scientific reports

Check for updates

OPEN

Muscle allele-specific expression QTLs may affect meat quality traits in *Bos indicus*

Jennifer Jessica Bruscadin[1], Marcela Maria de Souza[2], Karina Santos de Oliveira[1], Marina Ibelli Pereira Rocha[1], Juliana Afonso[5], Tainã Figueiredo Cardoso[3], Adhemar Zerlotini[4], Luiz Lehmann Coutinho[5], Simone Cristina Méo Niciura[3] & Luciana Correia de Almeida Regitano[3]✉

Single nucleotide polymorphisms (SNPs) located in transcript sequences showing allele-specific expression (ASE SNPs) were previously identified in the *Longissimus thoracis* muscle of a Nelore (*Bos indicus*) population consisting of 190 steers. Given that the allele-specific expression pattern may result from *cis*-regulatory SNPs, called allele-specific expression quantitative trait loci (aseQTLs), in this study, we searched for aseQTLs in a window of 1 Mb upstream and downstream from each ASE SNP. After this initial analysis, aiming to investigate variants with a potential regulatory role, we further screened our aseQTL data for sequence similarity with transcription factor binding sites and microRNA (miRNA) binding sites. These aseQTLs were overlapped with methylation data from reduced representation bisulfite sequencing (RRBS) obtained from 12 animals of the same population. We identified 1134 aseQTLs associated with 126 different ASE SNPs. For 215 aseQTLs, one allele potentially affected the affinity of a muscle-expressed transcription factor to its binding site. 162 aseQTLs were predicted to affect 149 miRNA binding sites, from which 114 miRNAs were expressed in muscle. Also, 16 aseQTLs were methylated in our population. Integration of aseQTL with GWAS data revealed enrichment for traits such as meat tenderness, ribeye area, and intramuscular fat . To our knowledge, this is the first report of aseQTLs identification in bovine muscle. Our findings indicate that various *cis*-regulatory and epigenetic mechanisms can affect multiple variants to modulate the allelic expression. Some of the potential regulatory variants described here were associated with the expression pattern of genes related to interesting phenotypes for livestock. Thus, these variants might be useful for the comprehension of the genetic control of these phenotypes.

Allele-specific expression (ASE) or allelic expression imbalance is a pattern that reflects the expression difference between copies of a gene from each chromosome (i.e., alleles)[1]. When ASE is parentally guided, it is named genomic imprinting, caused by epigenetic modifications with *cis* action, which result in the monoallelic expression of genes essential for mammal growth and development[2].

Genes showing ASE in their transcripts, as evidenced by the analysis of single nucleotide polymorphisms (SNP) alleles within them, are called ASE genes. ASE genes have been shown to affect phenotypes for livestock, such as adipogenesis and lipid metabolism in pigs[3]. The knowledge about regions with ASE and their regulatory mechanisms is relevant to improve animal breeding programs by increasing the accuracy of predictive models[4]. De Souza et al.[5] identified SNPs in transcribed regions showing allele-specific expression patterns (ASE SNPs) in the muscle of Nelore (*Bos indicus*) related to meat tenderness, an important trait for meat consumer's approval[6,7]. Many of these SNPs were located within genes responsible for essential biological functions for muscle development and meat tenderness[5].

Recent studies have explored putative regulatory mechanisms that may be responsible for ASE and genomic imprinting[8–10]. For instance, ASE can be caused by differential methylation in regulatory regions of the two parental alleles[11]. DNA methylation is a well-known imprinting mark in CpG islands[12], DNA regions with clusters of cytosine followed by guanine dinucleotides. This type of epigenetic modification, when presented in

