet* 5(2): e1000392.

da Rocha ST, Edwards CA, Ito M, Ogata T, Ferguson-Smith AC. 2008. Genomic imprinting at the mammalian Dlk1-Dio3 domain. *Trends Genet* 24(6): 306-316.

Davis E, Caiment F, Tordoir X, Cavaille J, Ferguson-Smith A, Cockett N, Georges M, Charlier C. 2005. RNAi-mediated allelic trans-interaction at the imprinted Rtl1/Peg11 locus. *Curr Biol* 15(8): 743-749.

Davis E, Jensen CH, Schroder HD, Farnir F, Shay-Hadfield T, Kliem A, Cockett N, Georges M, Charlier C. 2004. Ectopic expression of DLK1 protein in skeletal muscle of padumnal heterozygotes causes the callipyge phenotype. *Curr Biol* 14(20): 1858-1862.

Dean A. 2006. On a chromosome far, far away: LCRs and gene expression. *Trends Genet* 22(1): 38-45.

Deguchi M, Hata Y, Takeuchi M, Ide N, Hirao K, Yao I, Irie M, Toyoda A, Takai Y. 1998. BEGAIN (brain-enriched guanylate kinase-associated protein), a novel neuronal PSD-95/SAP90-binding protein. *J Biol Chem* 273(41): 26269-26272.

Dekker J, Rippe K, Dekker M, Kleckner N. 2002. Capturing chromosome conformation. *Science* 295(5558): 1306-1311.

Delgado EF, Geesink GH, Marchello JA, Goll DE, Koohmaraie M. 2001. The calpain system in three muscles of normal and callipyge sheep. *J Anim Sci* 79(2): 398-412.

Duggleby RG, Kinns H, Rood JI. 1981. A computer program for determining the size of DNA restriction fragments. *Anal Biochem* 110(1): 49-55.

Elcheva I, Goswami S, Noubissi FK, Spiegelman VS. 2009. CRD-BP protects the coding region of betaTrCP1 mRNA from miR-183-mediated degradation. *Mol Cell* 35(2): 240-246.

Ender C, Krek A, Friedlander MR, Beitzinger M, Weinmann L, Chen W, Pfeffer S, Rajewsky N, Meister G. 2008. A human snoRNA with microRNA-like functions. *Mol Cell* 32(4): 519-528.

Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30(7): 1575-1584.

Everts AK, Wulf DM, Wheeler TL, Everts AJ, Weaver AD, Daniel JA. 2010. Enhancement Technology Improves Palatability of Normal and Callipyge Lamb. *J Anim Sci*.

Fabian MR, Mathonnet G, Sundermeier T, Mathys H, Zipprich JT, Svitkin YV, Rivas F, Jinek M, Wohlschlegel J, Doudna JA et al. 2009. Mammalian miRNA RISC recruits CAF1 and PABP to affect PABP-dependent deadenylation. *Mol Cell* 35(6): 868-880.

Fejes AP, Robertson G, Bilenky M, Varhol R, Bainbridge M, Jones SJ. 2008. FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. *Bioinformatics* 24(15): 1729-1730.

Filipowicz W, Bhattacharyya SN, Sonenberg N. 2008. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? *Nat Rev Genet* 9(2): 102-114.

Floridon C, Jensen CH, Thorsen P, Nielsen O, Sunde L, Westergaard JG, Thomsen SG, Teisner B. 2000. Does fetal antigen 1 (FA1) identify cells with regenerative, endocrine and neuroendocrine potentials? A study of FA1 in embryonic, fetal, and placental tissue and in maternal circulation. *Differentiation* 66(1): 49-59.

Flynt AS, Lai EC. 2008. Biological principles of microRNA-mediated regulation: shared themes amid diversity. *Nat Rev Genet* 9(11): 831-842.

Freking BA, Keele JW, Beattie CW, Kappes SM, Smith TP, Sonstegard TS, Nielsen MK, Leymaster KA. 1998. Evaluation of the ovine callipyge locus: I. Relative chromosomal position and gene action. *J Anim Sci* 76(8): 2062-2071.

Freking BA, Keele JW, Nielsen MK, Leymaster KA. 1998. Evaluation of the ovine callipyge locus: II. Genotypic effects on growth, slaughter, and carcass traits. *J Anim Sci* 76(10): 2549-2559.

