DNA Research, 2017, 24(3), 221–233
doi: 10.1093/dnares/dsx004
Advance Access Publication Date: 23 February 2017
Full Paper

OXFORD

Full Paper

Genetic variability of the activity of bidirectional promoters: a pilot study in bovine muscle

Cédric Meersseman[1,2], Rabia Letaief[1], Véronique Léjard[1], Emmanuelle Rebours[1], Gabriel Guillocheau[1], Diane Esquerré[3], Anis Djari[4], Amanda Chamberlain[5,6], Christy Vander Jagt[5,6], Christophe Klopp[4], Mekki Boussaha[1], Gilles Renand[1], Abderrahman Maftah[2], Daniel Petit[2], and Dominique Rocha[1,*]

[1]GABI, INRA, AgroParisTech, Université Paris-Saclay, 78350 Jouy-en-Josas, France, [2]GMA, INRA, Université de Limoges, 87060 Limoges, France, [3]GenPhySE, Université de Toulouse, INRA, INPT, ENVT, 31326 Castanet Tolosan, France, [4]SIGENAE, UR 875, INRA, 31362 Castanet-Tolosan, France, [5]Dairy Futures Cooperative Research Centre, AgriBio, Bundoora, Victoria, Australia, and [6]AgriBio, Department of Economic Development, Jobs, Transport & Resources, Victoria, Australia

*To whom correspondence should be addressed. Tel. +33 134 652 432. Fax. +33 134 652 478.
Email: dominique.rocha@inra.fr

Edited by Dr. Takao Sekiya

Received 2 September 2016; Editorial decision 22 January 2017; Accepted 24 January 2017

Abstract

Bidirectional promoters are regulatory regions co-regulating the expression of two neighbouring genes organized in a head-to-head orientation. In recent years, these regulatory regions have been studied in many organisms; however, no investigation to date has been done to analyse the genetic variation of the activity of this type of promoter regions. In our study, we conducted an investigation to first identify bidirectional promoters sharing genes expressed in bovine *Longissimus thoracis* and then to find genetic variants affecting the activity of some of these bidirectional promoters. Combining bovine gene information and expression data obtained using RNA-Seq, we identified 120 putative bidirectional promoters active in bovine muscle. We experimentally validated *in vitro* 16 of these bidirectional promoters. Finally, using gene expression and whole-genome genotyping data, we explored the variability of the activity in muscle of the identified bidirectional promoters and discovered genetic variants affecting their activity. We found that the expression level of 77 genes is correlated with the activity of 12 bidirectional promoters. We also identified 57 single nucleotide polymorphisms associated with the activity of 5 bidirectional promoters. To our knowledge, our study is the first analysis in any species of the genetic variability of the activity of bidirectional promoters.

Key words: bidirectional promoter, cattle genome, genetic variability, muscle

1. Introduction

Whole-genome analyses have identified a class of regulatory regions that contain elements that initiate transcription of two different genes positioned with a head-to-head (5′–5′) orientation on opposite strands of the DNA. These regions termed 'bidirectional promoters'[1–3] (BiPs) often have fewer than 1,000 base pairs separating the transcription start sites (TSSs) of the two genes. Bidirectional promoters have been known for years;[4] however, recent genome-scale studies have shown that the regulation of the expression of up to 10% of the genes is controlled by bidirectional promoters.[3] Bidirectional promoters are a common feature within not only the human genome but are also present in many other genomes, including yeast, plants, Invertebrates and Vertebrates.[1–3,5–17] It has been shown that some bidirectional promoters, such as those of the histone genes, regulate the transcription of pair of genes that need coordinated expression to maintain stoichiometric relationships,[18] or regulate the coexpression of genes that function in the same biological pathway.[19] Detailed studies on bidirectional promoters have also shown that some are widely conserved among Vertebrates.[20–23] The presence of such genomic architecture in many metazoans and the high inter-species conservation seen for some bidirectional promoters suggest that they are functionally important.

Despite substantial interest in bidirectional promoters, the functional mechanisms underlying the activation of bidirectional promoters are currently not well characterized and, to our knowledge, no analysis of the variability of the activity of bidirectional promoters has been done so far.

The aim of our study was to explore bidirectional promoter activity variation and to identify if genetic variants contribute to this variability. We performed this study in bovine *Longissimus thoracis* (LT). First, we identified bovine bidirectional promoters active in muscle, combining bovine gene information and RNA-Seq data. We then experimentally validated *in vitro* a subset of these bidirectional promoters. Finally, using gene expression and whole-genome genotyping data generated for 20 samples, we explore the variability of the activity in muscle of the identified bidirectional promoters and discovered genetic variants affecting this activity. We identified 120 bovine bidirectional promoters active in LT. A total of 16 bidirectional promoters were selected and all were validated experimentally *in vitro*. As we had RNA-Seq data for 20 different samples from the same tissue, it is possible to analyse the inter-individual variability of the activity of these 120 bidirectional promoters. As previously described, expression levels of gene pairs sharing a BiP are more correlated than those of neighbouring gene pairs in a unidirectional or convergent configuration.

In addition, we found that the expression level of 77 genes is highly correlated with the activity of 12 bidirectional promoters active in muscle. We also identified 57 SNPs associated significantly with the activity of 5 bidirectional promoters.

To our knowledge, our study is the first analysis in any species of the genetic variability of the activity of bidirectional promoters.

2. Materials and methods

2.1. Animals and tissue samples

The study was conducted with 11 Limousin bull calves from a large study on the genetic determinism of beef and meat quality traits.[24] These eleven bull calves were not closely related to one another (for at least four generations). They were fattened in a single feedlot and fed *ad libidum* with wet corn silage. They were humanely slaughtered in an accredited commercial slaughterhouse when they reached 16 months. LT muscle samples were dissected immediately after death and tissue samples were snap frozen in liquid nitrogen and stored at −80 °C until analysis.

2.2. RNA sequencing and data analysis

RNA extraction and sequencing were performed as previously described.[25,26] Briefly, after transfer to ice-cold RNeasy RLT lysis buffer (Qiagen), LT tissue samples were homogenized using a Precellys tissue homogeniser (Bertin Technologie). Total RNA was isolated using RNeasy Midi columns (Qiagen) and then treated with RNAse-free DNase I (Qiagen) for 15 min at room temperature according to the manufacturer's protocols. The concentration of total RNA was measured with a Nanodrop ND-100 instrument (Thermo Scientific) and the quality was assessed with an RNA 6000 Nano Labchip kit using an Agilent 2100 Bioanalyzer (Agilent Technologies). All 11 samples had an RNA integrity number (RIN) value greater than eight.

The mRNA-Seq libraries were prepared using the TruSeq RNA Sample Preparation Kit (Illumina) according to the manufacturer's instructions. Briefly, Poly-A containing mRNA molecules were purified from 4 µg total RNA of each sample using oligo(dT) magnetic beads and fragmented into 150–400 bp pieces using divalent cations at 94 °C for 8 min. The cleaved mRNA fragments were converted to double-stranded cDNA using SuperScript II reverse transcriptase (Life Technologies) and primed by random primers. The resulting cDNA was purified using Agencourt AMPure® XP beads (Beckman Coulter). Then, cDNA was subjected to end-repair and phosphorylation and subsequent purification was performed using Agencourt AMPure® XP beads (Beckman Coulter). These repaired cDNA fragments were 3′-adenylated producing cDNA fragments with a single 'A' base overhung at their 3′-ends for subsequent adapter-ligation. Illumina adapters containing indexing tags were ligated to the ends of these 3′-adenylated cDNA fragments followed by two purification steps using Agencourt AMPure® XP beads (Beckman Coulter). Ten rounds of PCR amplification were performed to enrich the adapter-modified cDNA library using primers complementary to the ends of the adapters. The PCR products were purified using Agencourt AMPure® XP beads (Beckman Coulter) and size-selected (200 ± 25 bp) on a 2% agarose Invitrogen E-Gel (Thermo Scientific). Libraries were then checked on an Agilent Technologies 2100 Bioanalyzer using the Agilent High Sensitivity DNA Kit and quantified by quantitative PCR with the QPCR NGS Library Quantification kit (Agilent Technologies). After quantification, three different tagged cDNA libraries were pooled in equal ratios and a final qPCR check was performed post-pooling. Each library pool was used for 2 × 100 bp paired-end sequencing on one lane of the Illumina HiSeq2000 with a TruSeq SBS v3-HS Kit (Illumina). After sequencing, the samples were demultiplexed and the indexed adapter sequences were trimmed using the CASAVA v1.8.2 software (Illumina).