[1]Post-Graduation Program of Evolutionary Genetics and Molecular Biology, Center of Biological Sciences and Health, Federal University of São Carlos, São Carlos, SP, Brazil. [2]Post-Doctoral Fellow, Department of Animal Science, Iowa State University, Ames, IA, USA. [3]Embrapa Pecuária Sudeste, P. O. Box 339, São Carlos, SP 13564-230, Brazil. [4]Embrapa Informática Agropecuária, Campinas, SP, Brazil. [5]Department of Animal Science, University of São Paulo/ESALQ, Piracicaba, SP, Brazil. ✉email: luciana.regitano@embrapa.br

promoters, can suppress transcription when the CpG islands are hypermethylated, while active transcription shows no methylation or few isolated methylation events[11]. Another mechanism involved with the ASE pattern is the presence of differences on microRNA (miRNA) binding sites between the alleles[13,14], once miRNAs can impair the translation by degrading the target mRNA sequence[15]. Similarly, sequences showing an affinity for a transcription factor (TF), i.e., a transcription factor binding site (TFBS), may promote or suppress the transcription of the allele in phase with the given TFBS. Thus, in a heterozygous locus, if the regulatory region of only one allele has an affinity with the transcription factor (TF), this *cis*-regulatory variant can also lead to ASE[16].

To discover regulatory mechanisms causing ASE, we can execute an initial screening analysis to identify SNPs associated with a given transcript's unequal allele expression pattern: the ASE-quantitative trait loci (aseQTLs)[17]. From this initial screening, aseQTLs that present any evidence of being related to interesting phenotypes and regulatory mechanisms affecting allelic expression can be prioritized for further analysis. AseQTLs can contribute to the understanding of expression regulation of genes associated with economic traits and how alleles under selection are expected to be expressed in the offspring[18–23].

Herein, we investigated aseQTLs associated with allelic expression imbalance, evidenced by the number of reads corresponding to different SNP alleles in transcripts, previously described in a Nelore experimental population[5]. We then analyzed its involvement with predicted regulatory mechanisms, such as DNA methylation, whether they contained TFBS or miRNA binding sites. This study aimed to add a new knowledge layer about genomic regulation in Nelore muscle, describing potential mechanisms involved in the allelic expression of genes related to relevant phenotypes in bovine.

Results

aseQTLs analysis. We retrieved a total of 82,384 SNPs after filtering the original genotype file (429,513 SNPs from 190 Nelore samples), considering a 1 Mb sized window around the 820 SNPs with significant ASE. Despite this initial dataset, only 22,470 tests were valid due to the requirement of 10 heterozygous samples for both the aseQTL candidate and the SNP marking allelic imbalance. In this context, only 192 SNPs with ASE were tested in the aseQTL analysis. Allelic imbalance ratio values were compared between groups of homozygous and heterozygous animals for each SNP to be tested as aseQTL. AseQTL analysis resulted in 1134 significant aseQTLs (P value ≤ 0.05), associated with 126 SNPs marking allelic imbalance of 85 genes. AseQTLs reference alleles have a mean frequency of 0.67, ranging from 0.49 to 0.94. Figure 1 represents ASE SNPs and aseQTLs distributions through the chromosomes.

Three SNPs marking ASE were also identified as aseQTLs: rs136209194, located in the intergenic region, which acts like an aseQTL for the rs132817153 SNP; rs109842586 (*XIRP2*), associated with the ASE pattern of the rs109372848 SNP; and rs136097891 (*MTUS1*), an aseQTL for the rs109005284 SNP *(MTUS1)*. AseQTLs association results can be found in Supplementary Table 1.

The average number of aseQTLs per ASE SNP was nine (ranging from one to 56). The rs134422650 marker was associated with the highest number of aseQTLs (56). The second one was the rs110850310, located in the *HSPA1A* gene, with 38 aseQTLs. Figure 1 shows that some genomic locations had several aseQTLs (red ring) because they had multiple SNPs with allelic imbalance (grey ring) nearby, for example, between 15 and 20 kb positions of the chromosome 2 or between 10 and 15 kb of the chromosome 10.

Most aseQTLs affected the ASE marked by a unique SNP. Still, we found the rs132798564 aseQTL simultaneously associated with four ASE SNP markers of the *CMYA5* gene, and the rs110663707 aseQTL was associated with three SNPs, also of the *CMYA5* ASE gene. Moreover, 54 aseQTLs were associated simultaneously with different SNP pairs.