Freking BA, Murphy SK, Wylie AA, Rhodes SJ, Keele JW, Leymaster KA, Jirtle RL, Smith TP. 2002. Identification of the single base change causing the callipyge muscle hypertrophy phenotype, the only known example of polar overdominance in mammals. *Genome Res* 12(10): 1496-1506.

Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N. 2008. Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol* 26(4): 407-415.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* 19(1): 92-105.

Ge B, Gurd S, Gaudin T, Dore C, Lepage P, Harmsen E, Hudson TJ, Pastinen T. 2005. Survey of allelic expression using EST mining. *Genome Res* 15(11): 1584-1591.

Georges M, Charlier C, Cockett N. 2003. The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. *Trends Genet* 19(5): 248-252.

Georges M, Charlier C, Smit M, Davis E, Shay T, Tordoir X, Takeda H, Caiment F, Cockett N. 2004. Toward molecular understanding of polar overdominance at the ovine callipyge locus. *Cold Spring Harb Symp Quant Biol* 69: 477-483.

Georges M, Gunawardana A, Threadgill DW, Lathrop M, Olsaker I, Mishra A, Sargeant LL, Schoeberlein A, Steele MR, Terry C et al. 1991. Characterization of a set of variable number of tandem repeat markers conserved in bovidae. *Genomics* 11(1): 24-32.

Georgiades P, Watkins M, Burton GJ, Ferguson-Smith AC. 2001. Roles for genomic imprinting and the zygotic genome in placental development. *Proc Natl Acad Sci U S A* 98(8): 4522-4527.

Georgiades P, Watkins M, Surani MA, Ferguson-Smith AC. 2000. Parental origin-specific developmental defects in mice with uniparental disomy for chromosome 12. *Development* 127(21): 4719-4728.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J et al. 2005. Galaxy: a platform for interactive large-scale genome analysis. *Genome Res* 15(10): 1451-1455.

Glazov EA, Kongsuwan K, Assavalapsakul W, Horwood PF, Mitter N, Mahony TJ. 2009. Repertoire of bovine miRNA and miRNA-like small regulatory RNAs expressed upon viral infection. *PLoS One* 4(7): e6349.

Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C. 2010. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 329(5992): 643-648.

Gregory PD, Barbaric S, Horz W. 1999. Restriction nucleases as probes for chromatin structure. *Methods Mol Biol* 119: 417-425.

Gregory RI, Feil R. 1999. Analysis of chromatin in limited numbers of cells: a PCR-SSCP based assay of allele-specific nuclease sensitivity. *Nucleic Acids Res* 27(22): e32.

Gregory RI, Khosla S, Feil R. 2001. Probing chromatin structure with nuclease sensitivity assays. *Methods Mol Biol* 181: 269-284.

Gribnau J, Diderich K, Pruzina S, Calzolari R, Fraser P. 2000. Intergenic transcription and developmental remodeling of chromatin subdomains in the human beta-globin locus. *Mol Cell* 5(2): 377-386.

Griffiths-Jones S. 2006. miRBase: the microRNA sequence database. *Methods Mol Biol* 342: 129-138.

Grimson A, Farh KK, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* 27(1): 91-105.

Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A. 2009. BioMart Central Portal--unified access to biological data. *Nucleic Acids Res* 37(Web Server issue): W23-27.

Haig D. 2000. Genomic imprinting, sex-biased dispersal, and social behavior. *Ann N Y Acad Sci* 907: 149-163.

Haig D. 2004. Genomic imprinting and kinship: how good is the evidence? *Annu Rev Genet* 38: 553-585.

Haig D, Wilczek A. 2006. Sexual conflict and the alternation of haploid and diploid generations. *Philos Trans R Soc Lond B Biol Sci* 361(1466): 335-343.

Haley CS, Knott SA, Elsen JM. 1994. Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* 136(3): 1195-1207.

Hatada I, Morita S, Obata Y, Sotomaru Y, Shimoda M, Kono T. 2001. Identification of a new imprinted gene, Rian, on mouse chromosome 12 by fluorescent differential display screening. *J Biochem* 130(2): 187-190.

Haussecker D, Proudfoot NJ. 2005. Dicer-dependent turnover of intergenic transcripts from the human beta-globin gene cluster. *Mol Cell Biol* 25(21): 9724-9733.