The *Bos taurus* reference transcriptome was downloaded from Ensembl (version 63, Bos_taurus.Btau_4.0.63.cdna.all.fa). To align the reads back to the assembled reference transcriptome the BWA programme (version 0.5.9-r16) was used.[27] Reads were mapped for each sample separately to the assembled transcriptome. The BWA default values were used for mapping. Properly paired reads with a mapping quality of at least 30 ($-q = 30$) were extracted from the resulting BAM file using SAMtools[28] for further analyses. Properly paired is defined as both left and right reads mapped in opposite

directions on the same transcript at a distance compatible with the expected mean size of the fragments (<500 bp). Custom scripts were developed to identify paired-reads mapping to single locations and with the expected distance. Read pairs mapping to separate chromosomes were discarded for the present study. The number of paired-reads uniquely aligning to transcribed regions of each transcript was calculated for all genes in the annotated transcriptome. The transcript paired-read count was calculated as the number of unique paired-reads that aligned within the exons of each transcript, based on the coordinates of mapped reads.

2.3. SNP identification

BWA was also used to map reads onto the bovine genome assembly UMD3.1 version.[29] Only reliable properly paired BWA mapped reads were considered for SNP calling. Indels were not considered because alternative splicing impedes reliable indel discovery. SNPs were called using the SAMtools software package. Genotype likelihoods were computed using the SAMtools utilities and variable positions in the aligned reads compared with the reference were called with the BCFtools utilities.[30] SNPs were called only for positions with a minimal mapping quality ($-Q$) of 30, a minimum coverage ($-d$) of 4 and a maximum read depth ($-D$) of 10,000,000.

2.4. Data and statistical analyses

Information on bovine genes were obtained from the Ensembl Genome Browser (version 84) using the BioMart tool.[31] Functional annotation analysis of genes sharing putative bidirectional promoters was done using the FATIGO tool of the online software suite Babelomics.[32] Genes were assigned their Ensembl identities as input for Babelomics. Only one copy of each gene was used. Default parameter settings were used for the analysis. Statistical assessment of annotation differences between the two sets of sequences (SNP-containing genes versus all the other bovine genes) was carried out for each FATIGO analysis, using the Fisher Exact Test with correction for multiple testing. Prediction of putative TFBSs was performed using the bioinformatics tools MATCH.[33] It uses a library of positional weight matrices from TRANSFAC 6.0 public version.[34] A cut-off selection of 0.7 and 0.75 was applied for matrix and core similarities, respectively. Spearman's rank correlation coefficients were calculated for the correlation studies using the statistical *R* package.

2.5. Bidirectional promoter validation

Each putative bidirectional promoter region was PCR amplified from bovine genomic DNA using a pair of primers. The genomic DNA sequences were retrieved from the UCSC Genome Database. Sequence repeats were masked using RepeatMasker[35] and Primer3[36] was then used to design primer pairs to amplify each putative bidirectional promoter region. PCR primers were synthesized by Eurofins MWG Operon. Restriction enzyme cutting sites for *Eco*RI (GAATTC) and *Bam*HI (GGATCC) were artificially added into the PCR primers to facilitate directional cloning. Primer sequences are presented in Supplementary Table S1. Polymerase chain reactions were performed in 50 µl using 120 ng genomic DNA, 1 × PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.3 µM of each primer and 1U AccuPrime GC-rich *Taq* DNA polymerase (Invitrogen). The following cycling protocol was used: 95 °C for 15 min, followed by 35 cycles of 95 °C for 1 min, 52 °C for 1 min and 72 °C for 1 min, and a final extension step at 72 °C for 10 min. To check the quality of the amplification, 5 µl of PCR products were then analysed by gel electrophoresis with a 1% agarose gel. The PCR products were purified using the Qiagen MinElute Gel Extraction kit (Qiagen), digested with *Eco*RI and *Bam*HI (New England Biolabs), purified with Qiagen Reaction Cleanup kit, quantified and then ligated using T4 DNA ligase (New England Biolabs) to the pBiP0 vector at a 3:1 ratio. The vector was previously digested with *Eco*RI/*Bam*HI and dephosphorylated with alkaline phosphatase (New England Biolabs). The ligation products were transformed into *Escherichia coli* DH5α competent cells (Invitrogen). Ten clones were then amplified and plasmids were purified with the Pureyield Plasmid Miniprep System DNA purification kit (Promega). Positive clones were identified by digesting plasmid DNA with *Eco*RI and *Bam*HI. Digestion products were then visualized after gel electrophoresis on a 1% agarose gel. One clone carrying the right plasmid construct was then amplified and plasmids were purified with the Macherey-Nagel Midi Endotoxin-free plasmid DNA purification kit. The plasmids were then sequenced bidirectionally using BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems) and the primers used for the PCR amplification. After purification on Sephadex G50 superfine column (GE Healthcare), the sequencing reaction products were analyzed using 3130 Genetic Analyzer sequencer (Applied Biosystems). One clone containing for each cloning orientation the right plasmid construct was then chosen.

Murine C2C12 myoblastic cells (ATCC CRL-1772) were grown in Dulbecco's Modified Eagle's Medium with Glutamax-I (4.5 g/l glucose, Invitrogen) supplemented with 1% penicillin/streptomycin and 20% heat-inactivated fetal bovine serum. Cells were cultured at 37 °C in a humidified atmosphere of 5% CO_2 in air. C2C12 (~25,000 cells/well) were seeded into 24-well plates 1 day before transfection. Transfections were performed on 40% confluent C2C12 cells with TurboFect transfection reagent (Fermentas) using 2 µl of reagent per µg of DNA plasmid, during 4 h according to the manufacturer's protocol. The cells were transfected with 0.8 µg of the vectors carrying unidirectional pCMV1/pCMV2 promoters or putative bidirectional promoters or with pBI-CMV2, pBiP-DsRed, pBiP0 control vectors.[37] All transient transfection experiments were done in triplicate and repeated three times. Thirty-six hours after transfection cells were washed two times with PBS 1×, and then viewed with an Axio Observer Z1 microscope (Zeiss) and images were acquired using a CoolSNAP HQ2 camera (Photometrics) driven by the Axiovision imaging system software. Analyses of micrographies were performed using the AxioVision 4.7.2 software (Zeiss). Filter sets 38HE [excitation BP470/40(HE) and emission BP525/50(HE)] and 43HE [excitation BP545/30(HE) and emission BP620/60(HE)] were used to visualize AcGFP1 and DsRed monomer signals, respectively. Each reporter gene assay was performed in triplicate and with three independent transfection experiments.

2.6. Whole-genome SNP genotyping

Muscle samples were obtained and genomic DNA extracted using the DNA Midi kit (Qiagen). Quality of DNA was checked using a Nanodrop ND-100 spectrophotometer (Thermo Scientific) and quantity was estimated with Quant-iT Picogreen dsDNA kit (Life Technologies) on an ABI 7900HT (Life Technologies). All DNA samples were standardized to 50 ng/µL.

DNA samples were genotyped on the BovineSNP50 Genotyping Beadchips (Illumina) at LABOGENA (Jouy-en-Josas, France) using the standard operating procedures recommended by the manufacturer.

3. Results and discussion

3.1. Bovine muscle whole-transcriptome analysis

To identify bidirectional promoters active in bovine muscle, we first obtained a global view of the bovine LT transcriptome using paired-end RNA sequencing (RNA-Seq) from twenty Limousin bull calves. We used already-published data from nine Limousin animals[25,26] and poly(A)-enriched mRNA from 11 new Limousin bull calves were retrotranscribed and subjected to high-throughput sequencing. The 11 RNA-Seq libraries were barcode-tagged and sequenced on five lanes (3 libraries per lane) of an Illumina HiSeq2000 sequencer. Sequencing of cDNA libraries generated a total of 1,303 million raw paired-end reads with a length of 100 bases, resulting in a total of 130 gigabases. The reads were then de-multiplexed to assign reads to each sequenced sample according to its barcode index. Approximately 17–48 million paired-end reads were obtained for each library. Reads from each sample were then mapped back to the bovine reference transcriptome. We used the set of *Bos taurus* Ensembl transcripts v63 RefSeq genes as the reference transcriptome. This set contains transcripts for 22,915 known or novel genes but also pseudogenes. Based on mappings done using the Burrows–Wheeler Aligner (BWA) programme, 67–81% of the mapped reads were aligned properly paired (Supplementary Table S2). A total of 22,025 transcripts (18,055 genes) were identified, with at least one paired-end read in all samples analysed. Similar RNA-Seq read mapping rate and the number of genes identified were obtained in other RNA-Seq bovine studies. For example, Chitwood et al.[38] found that ~69% of the RNA-Seq reads they generated while sequencing the transcriptomes of single bovine blastocysts mapped uniquely onto the bovine genome. Raw gene expression levels were estimated by measuring the normalized count number for each transcript (number of reads per transcript divided by the total number of mapped reads, for each sample). The five most frequent transcripts are shown in Table 1. These five genes represented nearly 17% of all sequencing reads mapped to the bovine genome and are associated with muscle cell metabolism or structure. These results were consistent with the physiological role of genes expected in the surveyed tissue. Not all genes were expressed among all samples. A total of 14,700 transcripts corresponding to 12,314 different genes were detected in all 20 samples, while 840 different genes (~4%) were only expressed in one sample at a very low level (with on average less than 2 reads).