An average distance of 404,983 bp (varying from 281 bp to 1.1 Mb) was observed between aseQTLs and the respective transcription start sites (TSSs) of affected ASE genes. Moreover, the average distance between an aseQTL and its associated SNP marker for ASE was 417,698 bp (ranging from 2238 bp to 999,650 bp). AseQTLs were more frequently located close to the TSS of ASE genes and their associated SNPs (from 0 to 100 kb), becoming less frequent as the distance increases, especially in intervals larger than 1 Mb. Figure 2 shows aseQTLs distribution compared with the ASE SNPs and the TSSs of genes showing ASE in their transcripts.

Linkage disequilibrium analysis. Firstly, we evaluated the linkage disequilibrium (LD) between all aseQTLs to investigate LD blocks with associated regulatory elements. A total of 3030 aseQTL pairs showed D′ greater than 0.8, and 1813 of them were in total disequilibrium (D′ = 1). In our aseQTL data, we identified 117 LD blocks (Supplementary Table 2) that included between two to 21 aseQTLs. The largest LD block contained 21 aseQTLs (Fig. 3) and affected 5 SNPs marking the ASE of the *CMYA5* gene.

Then, we performed the same analysis considering all SNPs located in transcripts showing ASE to observe if LD should be regarded on accounting for redundant aseQTL affecting the same allelic transcript. A total of 32 ASE SNPs pairs showed D′ > 0.8, being 31 with D′ = 1, distributed in 11 LD blocks (Supplementary Table 3). Seven blocks consisted of two neighboring SNPs marking ASE within the same gene, i.e., *ASB5*, *ATP1A2*, *TXNIP*, *NXN*, *MAP4*, *EIF5,* and *CAVIN4* genes. For the gene *XIRP2*, two ASE SNP pairs were within two LD blocks. Three SNPs inside unknown ASE genes were within the same LD block, whereas the larger LD block involved five SNPs, all of them inside the *CMYA5* gene.

Lastly, we evaluated LD values between SNPs marking ASE and their respective associated aseQTLs. From 1134 SNP pairs submitted to LD analysis, 238 showed D' higher than 0.8, and 71 aseQTL/ASE SNP pairs were in total disequilibrium (D' = 1) (Supplementary Table 4) corresponding to 35 ASE genes. The gene *CMYA5* was identified as presenting more aseQTL/ASE SNP interactions (38), followed by *EIF5* (33) and *VCP* (27).