Henikoff S, Ahmad K. 2005. Assembly of variant histones into chromatin. *Annu Rev Cell Dev Biol* 21: 133-153.

Hernandez A, Fiering S, Martinez E, Galton VA, St Germain D. 2002. The gene locus encoding iodothyronine deiodinase type 3 (Dio3) is imprinted in the fetus and expresses antisense transcripts. *Endocrinology* 143(11): 4483-4486.

Hernandez A, Martinez ME, Fiering S, Galton VA, St Germain D. 2006. Type 3 deiodinase is critical for the maturation and function of the thyroid axis. *J Clin Invest* 116(2): 476-484.

Hernandez A, Park JP, Lyon GJ, Mohandas TK, St Germain DL. 1998. Localization of the type 3 iodothyronine deiodinase (DIO3) gene to human chromosome 14q32 and mouse chromosome 12F1. *Genomics* 53(1): 119-121.

Hill JJ, Davies MV, Pearson AA, Wang JH, Hewick RM, Wolfman NM, Qiu Y. 2002. The myostatin propeptide and the follistatin-related gene are inhibitory binding proteins of myostatin in normal serum. *J Biol Chem* 277(43): 40735-40741.

Hiraizumi Y, Crow JF. 1960. Heterozygous Effects on Viability, Fertility, Rate of Development, and Longevity of Drosophila Chromosomes That Are Lethal When Homozygous. *Genetics* 45(8): 1071-1083.

Hirsinger E, Malapert P, Dubrulle J, Delfini MC, Duprez D, Henrique D, Ish-Horowicz D, Pourquie O. 2001. Notch signalling acts in postmitotic avian myogenic cells to control MyoD activation. *Development* 128(1): 107-116.

Jackson SP, Miller MF, Green RD. 1997. Phenotypic characterization of rambouillet sheep expression the callipyge gene: III. Muscle weights and muscle weight distribution. *J Anim Sci* 75(1): 133-138.

John B, Enright AJ, Aravin A, Tuschl T, Sander C, Marks DS. 2004. Human MicroRNA targets. *PLoS Biol* 2(11): e363.

Johnson PL, McEwan JC, Dodds KG, Purchas RW, Blair HT. 2005. A directed search in the region of GDF8 for quantitative trait loci affecting carcass traits in Texel sheep. *J Anim Sci* 83(9): 1988-2000.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110(1-4): 462-467.

Kagami M, O'Sullivan MJ, Green AJ, Watabe Y, Arisaka O, Masawa N, Matsuoka K, Fukami M, Matsubara K, Kato F et al. 2010. The IG-DMR and the MEG3-DMR at human chromosome 14q32.2: hierarchical interaction and distinct functional properties as imprinting control centers. *PLoS Genet* 6(6): e1000992.

Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. 2007. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* 315(5815): 1137-1140.

Kennison JA, Southworth JW. 2002. Transvection in Drosophila. *Adv Genet* 46: 399-420.

Keys JR, Tallack MR, Zhan Y, Papathanasiou P, Goodnow CC, Gaensler KM, Crossley M, Dekker J, Perkins AC. 2008. A mechanism for Ikaros regulation of human globin gene switching. *Br J Haematol* 141(3): 398-406.

Kim DD, Kim TT, Walsh T, Kobayashi Y, Matise TC, Buyske S, Gabriel A. 2004. Widespread RNA editing of embedded alu elements in the human transcriptome. *Genome Res* 14(9): 1719-1725.

Koohmaraie M, Shackelford SD, Wheeler TL. 1998. Effect of prerigor freezing and postrigor calcium chloride injection on the tenderness of callipyge longissimus. *J Anim Sci* 76(5): 1427-1432.

Koohmaraie M, Shackelford SD, Wheeler TL, Lonergan SM, Doumit ME. 1995. A muscle hypertrophy condition in lamb (callipyge): characterization of effects on muscle growth and meat quality traits. *J Anim Sci* 73(12): 3596-3607.

Laborda J. 2000. The role of the epidermal growth factor-like protein dlk in cell differentiation. *Histol Histopathol* 15(1): 119-129.