3.2. Identification of putative bidirectional promoters shared by genes co-expressed in bovine LT

In order to identify putative bidirectional promoters active in bovine muscle, the chromosomal locations of all the bovine genes were retreived from the Ensembl database (version 84).[31] The Ensembl gene start sites were used as the TSS. Pair of genes in a head-to-head arrangement and separated by less than 1 kb between their TSSs were identified and the region between paired-genes were considered as putative bidirectional promoters. Among the 24,616 bovine genes annotated on the UMD3.1 genome assembly, we found 563 putative bidirectional promoters (Supplementary Table S3). The number of bovine bidirectional promoters we predicted is very similar to the one calculated by Xu et al.,[39] who found 574 putative bidirectional promoters in cattle.

To overcome confusion during the subsequent analyses, we then used four filters (Supplementary Fig. S1). First, we removed all gene pairs for which at least one gene had paralogous sequences identified in the Ensembl database and predicted 192 bidirectional promoters. Second, we retained only gene pairs associated with these putative BiPs and for which we could detect the expression for both paired-genes in our RNA-Seq dataset, leaving 171 cases. Third, to avoid quantification problems relative to transcript isoforms, genes with known alternative transcripts were excluded from the study. Finally, for validation purposes, we only selected bidirectional promoters between protein-coding gene pairs. In total, 120 BiPs remained, each with both genes expressed in muscle (Supplementary Table S4).

3.3. Characteristics of genes shared by bidirectional promoters active in bovine LT

To characterise functions that are significantly enriched in genes sharing bidirectional promoters active in bovine muscle, a gene ontology (GO) analysis was performed. All the 120 co-expressed gene pairs could be analysed for functional classifications and were assigned to one or more GO annotations. GO term analysis showed a significant enrichment of specific GO terms when comparing the annotations of gene pairs associated with bidirectional promoters active in muscle against all transcripts from the bovine reference transcriptome. A summary of the classification of these genes into major biological process, cell component and molecular function categories is presented in Supplementary Table S5. Genes encoding proteins found in mitochondria, ribonucleoprotein complexes or organelle inner membrane or involved in chaperone binding and cellular response to DNA damage stimulus are significantly over-represented. No significant enrichment in KEGG terms/pathways was found. These results are in agreement with previous studies.[1–3,5–7,12,40]

The chromosomal location of all bovine genes, of all the predicted bidirectional promoters and of the 120 BiPs active in muscle is presented in Fig. 1. Interestingly, the chromosomal distribution of these 120 bidirectional promoters does not reflect the gene content of the chromosomes (Mann–Whitney *U*-test, P-value <0.05). For example, we found only one (0.83%) predicted BiP active in muscle on BTA4, whereas this chromosome harbours 855 (3.48%) bovine genes. BTA11 and BTA19 have the highest number of bidirectional promoters active in muscle (both 11.67%) whereas these chromosomes have only ~4 and ~5%, respectively, of all bovine genes. The chromosomal location of the 120 BiPs active in muscle rather follows the chromosomal distribution of all bidirectional promoters

Table 1. Top five transcripts with the most assigned reads

Ensembl gene ID	Ensembl transcript ID	Description	Gene symbol	% total number of reads
ENSBTAG00000026986	ENSBTAT00000061449	Titin	TTN	4.46
ENSBTAG00000018204	ENSBTAT00000009327	Myosin heavy chain 1	MYH1	4.30
ENSBTAG00000043561	ENSBTAT00000060569	Cytochrome c oxidase subunit I	COX1	3.61
ENSBTAG00000007090	ENSBTAT00000012797	Myosin heavy chain 2	MYH2	2.36
ENSBTAG00000004965	ENSBTAT00000006534	Nucleoporin 133 kDa	NUP133	1.99

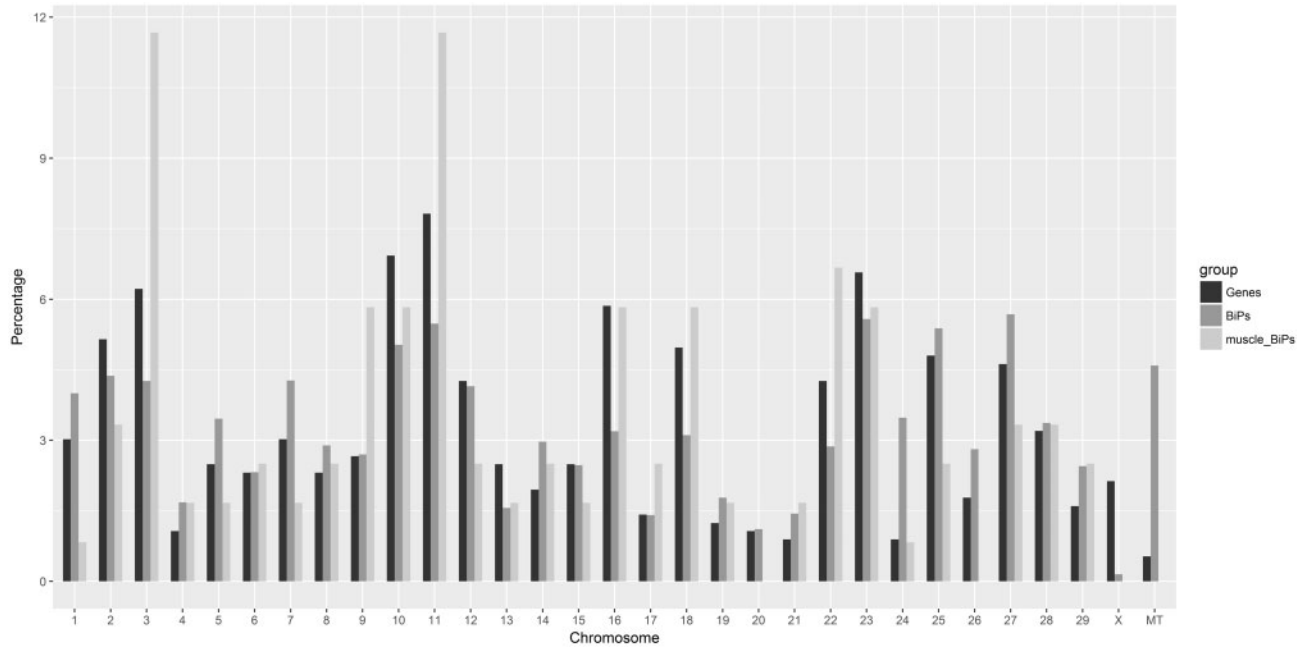

Figure 1. Chromosomal distribution of the 120 putative bovine bidirectional promoters active in muscle.

predicted in cattle (Pearson correlation coefficient $\rho = 0.802$). For example, we found 0.89 and 0.83% predicted bidirectional promoters and BiPs active in muscle, respectively, on chromosome 4.

3.4. Experimental validation of some predicted bidirectional promoters

To assess the promoter activity of the predicted bidirectional promoters, some bidirectional promoters were selected and cloned into the promoterless pBiP0 reporter vector we recently constructed.[37] This plasmid is a dual reporter vector with the *Green Fluorescent Protein* gene from the *Aequorea coerulescens* jellyfish (*AcGFP1*) inserted in a head-to-head configuration relative to a gene expressing a monomer mutant form of the *Red Fluorescent Protein* gene from the *Discosoma striata* reef coral (*DsRed monomer*). Both fluorescent protein genes are devoid of promoter and separated by a short multiple cloning region.

The sequences of predicted bidirectional promoters were downloaded from the UCSC database.[41] To ensure that all regulatory elements were included each sequence was extended until the first ATG (ATG not included) of both paired genes. Among the 120 promoter sequences 18 contain cutting sites for *Eco*RI or *Bam*HI and were eliminated, since these two restriction enzymes are used for cloning into the dual reporter vector. We previously validated one of these bidirectional promoters.[23] Sixteen new bidirectional promoters were then randomly selected for validation. Information relative to these 16 predicted bidirectional promoters is indicated in Table 2. The selected promoter regions were amplified by PCR and then cloned into the pBiP0 vector. Inserts were checked by sequencing.

The promoter activities of these sequences were then assayed after transient transfections in murine myoblastic cells (C2C12) as no bovine muscle cell line was available. The pBI-CMV2 and pBiP-DsRed vectors were used as positive controls, whereas the empty pBiP0 vector was used as a negative control to verify that no reporter gene expression could be observed without promoter. The expression of the reporter genes was assessed by fluorescence microscopy 36 h after transfection to evaluate transfection efficiencies and check reporter activities. No significant reporter gene expression was observed with the empty pBiP0 transfected cells and positive signals were therefore attributed to the inserted sequences. Fluorescence microscopy analyses showed that all the 16 selected predicted bidirectional promoters seem to drive bidirectional expression in C2C12 cells, although variability in the reporter signals was observed (Fig. 2; Supplementary Fig. S2). These good results suggest that the procedure we developed to detect bovine bidirectional promoters active in muscle is accurate and has a low false-positive rate.