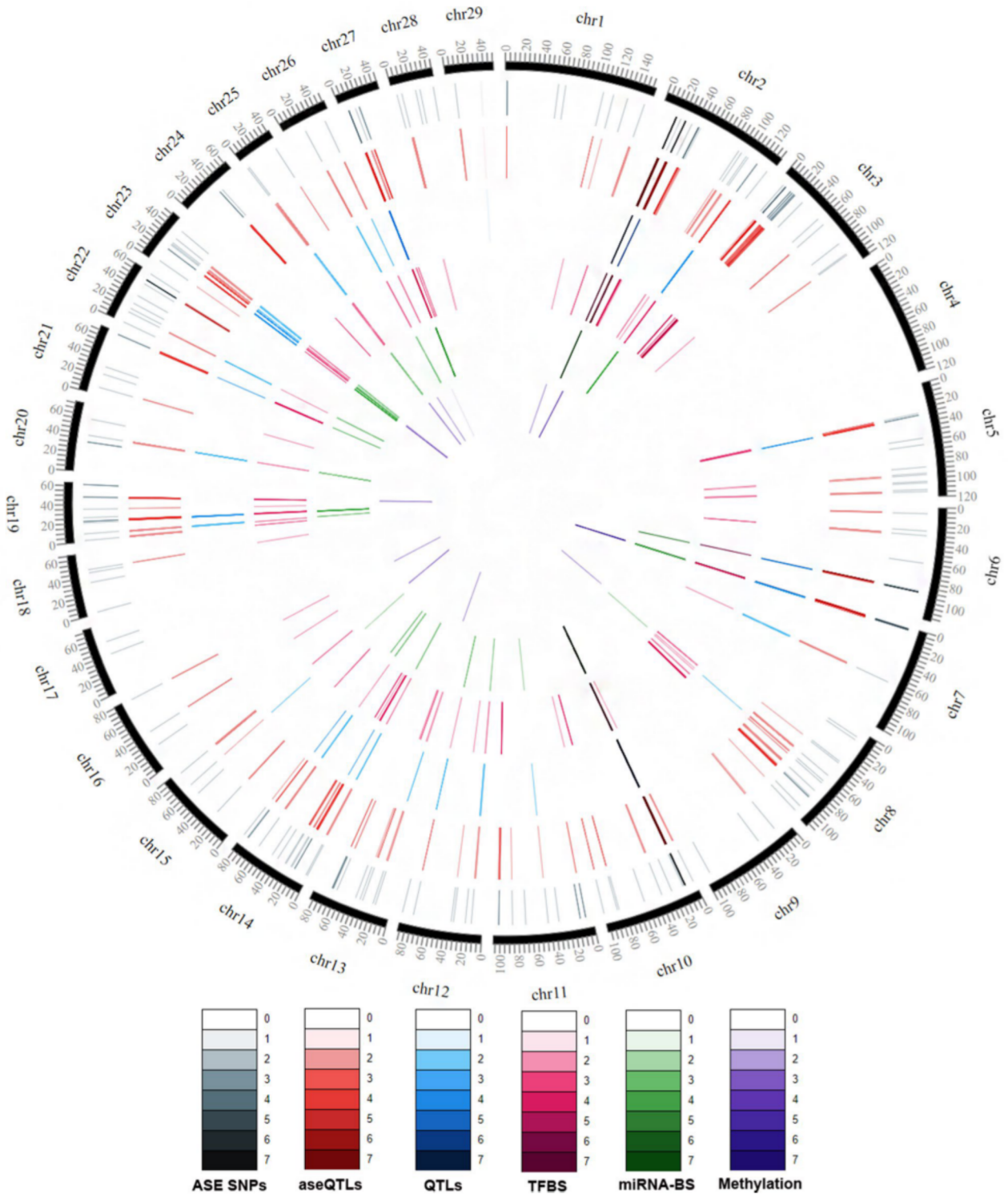

Figure 1. SNPs and aseQTLs distribution through the bovine chromosomes, overlap with previous QTL studies and predicted aseQTL regulatory mechanisms (TFBS, miRNA binding site, or methylated site). ASE SNPs: SNPs in transcripts with allele-specific expression, only the ASE SNPs with significant aseQTLs are displayed. aseQTL: identified allele-specific quantitative trait loci (aseQTLs). Overlapped: aseQTLs that overlaps with previous GWAS data made from our group. TFBS: aseQTLs that possibly enable the TF binding in only one allele. miRNA-BS: aseQTLs that potentially modify miRNA binding sites in only one allele. Methylated: methylated aseQTLs. The color intensity increases with increasing SNPs density in the plot scale (1 bp windows). This figure was made using the software CircosVCF[82].