Laborda J, Sausville EA, Hoffman T, Notario V. 1993. dlk, a putative mammalian homeotic gene differentially expressed in small cell lung carcinoma and neuroendocrine tumor cell line. *J Biol Chem* 268(6): 3817-3820.

Lander ES, Green P. 1987. Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci U S A* 84(8): 2363-2367.

Landgraf P, Rusu M, Sheridan R, Sewer A, Iovino N, Aravin A, Pfeffer S, Rice A, Kamphorst AO, Landthaler M et al. 2007. A mammalian microRNA expression atlas based on small RNA library sequencing. *Cell* 129(7): 1401-1414.

Laville E, Bouix J, Sayd T, Bibe B, Elsen JM, Larzul C, Eychenne F, Marcq F, Georges M. 2004. Effects of a quantitative trait locus for muscle hypertrophy from Belgian Texel sheep on carcass conformation and muscularity. *J Anim Sci* 82(11): 3128-3137.

Lee I, Ajay SS, Yook JI, Kim HS, Hong SH, Kim NH, Dhanasekaran SM, Chinnaiyan AM, Athey BD. 2009. New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites. *Genome Res* 19(7): 1175-1183.

Lee RC, Feinbaum RL, Ambros V. 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75(5): 843-854.

Lee SJ, McPherron AC. 1999. Myostatin and the control of skeletal muscle mass. *Curr Opin Genet Dev* 9(5): 604-607.

Lee Y, Ahn C, Han J, Choi H, Kim J, Yim J, Lee J, Provost P, Radmark O, Kim S et al. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* 425(6956): 415-419.

Lestrade L, Weber MJ. 2006. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* 34(Database issue): D158-162.

Levanon EY, Eisenberg E, Yelin R, Nemzer S, Hallegger M, Shemesh R, Fligelman ZY, Shoshan A, Pollock SR, Sztybel D et al. 2004. Systematic identification of abundant A-to-I editing sites in the human transcriptome. *Nat Biotechnol* 22(8): 1001-1005.

Li Q, Peterson KR, Fang X, Stamatoyannopoulos G. 2002. Locus control regions. *Blood* 100(9): 3077-3086.

Li Q, Zhang M, Han H, Rohde A, Stamatoyannopoulos G. 2002. Evidence that DNase I hypersensitive site 5 of the human beta-globin locus control region functions as a chromosomal insulator in transgenic mice. *Nucleic Acids Res* 30(11): 2484-2491.

Li R, Li Y, Kristiansen K, Wang J. 2008. SOAP: short oligonucleotide alignment program. *Bioinformatics* 24(5): 713-714.

Lim LP, Lau NC, Garrett-Engele P, Grimson A, Schelter JM, Castle J, Bartel DP, Linsley PS, Johnson JM. 2005. Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* 433(7027): 769-773.

Lim LP, Lau NC, Weinstein EG, Abdelhakim A, Yekta S, Rhoades MW, Burge CB, Bartel DP. 2003. The microRNAs of Caenorhabditis elegans. *Genes Dev* 17(8): 991-1008.

Lin SP, Youngson N, Takada S, Seitz H, Reik W, Paulsen M, Cavaille J, Ferguson-Smith AC. 2003. Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the Dlk1-Gtl2 imprinted cluster on mouse chromosome 12. *Nat Genet* 35(1): 97-102.

Linsen SE, de Wit E, Janssens G, Heater S, Chapman L, Parkin RK, Fritz B, Wyman SK, de Bruijn E, Voest EE et al. 2009. Limitations and possibilities of small RNA digital gene expression profiling. *Nat Methods* 6(7): 474-476.

Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y et al. 2009. Bos taurus genome assembly. *BMC Genomics* 10: 180.

Luedi PP, Hartemink AJ, Jirtle RL. 2005. Genome-wide prediction of imprinted murine genes. *Genome Res* 15(6): 875-884.

Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U. 2004. Nuclear export of microRNA precursors. *Science* 303(5654): 95-98.

Mansfield JH, Harfe BD, Nissen R, Obenauer J, Srineel J, Chaudhuri A, Farzan-Kashani R, Zuker M, Pasquinelli AE, Ruvkun G et al. 2004. MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. *Nat Genet* 36(10): 1079-1083.

Masternak K, Peyraud N, Krawczyk M, Barras E, Reith W. 2003. Chromatin remodeling and extragenic transcription at the MHC class II locus control region. *Nat Immunol* 4(2): 132-137.