3.5. Co-expression analysis of genes shared by bidirectional promoters active in bovine muscle

We have in our study the expression levels of genes expressed in LT for twenty samples. It is therefore possible to analyse the co-expression of each pair of genes sharing a bidirectional promoter. Using the normalized expression levels, we calculated the Spearman's rank correlation coefficient for each gene pair. We found that 34% (41/120) of the genes shared by bidirectional promoters are co-expressed (P-value < 0.05%) (Supplementary Table S6). In order to evaluate whether gene pairs sharing bidirectional promoters were significantly enriched for co-expressed genes, we then performed a similar co-expression analysis for pairs of non-overlapping neighbouring genes in a tail–tail orientation (convergent configuration) or in the same orientation (unidirectional configuration). We identified these two different types of gene pairs using the chromosomal locations of all the bovine genes from the Ensembl database and using the same filters as for the identification of gene pairs sharing BiPs (Supplementary Fig. S1). We found 218 and 163 gene pairs in the convergent and unidirectional configurations (Supplementary Tables S7 and S8), respectively. The Spearman's rank correlation coefficient for each gene pair was then calculated using the normalized expressed levels measured in the 20 samples. Thirty-five percent (76/218) and 30.1% (49/163) of the gene pairs in the convergent and unidirectional configurations, respectively, were co-expressed

Table 2. Information on the 16 tested predicted bovine bidirectional promoters

BiP #	Ensembl gene ID	BTA	Gene start	Gene end	Strand	Description	Gene name	Distance
2	ENSBTAG00000016337	23	32835596	32849025	−1	Acyl-CoA thioesterase 13	THEM2	293
	ENSBTAG00000000365	23	32849318	32861445	1	tyrosyl-DNA phosphodiesterase 2	TDP2	
12	ENSBTAG00000014646	17	74392831	74397072	−1	Mitotic spindle organizing protein 2B	MZT2	143
	ENSBTAG00000002130	17	74397215	74412807	1	sphingomyelin phosphodiesterase 4, neutral membrane (neutral sphingomyelinase-3)	SMPD4	
19	ENSBTAG00000003550	29	1032168	1056464	−1	Chromosome 11 open reading frame 54	C11orf54	163
	ENSBTAG00000003545	29	1056627	1062580	1	TATA box binding protein (TBP)-associated factor, RNA polymerase I, D, 41kDa	TAF1D	
24	ENSBTAG00000004295	11	93011815	93029730	−1	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8, 19kDa	NDUFA8	144
	ENSBTAG00000004296	11	93029874	93064357	1	MORN repeat containing 5	MORN5	
27	ENSBTAG00000004991	8	100163708	100212829	−1	Inhibitor of kappa light polypeptide gene enhancer in B-cells, kinase complex-associated protein	IKBKAP	324
	ENSBTAG00000004996	8	100213153	100219346	1	family with sequence similarity 206, member A	FAM206A	
31	ENSBTAG00000006195	29	44691682	44693901	−1	Chromosome 29 open reading frame, human C11orf68	C29H11orf68	270
	ENSBTAG00000006199	29	44694171	44696295	1	DR1-associated protein 1 (negative cofactor 2 alpha)	DRAP1	
61	ENSBTAG00000023018	28	44283041	44397693	−1	Poly(ADP-ribose) glycohydrolase	PARG	121
	ENSBTAG00000011694	28	44397814	44418582	1	translocase of inner mitochondrial membrane 23 homolog (yeast)	TIMM23	
62	ENSBTAG00000011930	17	63448841	63466323	−1	DEAD (Asp-Glu-Ala-Asp) box polypeptide 54	DDX54	327
	ENSBTAG00000011937	17	63466650	63472165	1	chromosome 17 open reading frame, human C12orf52 (C17H12orf52)	RITA	
66	ENSBTAG00000012586	2	86438979	86449372	−1	Heat shock 60kDa protein 1 (chaperonin)	HSPD1	153
	ENSBTAG00000012589	2	86449525	86451564	1	heat shock 10kDa protein 1 (chaperonin 10)	HSPE1	
81	ENSBTAG00000015667	11	2324293	2335178	−1	Transmembrane protein 127	TMEM127	204
	ENSBTAG00000015659	11	2335382	2342308	1	cytosolic iron-sulfur protein assembly 1	CIAO1	
85	ENSBTAG00000016559	29	43146856	43149460	−1	Nudix (nucleoside diphosphate linked moiety X)-type motif 22	NUDT22	150
	ENSBTAG00000016555	29	43149610	43152143	1	tRNA phosphotransferase 1	TRPT1	
86	ENSBTAG00000016558	25	1354553	1360122	−1	splA/ryanodine receptor domain and SOCS box containing 3	SPSB3	457
	ENSBTAG00000016561	25	1360579	1365802	1	nucleotide binding protein 2	NUBP2	
100	ENSBTAG00000025028	12	47768049	47781910	−1	Mitotic spindle organizing protein 1	MZT1	205
	ENSBTAG00000019886	12	47782115	47809879	1	bora, aurora kinase A activator	BORA	
104	ENSBTAG00000021547	3	103353700	103422206	−1	WD repeat domain 65	WDR65	399
	ENSBTAG00000021544	3	103422605	103429520	1	EBNA1 binding protein 2	EBNA1BP2	
115	ENSBTAG00000016679	17	41262754	41307126	−1	Electron-transferring-flavoprotein dehydrogenase	ETFDH	252
	ENSBTAG00000033486	17	41307378	41310601	1	chromosome 17 open reading frame, human C4orf46	C17H4orf46	
118	ENSBTAG00000000489	11	10186102	10189962	−1	WD repeat domain 54	WDR54	90
	ENSBTAG00000040226	11	10190052	10195932	1	chromosome 11 open reading frame, human C2orf81	C11H2orf81	

Distance, distance between transcription start sites, in bp.

(Supplementary Tables S9 and S10). The percentage of co-expressed genes is comparable for the bidirectional, convergent and unidirectional configurations, suggesting that co-expression is most likely mediated by shared chromatin environment rather than only by specific regulatory regions. Similar findings were previously found in other species, including yeast.[42] In addition, on average the expression levels in muscle of gene pairs sharing a BiP were more correlated compared with pairs of neighbouring genes in a unidirectional or convergent configuration (Z-score test, P-value $<0.05\%$). Using normalized gene expression counts previously measured in 18 bovine tissues and for three different cows,[43,44] we calculated the Pearson correlation coefficient between the expression levels of the 120 gene pairs sharing a BiP, the 163 gene pairs in an unidirectional configuration and the 218 gene pairs in a convergent configuration. We then