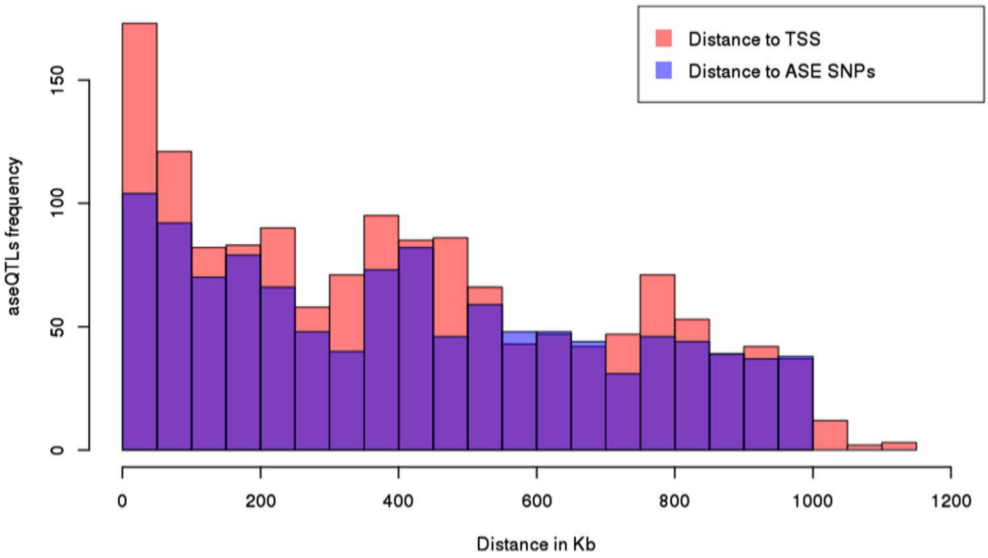

Figure 2. Distribution of aseQTLs concerning the distance to each associated ASE gene's transcription start site (TSS) in kb (in pink) and each associated SNP in kb (in blue). X-axis: distance in kb. Y-axis: frequency of aseQTL. Histogram plotted using R basic functions.

aseQTLs overlapping with QTLs. After converting the QTLs retrieved from the Cattle QTL Database to the current genome assembly build, ARS-UCD1.2, we identified overlaps of 173 aseQTLs within QTL regions for 21 traits, distributed in 215 occurrences (Supplementary Table 5). The permutation approach applied to test if the probability of the aseQTLs location within QTL regions was higher than chance, was significant (P value = 0.0001). The three traits with the highest number of aseQTLs within a QTL region were body weight (81 aseQTLs), milk palmitoleic acid content (42 aseQTLs), and intramuscular fat (IMF) (28 aseQTLs). The rs136717535 aseQTL overlapped with regions of more QTL traits: milk kappa-casein percentage, milk protein percentage, shear force, calf size, and birth index.

We also integrated data with other studies produced by our research group, with a broader sample of the same experimental population. All permutation tests performed for QTL ranges data were significant (P value ≤ 0.05). We identified 234 aseQTLs distributed in 894 overlaps with 25 traits (Supplementary Table 6). The aseQTLs that integrate with GWAS data from our research group are distributed in the blue ring of Fig. 1. Of these, we identified 847 overlaps with meat quality traits[21]. Table 1 shows the number of overlapping aseQTLs with QTLs associated with meat quality traits and the complete name of each trait.

The rs109550233 SNP, marker of allelic imbalance, presented more associated aseQTLs overlapping with QTLs for meat quality traits than any other in this study, with nine aseQTLs within regions of QTLs for nine traits, b*FAT, BFT, CL, L*FAT, L*MUSCLE, REA, WBSF0, WBSF7, and WHC, followed by the rs135906938 SNP (*ACOT13* gene), with eight aseQTLs in QTLs for eight meat quality traits, i.e., b*FAT, BFT, CL, L*FAT, L*MUSCLE, REA, WBSF7, and WHC. Thirty-five aseQTLs associated with six SNPs marking ASE of *CMYA5* overlapped with seven meat quality traits: b*FAT, CL, L*FAT, REA, WBSF0, WBSF7, and WHC.

Ten aseQTLs overlapped with mineral content QTLs for Co, Mn, Zn, Ca, Cr, Ar, K, Mg, and S[24]; seven overlapped with QTLs for IMF composition, which were also associated with octadecenoic acid[25]; Six overlapped with feed efficiency associated traits: efficiency of gain (EG), maintenance efficiency (ME) and partial efficiency of growth (PEG)[26]. Additionally, 20 aseQTLs overlapped with *cis*-eQTL regions and four with trans-eQTL regions[27].

Annotation of the aseQTL SNPs. According to aseQTL locations in the genome, we predicted SNP consequences with the VEP software[28] (Supplementary Table 7). AseQTLs were predominantly distributed in intronic regions (68%), followed by intergenic regions (27%). The remaining approximately 5% were distributed equally in non-coding transcription regions, 3′UTR variants, missense variants, and synonymous variants. Four SNPs, rs132671408 (in the *AOPEP* gene), rs135305605 (in the *RUSC2* gene), rs136155631 (in the *RAI14* gene) and rs43291678 (in the *CCDC141* gene) were within the 5′UTR regions and three were located in splice regions: rs109091165 (inside *PECAM1* gene), rs133325919 (inside *FAIM2* gene) and rs137526254 (inside *MYO6* gene).