McCaskill JS. 1990. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers* 29(6-7): 1105-1119.

McGrath J, Solter D. 1984. Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* 37(1): 179-183.

Moon YS, Smas CM, Lee K, Villena JA, Kim KH, Yun EJ, Sul HS. 2002. Mice lacking paternally expressed Pref-1/Dlk1 display growth retardation and accelerated adiposity. *Mol Cell Biol* 22(15): 5585-5592.

Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M et al. 2008. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res* 18(4): 610-621.

Murphy SK, Freking BA, Smith TP, Leymaster K, Nolan CM, Wylie AA, Evans HK, Jirtle RL. 2005. Abnormal postnatal maintenance of elevated DLK1 transcript levels in callipyge sheep. *Mamm Genome* 16(3): 171-183.

Murphy SK, Nolan CM, Huang Z, Kucera KS, Freking BA, Smith TP, Leymaster KA, Weidman JR, Jirtle RL. 2006. Callipyge mutation affects gene expression in cis: a potential role for chromatin structure. *Genome Res* 16(3): 340-346.

Nelson PT, Hatzigeorgiou AG, Mourelatos Z. 2004. miRNP:mRNA association in polyribosomes in a human neuronal cell line. *RNA* 10(3): 387-394.

Nilsen TW. 2008. Endo-siRNAs: yet another layer of complexity in RNA silencing. *Nat Struct Mol Biol* 15(6): 546-548.

Ohler U, Yekta S, Lim LP, Bartel DP, Burge CB. 2004. Patterns of flanking sequence conservation and a characteristic upstream motif for microRNA gene identification. *RNA* 10(9): 1309-1322.

Okamoto M, Takemori H, Halder SK, Hatano O. 1997. Zona glomerulosa-specific factor: cloning and function. *Steroids* 62(1): 73-76.

Ozsolak F, Poling LL, Wang Z, Liu H, Liu XS, Roeder RG, Zhang X, Song JS, Fisher DE. 2008. Chromatin structure analyses identify miRNA promoters. *Genes Dev* 22(22): 3172-3183.

Pashev IG, Dimitrov SI, Angelov D. 1991. Crosslinking proteins to nucleic acids by ultraviolet laser irradiation. *Trends Biochem Sci* 16(9): 323-326.

Paulsen M, Ferguson-Smith AC. 2001. DNA methylation in genomic imprinting, development, and disease. *J Pathol* 195(1): 97-110.

Paz-Yaacov N, Levanon EY, Nevo E, Kinar Y, Harmelin A, Jacob-Hirsch J, Amariglio N, Eisenberg E, Rechavi G. 2010. Adenosine-to-inosine RNA editing shapes transcriptome diversity in primates. *Proc Natl Acad Sci U S A* 107(27): 12174-12179.

Perkins AC, Kramer LN, Spurlock DM, Hadfield TS, Cockett NE, Bidwell CA. 2006. Postnatal changes in the expression of genes located in the callipyge region in sheep skeletal muscle. *Anim Genet* 37(6): 535-542.

Rathjen T, Pais H, Sweetman D, Moulton V, Munsterberg A, Dalmay T. 2009. High throughput sequencing of microRNAs in chicken somites. *FEBS Lett* 583(9): 1422-1426.

Reik W, Constancia M, Fowden A, Anderson N, Dean W, Ferguson-Smith A, Tycko B, Sibley C. 2003. Regulation of supply and demand for maternal nutrients in mammals by imprinted genes. *J Physiol* 547(Pt 1): 35-44.

Renard JP, Baldacci P, Richoux-Duranthon V, Pournin S, Babinet C. 1994. A maternal factor affecting mouse blastocyst formation. *Development* 120(4): 797-802.

Rice P, Longden I, Bleasby A. 2000. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 16(6): 276-277.

Rogan DF, Cousins DJ, Santangelo S, Ioannou PA, Antoniou M, Lee TH, Staynov DZ. 2004. Analysis of intergenic transcription in the human IL-4/IL-13 gene cluster. *Proc Natl Acad Sci U S A* 101(8): 2446-2451.

Runge S, Nielsen FC, Nielsen J, Lykke-Andersen J, Wewer UM, Christiansen J. 2000. H19 RNA binds four molecules of insulin-like growth factor II mRNA-binding protein. *J Biol Chem* 275(38): 29562-29569.