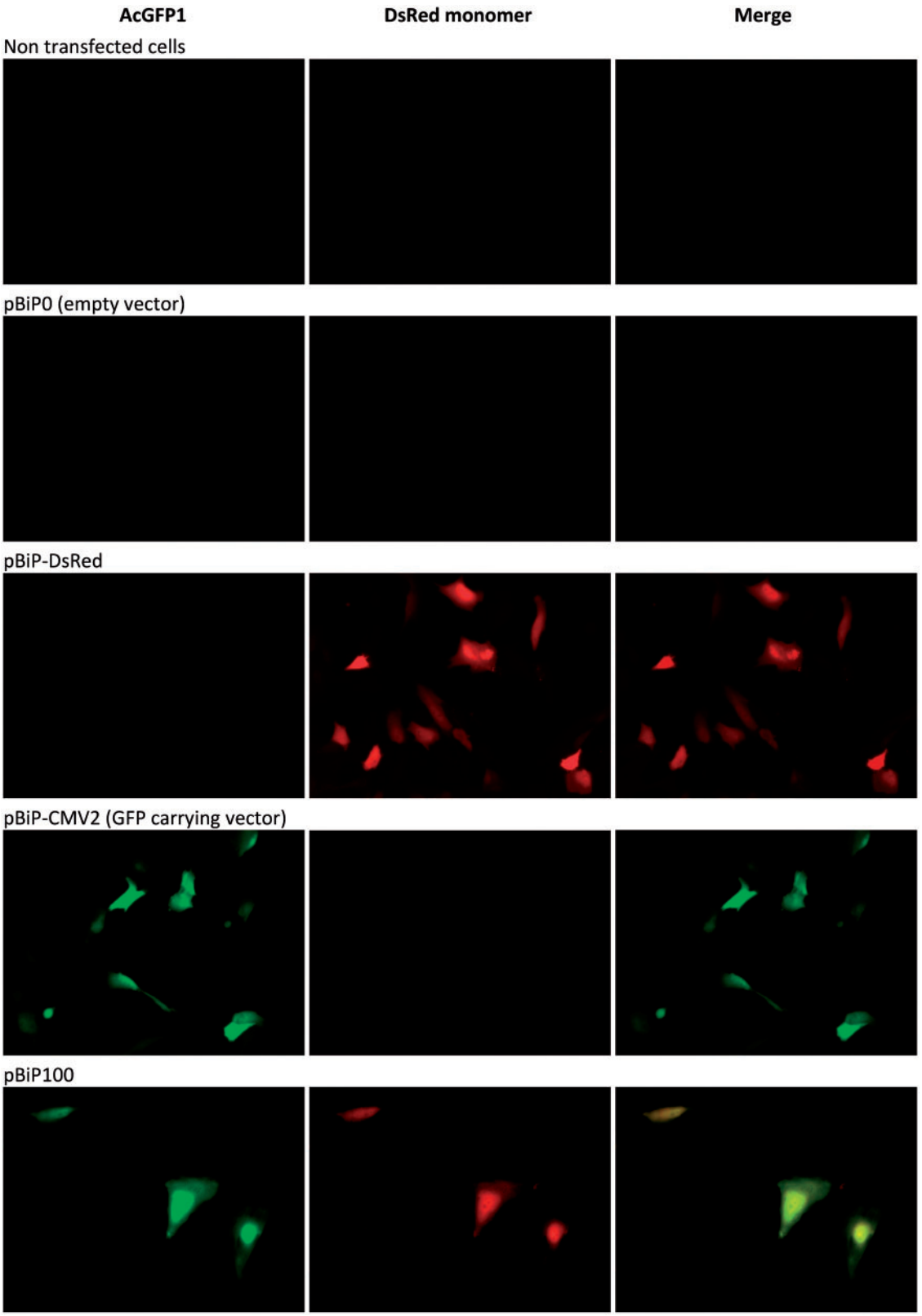

Figure 2. Fluorescence microscopy analysis of bidirectional promoter BiP100 in C2C12 cells. Images were taken at x20 magnification, 36h after transfection.

compared the average expression correlation between these three types of gene pairs and found no significant differences between gene pairs in unidirectional and convergent configurations, but identified significant differences between gene pairs sharing a BiP and the other two types of gene pairs (*Z*-score test, *P*-value $<0.05\%$), The higher correlation between expression levels of genes sharing BiPs has previously been described and is a common feature of bidirectional promoters.[39]

3.6. The activity of the bidirectional promoters is highly variable

Not all the 240 genes sharing the 120 active bidirectional promoters were expressed in all the twenty samples. Among these genes, 227 genes (corresponding to 114 bidirectional promoters) were expressed in all 20 samples; whereas 13 genes (from 13 different bidirectional promoters) were expressed in all but one sample. In order to evaluate inter-individual variability of the activity of the bidirectional promoters active in muscle, we calculated for each sample the expression ratio of each gene pair (using normalized gene expression levels). We considered that the value of this ratio is a measure of the activity of the bidirectional promoter. We then calculated the coefficient of variation for the activity of each bidirectional promoter. We could calculate the activity (paired-gene ratio) only for 107 bidirectional promoters, as some genes were not expressed in some of the 20 samples. Interestingly there were significant differences when we compared the activity variability of these 107 bidirectional promoters. Thirteen bidirectional promoters had a thightly controled activity (CV $<50\%$) whereas 41 had a very variable activity (CV $>100\%$; Supplementary Table S11). These results suggest a strong or a loosened gene expression regulation of the gene pairs depending on the shared bidirectional promoter. This highly variable activity could be due to polymorphisms within the regulatory elements located within the bidirectional promoter regions.

3.7. The expression level of some genes is highly correlated with the activity of some bidirectional promoters active in muscle

Thanks to the availability of the expression levels of genes expressed in LT for 20 samples, it is also possible to analyse the effect of these genes on the activity of the identified bidirectional promoters. Using the normalized expression levels, we calculated the Spearman's rank correlation coefficient for each bidirectional promoter ratio with each gene expressed within the 20 muscle samples. We found after correction for multiple testing 77 significant correlations ($P < 1.9 \times 10^{-8}$), between 77 different genes and the activity of 12 different bidirectional promoters (Supplementary Table S12). We found 74 positive correlations and only 3 negative correlations (anti-correlations). Among these 77 correlations, we could find 9 significant correlations between the activity of a bidirectional promoter and the expression level of one of the two paired-genes sharing the bidirectional promoter. To characterize the functions of the genes showing expression levels highly correlated with the activity of these 12 bidirectional promoters, we performed a GO annotation. No significant enrichment in GO terms and KEGG terms/pathways was found.

Interestingly, with the exception of the paired-genes sharing a bidirectional promoter, no highly correlated genes were in close vicinity of the bidirectional promoter, indicating that the relationship between these genes and the bidirectional promoters are *trans*-effects. In addition, no genes are associated with more than a single bidirectional promoter, suggesting that the regulation of bidirectional promoters is highly variable and no master regulator genes may be at play.

We have to be cautious with our results, as a significant correlation does not relate necessarily to an effect of the gene showing a high correlation on the activity of the bidirectional promoter. Indeed, the paired-genes sharing the bidirectional promoter could themselves have an impact on the regulation of the expression of the gene, for which the expression level shows a strong correlation. However, some of the genes showing a strong correlation with the activity of some of the bidirectional promoters encode for proteins involved in gene regulation, such as histones H2a.1 and H3.1 and transcription factors Hnf1a and Zfn133 and therefore might have direct effect on the activity of these bidirectional promoters. However, no binding sites for Hfn1a or Zfn133 were predicted within the sequence of bidirectional promoter #118 or #81, respectively. More work is needed to establish the directionality of the relationships detected with the correlation and the true involvement of identifed genes with regulatory functions.

3.8. Some genetic variants might alter the activity of the bidirectional promoters active in muscle

After assessing if some genes expressed in muscle influence the activity of the identified bidirectional promoters, we performed a similar analysis to this time identify DNA polymorphisms altering the activity of these bidirectional promoters.

DNA samples from the twenty animals were genotyped with the Illumina BovineSNP50 Genotyping Beadchips. Among the 54,001 SNPs included in the chip, only polymorphic single nucleotide polymorphisms (SNPs) mapping to bovine autosomes on the bovine genome assembly UMD3.1 version,[29] genotyped on $\geq 90\%$ of the individuals, without a low minor allele frequency (MAF ≥ 0.05) and in Hardy–Weinberg Equilibrium (*P*-value <0.001) were retained for further analysis. 21,470 informative SNPs were left after all the filtering steps. We calculated the Spearman's rank correlation coefficient for each bidirectional promoter ratio with the genotypes of each SNP. We found after correction for multiple testing 27 significant correlations (11 and 16 positive and negative correlations respectively, *P*-value $<1.94 \times 10^{-8}$), between 27 different SNPs and the activity of 4 different bidirectional promoters. Annotation of these variants indicates that 20 SNPs are intergenic while the remaining 7 SNPs are within genes but none introduce missense changes in the corresponding proteins or are located in putative regulatory regions (Table 3). GO annotation of these seven genes did not reveal any significant enrichment. None of the intergenic or intronic SNPs were located within known genes encoding bovine long non-coding RNAs.[26,45] The 27 SNPs were neither in close vicinity of the bidirectional promoter and they are all associated with only one bidirectional promoter. This suggest, as seen with the genes whose expression were significantly correlated with the activity of some bidirectional promoters that the effect of these SNPs are *trans*-effects and there might not be master regulator genes. In addition, annotation of these variants indicates that most SNPs are intergenic and it is unlikely that these 27 SNPs have any direct effect on the activity of the bidirectional promoter.

In order to identify SNPs altering directly the activity of some of the bidirectional promoters active in muscle, we performed another correlation analysis between the activity of the bidirectional promoters and this time the genotypes from coding SNPs identified with the RNA-Seq data. Mapped paired-end reads from each sample were

Table 3. List of SNPs from the Illumina BovineSNP50 Beadchip with genotypes highly correlated to the activity of some bidirectional promoters