Methylated aseQTLs. To analyze the methylation pattern, we compared the positions of aseQTLs with the methylation percentage of SNPs identified by RRBS in 12 samples. We identified 69 methylated cytosines in 16 aseQTLs, whose distribution is shown in the purple ring of Fig. 1. The rs110532113 aseQTL, associated with an ASE SNP on the *MSRB1* gene, was methylated in 11 of the 12 tested animals. Thirteen aseQTLs had 37 cytosines with 100% of methylation in the samples. Eight animals showed 100% of methylation in the rs134229733 aseQTL, and seven animals in the rs137496307 aseQTL, associated with an SNP marker of ASE of the *VIM* gene (Fig. 4; Supplementary Table 8).

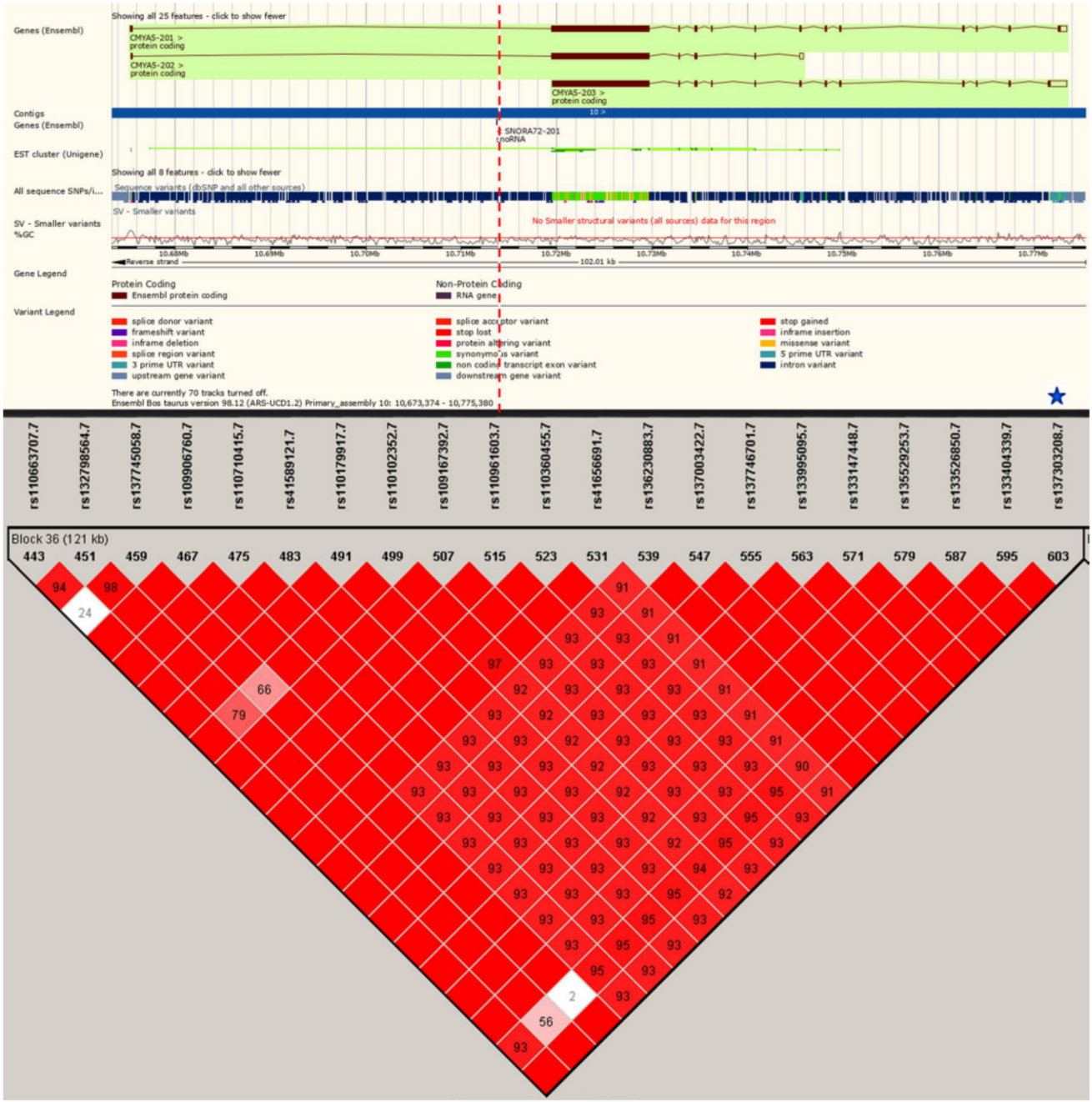

Figure 3. The biggest aseQTLs-LD block, located on chromosome 10, containing 21 aseQTLs. Red squares without numbers are SNPs in total LD, and the numbers within them indicate the intensity of the LD from 0 to 99. The intensity of the reddish tones increases proportionally to the LD values. The genomic position of the SNP rs137303208 is indicated with the dashed red line. All the SNPs in the block were within miRNAs binding sites. The blue star indicates that the SNP rs137303208 also was identified within TFBS. Gene representation was obtained in the Ensembl database and the LD plot was made using Haploview[76]. The two graphics were joined without any scale.

Six methylated aseQTLs were associated with SNPs inside transcribed regions with the allelic imbalance of seven genes: *HSPA1A*, *MSRB1*, *VCP*, *VIM*, *SCN4A*, *PECAM1,* and *ZDHHC4*, being the rs137668888 aseQTL simultaneously associated with the ASE marked by SNPs of *SCN4A* and *PECAM1* genes. The rs110407852 was the methylated aseQTL