Salvatore D, Low SC, Berry M, Maia AL, Harney JW, Croteau W, St Germain DL, Larsen PR. 1995. Type 3 lodothyronine deiodinase: cloning, in vitro expression, and functional analysis of the placental selenoenzyme. *J Clin Invest* 96(5): 2421-2430.

Schmidt JV, Matteson PG, Jones BK, Guan XJ, Tilghman SM. 2000. The Dlk1 and Gtl2 genes are linked and reciprocally imprinted. *Genes Dev* 14(16): 1997-2002.

Schmitt S, Prestel M, Paro R. 2005. Intergenic transcription through a polycomb group response element counteracts silencing. *Genes Dev* 19(6): 697-708.

Schuelke M, Wagner KR, Stolz LE, Hubner C, Riebel T, Komen W, Braun T, Tobin JF, Lee SJ. 2004. Myostatin mutation associated with gross muscle hypertrophy in a child. *N Engl J Med* 350(26): 2682-2688.

Schuster-Gossler K, Simon-Chazottes D, Guenet JL, Zachgo J, Gossler A. 1996. Gtl2lacZ, an insertional mutation on mouse chromosome 12 with parental origin-dependent phenotype. *Mamm Genome* 7(1): 20-24.

Seaton G, Haley CS, Knott SA, Kearsey M, Visscher PM. 2002. QTL Express: mapping quantitative trait loci in simple and complex pedigrees. *Bioinformatics* 18(2): 339-340.

Segers K, Vaiman D, Berghmans S, Shay T, Meyers S, Beever J, Cockett N, Georges M, Charlier C. 2000. Construction and characterization of an ovine BAC contig spanning the callipyge locus. *Anim Genet* 31(6): 352-359.

Seitz H, Royo H, Bortolin ML, Lin SP, Ferguson-Smith AC, Cavaille J. 2004. A large imprinted microRNA gene cluster at the mouse Dlk1-Gtl2 domain. *Genome Res* 14(9): 1741-1748.

Seitz H, Youngson N, Lin SP, Dalbert S, Paulsen M, Bachellerie JP, Ferguson-Smith AC, Cavaille J. 2003. Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nat Genet* 34(3): 261-262.

Sekita Y, Wagatsuma H, Nakamura K, Ono R, Kagami M, Wakisaka N, Hino T, Suzuki-Migishima R, Kohda T, Ogura A et al. 2008. Role of retrotransposon-derived imprinted gene, Rtl1, in the feto-maternal interface of mouse placenta. *Nat Genet* 40(2): 243-248.

Selbach M, Schwanhausser B, Thierfelder N, Fang Z, Khanin R, Rajewsky N. 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* 455(7209): 58-63.

Shackelford SD, Wheeler TL, Koohmaraie M. 2004. Evaluation of sampling, cookery, and shear force protocols for objective evaluation of lamb longissimus tenderness. *J Anim Sci* 82(3): 802-807.

Shay TL, Berghmans S, Segers K, Meyers S, Beever JE, Womack JE, Georges M, Charlier C, Cockett NE. 2001. Fine-mapping and construction of a bovine contig spanning the ovine callipyge locus. *Mamm Genome* 12(2): 141-149.

Sheng M. 2001. The postsynaptic NMDA-receptor--PSD-95 signaling complex in excitatory synapses of the brain. *J Cell Sci* 114(Pt 7): 1251.

Shin C, Nam JW, Farh KK, Chiang HR, Shkumatava A, Bartel DP. 2010. Expanding the microRNA targeting code: functional sites with centered pairing. *Mol Cell* 38(6): 789-802.

Smit M, Segers K, Carrascosa LG, Shay T, Baraldi F, Gyapay G, Snowder G, Georges M, Cockett N, Charlier C. 2003. Mosaicism of Solid Gold supports the causality of a noncoding A-to-G transition in the determinism of the callipyge phenotype. *Genetics* 163(1): 453-456.

Smit MA, Tordoir X, Gyapay G, Cockett NE, Georges M, Charlier C. 2005. BEGAIN: a novel imprinted gene that generates paternally expressed transcripts in a tissue- and promoter-specific manner in sheep. *Mamm Genome* 16(10): 801-814.