SNP ID	BTA	Position	Alleles	Consequence	Ensembl gene ID	Ensembl transcript ID	Gene symbol	BiP #	*Rho*	*P*-value
Hapmap42177-BTA-31679	1	61,125,554	A/G	intergenic_variant				118	0.99	1.78E-24
Hapmap42933-BTA-40454	1	85,731,453	A/G	intergenic_variant				101	−0.99	1.95E-20
ARS-BFGL-NGS-108514	2	105,521,127	A/G	intergenic_variant				118	−0.99	1.78E-24
Hapmap54229-rs29017613	4	30,200,987	A/G	intergenic_variant				118	0.99	1.78E-24
ARS-BFGL-NGS-35869	5	21,707,809	A/G	intergenic_variant				118	−0.99	1.78E-24
ARS-BFGL-NGS-93953	5	22,737,219	A/G	intergenic_variant				118	0.99	1.78E-24
Hapmap49852-BTA-107572	6	29,790,005	A/G	intergenic_variant				104	−0.99	1.78E-24
Hapmap25168-BTC-033275	6	33,713,818	A/G	intergenic_variant				104	0.99	1.78E-24
Hapmap36813-SCAFFOLD50174_9004	6	33,768,128	A/G	intergenic_variant				104	0.99	1.78E-24
Hapmap23923-BTC-066021	6	39,721,727	A/C	intergenic_variant				104	−0.99	1.78E-24
Hapmap36286-SCAFFOLD260285_24265	7	111,161,115	A/C	intergenic_variant				104	−0.99	1.78E-24
ARS-BFGL-NGS-99031	8	101,251,865	A/C	intergenic_variant				118	0.99	1.78E-24
ARS-BFGL-NGS-60678	9	85,454,475	A/G	intergenic_variant				104	0.99	1.78E-24
ARS-BFGL-NGS-4488	10	6,476,252	A/G	intron_variant	ENSBTAG00000024878	ENSBTAT00000048409	ANKRD31	104	−1	0
ARS-BFGL-NGS-5819	10	10,428,184	A/G	intron_variant	ENSBTAG00000025853	ENSBTAT00000017647	HOMER1	118	0.99	1.78E-24
ARS-BFGL-NGS-27341	11	91,438,914	A/G	intergenic_variant				118	−0,99	1.78E-24
ARS-BFGL-NGS-115889	12	79,814,959	A/C	intron_variant	ENSBTAG00000010395	ENSBTAT00000013726	DOCK9	64	0.99	1.95E-20
ARS-BFGL-NGS-23509	14	70,636,087	A/G	intergenic_variant				118	−0.99	1.78E-24
ARS-BFGL-NGS-859	14	73,919,098	A/G	intergenic_variant				104	−0.99	1.78E-24
ARS-BFGL-NGS-99802	16	74,999,809	A/C	5_prime_UTR_variant	ENSBTAG00000010850	ENSBTAT00000014402	SERTAD4	104	0.99	1.78E-24
Hapmap52466-rs29015577	19	3,617,183	A/C	intergenic_variant				118	−0.99	1.78E-24
Hapmap47625-BTA-44726	19	21,878,635	A/G	intergenic_variant				104	−0.99	1.78E-24
ARS-BFGL-NGS-14187	19	25,165,920	A/G	intron_variant	ENSBTAG00000014806	ENSBTAT00000019703	ATP2A3	118	−0.99	1.78E-24
ARS-BFGL-NGS-109291	23	13,517,193	T/A	intron_variant	ENSBTAG00000027197	ENSBTAT00000064399	KIF6	64	−0.99	1.95E-20
ARS-BFGL-NGS-57958	26	23,000,155	A/G	intergenic_variant				104	−0.99	1.78E-24
ARS-BFGL-NGS-42033	28	14,243,381	C/G	intergenic_variant				118	−0.99	1.78E-24
ARS-BFGL-NGS-77028	28	26,689,199	A/G	intron_variant	ENSBTAG00000005666	ENSBTAT00000007444	LRRC20	118	0.99	1.78E-24

Table 4. List of coding SNPs with genotypes highly correlated to the activity of some bidirectional promoters

cSNP ID	BTA	Position	dbSNP ID	Alleles	Consequence	Ensembl gene ID	Ensembl transcript ID	Gene symbol	BiP #	*Rho*	*P*-value
ENSBTAT00000002157_515	9	61,291,956		G/A	intron_variant	ENSBTAG00000001644	ENSBTAT00000002157	MDN1	104	−0.99	1.78E-24
ENSBTAT00000003078_106	15	82,337,162		G/A	intron_variant	ENSBTAG00000002381	ENSBTAT00000003078	ZDHHC5	118	0.99	1.78E-24
ENSBTAT00000003826_479	21	22,218,555		C/T	upstream_gene_variant	ENSBTAG00000002939	ENSBTAT00000003826	FURIN	104	−0.99	1.78E-24
ENSBTAT00000005923_585	25	2,977,744		T/C	upstream_gene_variant	ENSBTAG00000004509	ENSBTAT00000005923	SLX4	118	−0.99	1.78E-24
ENSBTAT00000008129_240	16	19,500,611		G/C	intron_variant	ENSBTAG00000006186	ENSBTAT00000008129	KCTD3	118	0.99	1.78E-24
ENSBTAT00000008507_123	15	82,265,194		T/C	intron_variant	ENSBTAG00000006493	ENSBTAT00000008507	CLP1	118	−0.99	1.78E-24
ENSBTAT00000008728_747	12	18,197,212		C/T	intron_variant	ENSBTAG00000006640	ENSBTAT00000008728	RB1	118	0.99	1.78E-24
ENSBTAT00000018163_665	21	66,866,347		T/G	upstream_gene_variant	ENSBTAG00000013666	ENSBTAT00000018163	SLC25A29	104	−0.99	1.78E-24
ENSBTAT00000020493_130	7	62,925,278		A/G	intron_variant	ENSBTAG00000015419	ENSBTAT00000020493	ARHGEF37	64	−0.99	1.95E-20
ENSBTAT00000025492_402	19	55,675,666		T/C	intron_variant	ENSBTAG00000019153	ENSBTAT00000025492	JMJD6	24	−1	6.38E-137
ENSBTAT00000028060_520	21	16,720,681		C/T	intron_variant	ENSBTAG00000037383	ENSBTAT00000028060	AKAP13	118	−0.99	1.78E-24
ENSBTAT00000028277_365	15	76,757,181		T/G	intron_variant	ENSBTAG00000021223	ENSBTAT00000028277	CRY2	118	−0.99	1.78E-24
ENSBTAT00000028656_569	14	75,708,914		T/G	intron_variant	ENSBTAG00000039968	ENSBTAT00000028656	TMEM55A	104	0.99	1.78E-24
ENSBTAT00000028865_103	25	39,432,034		G/A	intron_variant	ENSBTAG00000037400	ENSBTAT00000028865	TNRC18	118	0.99	1.78E-24
ENSBTAT00000029400_195	17	73,658,214		C/T	upstream_gene_variant	ENSBTAG00000021656	ENSBTAT00000029400	SPECC1L	104	0.99	1.78E-24
ENSBTAT00000029403_203	18	62,789,937		A/G	intron_variant	ENSBTAG00000030393	ENSBTAT00000029403	RDH13	104	−0.99	1.78E-24
ENSBTAT00000033704_233	6	113,701,647	rs442770236	G/T	upstream_gene_variant	ENSBTAG00000004316	ENSBTAT00000033704	BOD1L	104	0.99	1.78E-24
ENSBTAT00000035362_466	20	10,305,279		G/A	downstream_gene_variant	ENSBTAG00000027980	ENSBTAT00000035362	TAF9	118	−0.99	1.78E-24
ENSBTAT00000037465_138	25	10,053,612		T/C	intron_variant	ENSBTAG00000026375	ENSBTAT00000037465	RMI2	118	−0.99	1.78E-24
ENSBTAT00000043122_702	10	86,384,804		A/G	upstream_gene_variant	ENSBTAG00000020379	ENSBTAT00000043122	AREL1	118	0.99	1.78E-24
ENSBTAT00000043260_900	10	80,246,761		A/G	upstream_gene_variant	ENSBTAG00000014334	ENSBTAT00000043260	ZFYVE26	118	0.99	1.78E-24
ENSBTAT00000043778_165	15	82,285,586		T/C	intron_variant	ENSBTAG00000002411	ENSBTAT00000043778	CTNND1	118	0.99	1.78E-24
ENSBTAT00000043778_492	15	82,284,426		G/A	intron_variant	ENSBTAG00000002411	ENSBTAT00000043778	CTNND1	118	−0.99	1.78E-24
ENSBTAT00000043778_760	15	82,284,694		C/T	intron_variant	ENSBTAG00000002411	ENSBTAT00000043778	CTNND1	118	−0.99	1.78E-24
ENSBTAT00000054096_509	8	103,426,705		G/A	upstream_gene_variant	ENSBTAG00000038335	ENSBTAT00000054096	IGBP1	118	0.99	1.78E-24
ENSBTAT00000056520_352	23	27,378,739		C/T	upstream_gene_variant	ENSBTAG00000039620	ENSBTAT00000056520	MGC151586	104	0.99	1.78E-24
ENSBTAT00000061199_185	14	20,989,316		A/G	upstream_gene_variant	ENSBTAG00000044106	ENSBTAT00000061199	SPIDR	118	−0.99	1.78E-24
ENSBTAT00000061451_269	23	3,434,098		G/A	intron_variant	ENSBTAG00000021237	ENSBTAT00000061451	DST	118	−0.99	1.78E-24
ENSBTAT00000061451_704	23	3,432,111		G/T	intron_variant	ENSBTAG00000021237	ENSBTAT00000061451	DST	118	0.99	1.78E-24
ENSBTAT00000034195_104	8	53,736,890		G/A	intron_variant	ENSBTAG00000017734	ENSBTAT00000034195	VPS13A	118	0.99	1.78E-24

used with the SAMtools package for variant calling. Using stringent parameters (e.g. minimum coverage of 8 reads and mapping quality of 20) we detected 8,121 different billalelic SNPs. Among the SNPs identified, similarly to the SNPs from the Illumina BovineSNP50 Genotyping Beadchip, we used only polymorphisms mapping to bovine autosomes, genotyped on ≥90% of the individuals, without a low minor allele frequency (MAF ≥0.05) and in Hardy–Weinberg Equilibrium (P-value <0.001). A total of 8,116 informative SNPs were left after these filtering steps. We calculated the Spearman's rank correlation coefficient for each bidirectional promoter ratio with the genotypes of each SNP. We found after correction for multiple testing 30 significant correlations (P-value $<5.13 \times 10^{-8}$), between 30 different SNPs (in 27 different genes) and the activity of 4 different bidirectional promoters (Table 4). Annotation of these variants indicates that all these SNPs are within protein-coding genes. GO annotation of the 27 genes containing the 30 coding SNPs (cSNPs) did not reveal any significant enrichment. None of the 30 cSNPs introduce missense changes in the corresponding proteins; however, 11 cSNPs are located in the upstream or downstream regions.

We found among these correlations a significant negative correlation between a cSNP within *Taf9*, a gene encoding the RNA polymerase II, TATA box binding protein-associated factor (TAF) 32kD (also known as *TafII32* or *Taf2g)*, and the activity of bidirectional promoter #118. TAF9 is involved in the initiation of the transcription by RNA polymerase II.[46] Interestingly, Taf1, another TATA box binding factor has motifs over-represented in human bidirectional promoters[47] and two-hybrid experiments performed in yeast have shown that Taf1 is able to bind to DNA but also to different TATA box binding factors, including Taf9.[48] Many bidirectional promoters lack a TATA box;[3,11] however, it has been shown that several TATA-less promoters require TAFs for transcription.[47–53] The cSNP found in *Taf9* is therefore an interesting variant and might point out an effect of Taf9 on the activity of bidirectional promoter #118.

The 11 cSNPs might have an impact on the regulation of the corresponding gene, which in turn might affect the activity of some bidirectional promoters. Spearman's rank correlation coefficients were calculated for the genotypes of these 11 cSNPs and the normalized expression level of the corresponding gene (Table 5). Significant correlations were found at P-value <0.05 for 4 cSNPs, suggesting that these cSNPs have an effect on the regulation of the expression of their corresponding gene. The seven other cSNPs, including the variant found in *Taf9*, might have most probably an effect on the post-transcriptional regulation of the corresponding gene, which consequently alter the amount of proteins which then affect the activity of the bidirectional promoter.

3.9. Some bidirectional promoter regions are located within quantitative trait loci for meat-related phenotypes

The positions of the 120 bidirectional promoters sharing genes expressed in bovine LT were compared with the position of known quantitative trait loci (QTLs) deposited in the public database AnimalQTLdb.[54] Fifty-four bidirectional promoters were located in 244 different QTL regions (Supplementary Table S14). For example, 7 different bidirectional promoters are found in 8 QTL regions for intramuscular fat; whereas 18 bidirectional promoters are within 18 QTLs for shear force. Karim et al.[55] have shown that a genetic variant located within a bidirectional promoter shared by *PLAG1* and

Table 5. Correlation scores between cSNPs and the expression level of the corresponding gene

cSNP ID	BTA	Position	dbSNP ID	Alleles	Consequence	Ensembl transcript ID	Gene symbol	*Rho*	*P*-value
ENSBTAT00000003826_4794	21	22,218,555		C/T	upstream_gene_variant	ENSBTAT00000003826	FURIN	−0.52	0.02
ENSBTAT00000018163_665	21	66,866,347		T/G	upstream_gene_variant	ENSBTAT00000018163	SLC25A29	−0.49	0.03
ENSBTAT00000043122_702	10	86,384,804		A/G	upstream_gene_variant	ENSBTAT00000043122	AREL1	0.49	0.03
ENSBTAT00000043260_900	10	80,246,761		A/G	upstream_gene_variant	ENSBTAT00000043260	ZFYVE26	0.46	0.04
ENSBTAT00000005923_585	25	2,977,744		T/C	upstream_gene_variant	ENSBTAT00000005923	SLX4	−0.43	0.06
ENSBTAT00000035362_4667	20	10,305,279		G/A	downstream_gene_variant	ENSBTAT00000035362	TAF9	−0.43	0.06
ENSBTAT00000029400_195	17	73,658,214		C/T	upstream_gene_variant	ENSBTAT00000029400	SPECC1L	0.35	0.14
ENSBTAT00000061199_1851	14	20,989,316		A/G	upstream_gene_variant	ENSBTAT00000061199	SPIDR	−0.29	0.22
ENSBTAT00000033704_233	6	113,701,647	rs442770236	G/T	upstream_gene_variant	ENSBTAT00000033704	BOD1L	0.17	0,46
ENSBTAT00000054096_509	8	103,426,705		G/A	upstream_gene_variant	ENSBTAT00000054096	IGBP1	0.14	0.54
ENSBTAT00000056520_3528	23	27,378,739		C/T	upstream_gene_variant	ENSBTAT00000056520	MGC151586	−0.12	0.63

CHCHD7 influences bovine stature. It will be interesting to investigate if variants within some of the bidirectional promoter regions we identified affect quantitative traits, including meat quality traits.

4. Conclusion

In this study, we identified 120 bidirectional promoters active in bovine muscle and validated *in vitro* 16 of them. Using RNA-seq data from 20 muscle samples, we found that the activity of these bidirectional promoters is highly variable and that the expression level of 77 genes is highly correlated with the activity of 12 of the bidirectional promoters. We also identified 57 SNPs associated significantly with the activity of 5 bidirectional promoters. Interestingly, we found that a coding SNP within *Taf9* has an effect on the activity of bidirectional promoter #118. To our knowledge, our study is the first analysis in any species of the genetic variability of the activity of bidirectional promoters.

Data availability

The sequencing data have been submitted to the European Nucleotide Archive (accession numbers ERP002220, E-MTAB-2646 and E-MTAB-4625).

Acknowledgements

We are grateful to Sébastien Legardinier (UMR1061, Limoges) for the gift of the C2C12 cells and Jérôme Chapuis (UR892 INRA VIM, Jouy-en-Josas) for his help with the fluorescence microscope.

Conflict of interest

None declared.

Supplementary data

Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

The work was supported by the French National Research Agency (Regulomix project, contract ANR-09-GENM-011), French National Institute for Agricultural Research and the Limousin Regional Council.