Song G, Wang L. 2008. MiR-433 and miR-127 arise from independent overlapping primary transcripts encoded by the miR-433-127 locus. *PLoS One* 3(10): e3574.

Stalder J, Larsen A, Engel JD, Dolan M, Groudine M, Weintraub H. 1980. Tissue-specific DNA cleavages in the globin chromatin domain introduced by DNAase I. *Cell* 20(2): 451-460.

Steshina EY, Carr MS, Glick EA, Yevtodiyenko A, Appelbe OK, Schmidt JV. 2006. Loss of imprinting at the Dlk1-Gtl2 locus caused by insertional mutagenesis in the Gtl2 5' region. *BMC Genet* 7: 44.

Sullivan CS, Grundhoff AT, Tevethia S, Pipas JM, Ganem D. 2005. SV40-encoded microRNAs regulate viral gene expression and reduce susceptibility to cytotoxic T cells. *Nature* 435(7042): 682-686.

Takada S, Paulsen M, Tevendale M, Tsai CE, Kelsey G, Cattanach BM, Ferguson-Smith AC. 2002. Epigenetic analysis of the Dlk1-Gtl2 imprinted domain on mouse chromosome 12: implications for imprinting control from comparison with Igf2-H19. *Hum Mol Genet* 11(1): 77-86.

Takada S, Tevendale M, Baker J, Georgiades P, Campbell E, Freeman T, Johnson MH, Paulsen M, Ferguson-Smith AC. 2000. Delta-like and gtl2 are reciprocally expressed, differentially methylated linked imprinted genes on mouse chromosome 12. *Curr Biol* 10(18): 1135-1138.

Takahashi N, Kobayashi R, Kono T. 2010. Restoration of Dlk1 and Rtl1 is necessary but insufficient to rescue lethality in intergenic differentially methylated region (IG-DMR)-deficient mice. *J Biol Chem* 285(34): 26121-26125.

Takahashi N, Okamoto A, Kobayashi R, Shirai M, Obata Y, Ogawa H, Sotomaru Y, Kono T. 2009. Deletion of Gtl2, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice. *Hum Mol Genet* 18(10): 1879-1888.

Takeda H, Caiment F, Smit M, Hiard S, Tordoir X, Cockett N, Georges M, Charlier C. 2006. The callipyge mutation enhances bidirectional long-range DLK1-GTL2 intergenic transcription in cis. *Proc Natl Acad Sci U S A* 103(21): 8119-8124.

Takeda H, Charlier C, Farnir F, Georges M. 2010. Demonstrating polymorphic miRNA-mediated gene regulation in vivo: application to the g+6223G->A mutation of Texel sheep. *RNA* 16(9): 1854-1863.

Tay Y, Zhang J, Thomson AM, Lim B, Rigoutsos I. 2008. MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation. *Nature* 455(7216): 1124-1128.

Taylor RG, Koohmaraie M. 1998. Effects of postmortem storage on the ultrastructure of the endomysium and myofibrils in normal and callipyge longissimus. *J Anim Sci* 76(11): 2811-2817.

Terwilliger JD. 1995. A powerful likelihood method for the analysis of linkage disequilibrium between trait loci and one or more polymorphic marker loci. *Am J Hum Genet* 56(3): 777-787.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22(22): 4673-4680.

Tierling S, Gasparoni G, Youngson N, Paulsen M. 2009. The Begain gene marks the centromeric boundary of the imprinted region on mouse chromosome 12. *Mamm Genome* 20(9-10): 699-710.

Tobin JF, Celeste AJ. 2005. Myostatin, a negative regulator of muscle mass: implications for muscle degenerative diseases. *Curr Opin Pharmacol* 5(3): 328-332.

Trivers R, Burt A. 1999. Kinship and genomic imprinting. *Results Probl Cell Differ* 25: 1-21.

Uejima H, Lee MP, Cui H, Feinberg AP. 2000. Hot-stop PCR: a simple and general assay for linear quantitation of allele ratios. *Nat Genet* 25(4): 375-376.

Vakoc CR, Letting DL, Gheldof N, Sawado T, Bender MA, Groudine M, Weiss MJ, Dekker J, Blobel GA. 2005. Proximity among distant regulatory elements at the beta-globin locus requires GATA-1 and FOG-1. *Mol Cell* 17(3): 453-462.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* 3(7): RESEARCH0034.

Vasudevan S, Tong Y, Steitz JA. 2007. Switching from repression to activation: microRNAs can up-regulate translation. *Science* 318(5858): 1931-1934.

Vuocolo T, Byrne K, White J, McWilliam S, Reverter A, Cockett NE, Tellam RL. 2007. Identification of a gene network contributing to hypertrophy in callipyge skeletal muscle. *Physiol Genomics* 28(3): 253-272.

Wagner RW, Smith JE, Cooperman BS, Nishikura K. 1989. A double-stranded RNA unwinding activity introduces structural alterations by means of adenosine to inosine conversions in mammalian cells and Xenopus eggs. *Proc Natl Acad Sci U S A* 86(8): 2647-2651.

Wakasugi N. 1974. A genetically determined incompatibility system between spermatozoa and eggs leading to embryonic death in mice. *J Reprod Fertil* 41(1): 85-96.

Wakasugi N. 2007. Embryologic, cytobiologic and genetic interpretations of DDK syndrome in mice. *Dev Growth Differ* 49(7): 555-559.

Wakasugi N, Tomita T, Kondo K. 1967. Differences of fertility in reciprocal crosses between inbred strains of mice. DDK, KK and NC. *J Reprod Fertil* 13(1): 41-50.

Wang Q, Khillan J, Gadue P, Nishikura K. 2000. Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. *Science* 290(5497): 1765-1768.

Wang Y, Sul HS. 2006. Ectodomain shedding of preadipocyte factor 1 (Pref-1) by tumor necrosis factor alpha converting enzyme (TACE) and inhibition of adipocyte differentiation. *Mol Cell Biol* 26(14): 5421-5435.

Wei B, Cai T, Zhang R, Li A, Huo N, Li S, Gu YQ, Vogel J, Jia J, Qi Y et al. 2009. Novel microRNAs uncovered by deep sequencing of small RNA transcriptomes in bread wheat (Triticum aestivum L.) and Brachypodium distachyon (L.) Beauv. *Funct Integr Genomics* 9(4): 499-511.

Wightman B, Ha I, Ruvkun G. 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. *Cell* 75(5): 855-862.

Wilkins JF, Haig D. 2003. Inbreeding, maternal care and genomic imprinting. *J Theor Biol* 221(4): 559-564.

Wilkins JF, Haig D. 2003. What good is genomic imprinting: the function of parent-specific gene expression. *Nat Rev Genet* 4(5): 359-368.

Wood AJ, Oakey RJ. 2006. Genomic imprinting in mammals: emerging themes and established theories. *PLoS Genet* 2(11): e147.

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES, Kellis M. 2005. Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* 434(7031): 338-345.

Yang W, Chendrimada TP, Wang Q, Higuchi M, Seeburg PH, Shiekhattar R, Nishikura K. 2006. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol* 13(1): 13-21.

Yekta S, Shih IH, Bartel DP. 2004. MicroRNA-directed cleavage of HOXB8 mRNA. *Science* 304(5670): 594-596.

Youngson NA, Kocialkowski S, Peel N, Ferguson-Smith AC. 2005. A small family of sushi-class retrotransposon-derived genes in mammals and their relation to genomic imprinting. *J Mol Evol* 61(4): 481-490.

Zamore PD, Haley B. 2005. Ribo-gnome: the big world of small RNAs. *Science* 309(5740): 1519-1524.

Zhao Y, Samal E, Srivastava D. 2005. Serum response factor regulates a muscle-specific microRNA that targets Hand2 during cardiogenesis. *Nature* 436(7048): 214-220.

Zhou Y, Cheunsuchon P, Nakayama Y, Lawlor MW, Zhong Y, Rice KA, Zhang L, Zhang X, Gordon FE, Lidov HG et al. 2010. Activation of paternally expressed genes and perinatal death caused by deletion of the Gtl2 gene. *Development* 137(16): 2643-2652.

Zimmers TA, Davies MV, Koniaris LG, Haynes P, Esquela AF, Tomkinson KN, McPherron AC, Wolfman NM, Lee SJ. 2002. Induction of cachexia in mice by systemically administered myostatin. *Science* 296(5572): 1486-1488.