References

1. Adachi, N. and Lieber, M. 2002, Bidirectional gene organization: a common architectural feature of the human genome. *Cell* **109**, 807–9.
2. Takai, D. and Jones, P.A. 2004, Origins of bidirectional promoters: computational analyses of intergenic distance in the human genome. *Mol. Biol. Evol.* **21**, 463–7.
3. Trinklein, N.D., Aldred, S.F., Hartman, S.J., Schroeder, D.I., Otillar, R.P. and Myers, R.M. 2004, An abundance of bidirectional promoters in the human genome. *Genome Res.* **14**, 62–6.
4. Williams, T.J. and Fried, M. 1986, The MES-1 murine enhancer element is closely associated with the heterogeneous 5′ ends of two divergent transcription units. *Mol. Cell. Biol.* **6**, 4558–69.
5. Herr, D.R. and Harris, G.L. 2004, Close head-to-head juxtaposition of genes favors their coordinate regulation in *Drosophila melanogaster*. *FEBS Lett.* **572**, 147–53.
6. Koyanagi, K.O., Hagiwara, M., Itoh, T., Gojobori, T. and Imanishi, T. 2005, Comparative genomics of bidirectional gene pairs and its implications for the evolution of a transcriptional regulation system. *Gene* **353**, 169–76.
7. Engstrom, P.G., Suzuki, H., Ninomiya, N., et al. 2006, Complex loci in human and mouse genomes. *PLoS Genet.* **2**, e47.
8. Li, Y.Y., Yu, H., Guo, Z.M., Guo, T.Q., Tu, K. and Li, X.Y. 2006, Systematic analysis of head-to-head gene organization: evolutionary conservation and potential biological relevance. *PLoS Comput. Biol.* **2**, e74.
9. Krom, N. and Ramakrishna, W. 2008, Comparative analysis of divergent and convergent gene pairs and their expression patterns in rice, *Arabidopsis*, and populus. *Plant Physiol.* **147**, 1763–73.
10. Yang, M.Q. and Elnitski, L.L. 2008, Prediction-based approaches to characterize bidirectional promoters in the mammalian genome. *BMC Genomics* **9** (**Suppl 1**), S2.
11. Yang, M.Q. and Elnitski, L.L. 2008, Diversity of core promoter elements comprising human bidirectional promoters. *BMC Genomics* **9** (**Suppl 2**), S3.
12. Yang, M.Q., Taylor, J. and Elnitski, L. 2008, Comparative analyses of bidirectional promoters in vertebrates. *BMC Bioinformatics* **9** (**Suppl 6**), S9.
13. Piontkivska, H., Yang, M.Q., Larkin, D.M., Lewin, H.A., Reecy, J. and Elnitski, L. 2009, Cross-species mapping of bidirectional promoters enables prediction of unannotated 5′ UTRs and identification of species-specific transcripts. *BMC Genomics* **10**, 189.
14. Wang, Q., Wan, L., Li, D., Zhu, L., Qian, M. and Deng, M. 2009, Searching for bidirectional promoters in *Arabidopsis thaliana*. *BMC Bioinformatics* **10** (**Suppl 1**), S29.
15. Krom, N. and Ramakrishna, W. 2010, Conservation, rearrangement, and deletion of gene pairs during the evolution of four grass genomes. *DNA Res.* **17**, 343–52.
16. Woo, Y.H. and Li, W.H. 2011, Gene clustering pattern, promoter architecture, and gene expression stability in eukaryotic genomes. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3306–11.
17. Behura, S.K. and Severson, D.W. 2015, Bidirectional promoters of insects: genome-wide comparison, evolutionary implication and influence on gene expression. *J. Mol. Biol.* **427**, 521–36.
18. Maxson, R., Cohn, R., Kedes, L. and Mohun, T. 1983, Expression and organization of histone genes. *Annu. Rev. Genet.* **17**, 239–77.
19. Momota, R., Sugimoto, M., Oohashi, T., Kigasawa, K., Yoshioka, H. and Ninomiya, Y. 1998, Two genes, COL4A3 and COL4A4 coding for the human α3(IV) and α4(IV) collagen chains are arranged head-to-head on chromosome 2q36. *FEBS Lett.* **424**, 11–6.
20. Harrison, L., Ascione, A.G., Takiguchi, Y., Wilson, D.M., Chen, D.J. and Temple, B. 1997, Comparison of the promoters of the mouse (APEX) and human (APE) apurinic endonuclease genes. *Mutat. Res.* **385**, 159–72.
21. Tan, M., Wei, C. and Price, C.M. 2003, The telomeric protein Rap1 is conserved in vertebrates and is expressed from a bidirectional promoter positioned between the Rap1 and KARS genes. *Gene* **323**, 1–10.
22. Uwanogho, D.A., Yasin, S.A., Starling, B. and Price, J. 2010, The intergenic region between the mouse Recql4 and Lrrc14 genes functions as an evolutionary conserved bidirectional promoter. *Gene* **449**, 103–17.
23. Meersseman, C., Lejard, V., Rebours, E., et al. 2014, Bovine TWINKLE and mitochondrial ribosomal protein L43 genes are regulated by an evolutionary conserved bidirectional promoter. *Gene* **537**, 154–63.
24. Allais, S., Leveziel, H., Hocquette, J.F., et al. 2014, Fine mapping of quantitative trait loci underlying sensory meat quality traits in three French beef cattle breeds. *J. Anim. Sci.* **92**, 4329–41.
25. Djari, A., Esquerre, D., Weiss, B., et al. 2013, Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics* **14**, 307.
26. Billerey, C., Boussaha, M., Esquerre, D., et al. 2014, Identification of large intergenic non-coding RNAs in bovine muscle using next-generation transcriptomic sequencing. *BMC Genomics* **15**, 499.
27. Li, H. and Durbin, R. 2009, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 175417–60.
28. Li, H., Handsaker, B., Wysoker, A., et al. 2009, The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–9.

29. Zimin, A.V., Delcher, A.L., Florea, L., et al. 2009, A whole-genome assembly of the domestic cow. *Bos taurus.Genome Biol.* **10**, R42.
30. Li, H; 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–93.
31. Birney, E., Andrews, D., Caccamo, M., et al. 2006, Ensembl 2006. *Nucleic Acids Res.* **34** (Database issue), D556–D561.
32. Al-Shahrour, F., Minguez, P., Vaquerizas, J.M., Conde, L. and Dopazo, J. 2005, BABELOMICS: a suite of web tools for functional annotation and analysis of groups of genes in high-throughput experiments. *Nucleic Acids Res.* **33** (Web Server issue), W460–4.
33. Kel, A.E., Gössling, E., Reuter, I., Cheremushkin, E., Kel-Margoulis, O.V. and Wingender E. 2003, MATCH: a tool for searching transcription factor binding sites in DNA sequences. *Nucleic Acids Res.* **31**, 3576–9.
34. Matys, V., Fricke, E., Geffers, R., et al. 2003, TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res.* **31**, 374–378.
35. Homepage Repeat Masker Server, 2013. http://www.repeatmasker.org.
36. Rozen, S. and Skaletsky, H. 2000, Primer3 on the WWW for general users and for biologist programmers. *Methods Mol. Biol.* **132**, 365–86.
37. Lejard, V., Rebours, E., Meersseman, C. and Rocha, D. 2014, Construction and validation of a novel dual reporter vector for studying mammalian bidirectional promoters. *Plasmid* **74**, 1–8.
38. Chitwood, J.L., Rincon, G., Kaiser, G.G., Medrano, J.F. and Ross, P.J. 2003, RNA-seq analysis of single bovine blastocysts. *BMC Genomics* **14**, 350.
39. Xu, C., Chen, J. and Shen, B. 2012, The preservation of bidirectional promoter architecture in eukaryotes: what is the driving force? *BMC Syst. Biol.* **6** (**Suppl 1**), S21.
40. Liu, B., Chen, J. and Shen, B. 2011, Genome-Wide analysis of the transcription factor binding preference of human bi-directional promoters and functional annotation of related gene pairs. *BMC Syst. Biol.* **5** (**Suppl 1**), S2.
41. Karolchik, D., Baertsch, R., Diekhans, M., et al. 2003, The UCSC genome browser database. *Nucleic Acids Res.* **31**, 51–54.
42. Tsai, H.K., Su, C.P., Lu, M.Y., Shih, C.H and Wang, D. 2007, Co-expression of adjacent genes in yeast cannot be simply attributed to shared regulatory system. *BMC Genomics* **8**, 352.
43. Koufariotis, L.T., Chen, Y.P., Chamberlain, A., Vander Jagt, C. and Hayes, B.J. 2015. A catalogue of novel bovine long noncoding RNA across 18 tissues. *PLoS One* **10**, e0141225.
44. Chamberlain, A.J., Vander Jagt, C.J., Hayes, B.J., et al. 2015. Extensive variation between tissues in allele specific expression in an outbred mammal. *BMC Genomics* **16**, 993.
45. Liu, C., Bai, B., Skogerbø, G., et al. 2005, NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Res.* **33** (Database issue), D112–5.
46. Hisatake, K., Ohta, T., Takada, R., et al. 1995, Evolutionary conservation of human TATA-binding-polypeptide-associated factors TAFII31 and TAFII80 and interactions of TAFII80 with other TAFs and with general transcription factors. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 8195–9.
47. Chen, Y., Li, Y., Wei, J. and Li, Y.Y. 2014, Transcriptional regulation and spatial interactions of head-to-head genes. *BMC Genomics* **15**, 519.
48. Yatherajam, G., Zhang, L., Kraemer, S.M. and Stargell, L.A. 2003, Protein-protein interaction map of yeast TFIID. *Nucleic Acids Res.* **31**, 1252–60.
49. Pugh, B.F. and Tjian, R. 1991, Transcription from a TATA-less promoter requires a multisubunit TFIID complex. *Genes Dev.* **5**, 1935–45.
50. Zhou, Q., Lieberman, P.M., Boyer, T.G. and Berk, A.J. 1992, Holo-TFIID supports transcriptional stimulation by diverse activators and from a TATA-less promoter. *Genes Dev.* **6**, 1964–74.
51. Sadowski, C.L., Henry, R.W., Lobo, S.M. and Hernandez, N. 1993, Targeting TBP to a non-TATA box cis-regulatory element: a TBP-containing complex activates transcription from snRNA promoters through the PSE. *Genes Dev.* **7**, 1535–48.
52. Zhou, Q., Boyer, T.G. and Berk, A.J. 1993, Factors (TAFs) required for activated transcription interact with TATA box-binding protein conserved core domain. *Genes Dev.* **7**, 180–7.
53. Martinez, E., Chiang, C.M., Ge, H. and Roeder, R.G. 1994, TATA-binding protein-associated factor(s) in TFIID function through the initiator to direct basal transcription from a TATA-less class II promoter. *EMBO J.* **13**, 3115–312.
54. Hu, Z.L., Fritz, E.R. and Reecy, J.M. 2007, AnimalQTLdb: a livestock QTL database tool set for positional QTL information mining and beyond. *Nucleic Acids Res.* **35** (Database issue), D604–9.
55. Karim, L., Takeda, H., Lin, L., et al. 2011, Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. *Nat. Genet.* **43**, 405–13.