closest to the TSS region of the respective associated gene (*VCP*), at a distance of 51,026 bp from it. The average distance of these methylated aseQTLs and the TSS of associated ASE genes was 216,169 bp. The other ten methylated aseQTLs were associated with 9 ASE SNPs. The rs42616057 and rs109332326 aseQTLs were both methylated and associated with the allelic imbalance of rs110095343 SNP.

aseQTLs within TFBS. Using 50 bp flanking sequences of all aseQTLs as input in TRAP software[29], we found 83 TFBS overrepresented in our sequences (FDR < 0.05, Supplementary Table 9). Of them, only 12 TFs

Meat quality trait[a]	Number of aseQTLs	Affected ASE genes
Ribeye area (REA)	192	*ACOT13, ASB5, CA2, CAB39, CMYA5, CUEDC1, DNAJC21, ITGB1, KLF10, MBNL2, NXN, PTP4A3, RGCC, SPARC, STBD1, XIRP1, XIRP2*
Lightness of fat (L*FAT)	127	*ACOT13, ASB5, CAB39, CMYA5, DNAJC21, ITGB1, KLF10, NXN, SPARC, STBD1*
Warner–Bratzler shear force 7 days after slaughter (WBSF7)	122	*ACOT13, ASB5, CAB39, CMYA5, DNAJC21, ITGB1, KLF10, SPARC, STBD1*
Water holding capacity (WHC)	113	*ACOT13, ASB5, CAB39, CMYA5, DNAJC21, ITGB1, KLF10, SPARC, STBD1*
Cooking loss (CL)	105	*ACOT13, ASB5, CMYA5, DNAJC21, ITGB1, SPARC, STBD1*
Yellowness of fat (b*FAT)	100	*ACOT13, CAB39, CMYA5, DNAJC21, ITGB1, KLF10, SPARC, STBD1*
Warner–Bratzler shear force 24 h after slaughter (WBSF0)	46	*CMYA5, PFKM*
Lightness muscle (L*MUSCLE)	23	*ACOT13, SPARC,*
Backfat thickness (BFT)	17	*ACOT13, STBD1*
Warner–Bratzler shear force 14 days after slaughter (WBSF14)	2	*SPARC*

Table 1. Allele-specific expression quantitative trait loci (aseQTLs) overlapping meat quality traits identified in Nelore muscle and allele specific expression (ASE) genes affected by aseQTLs.

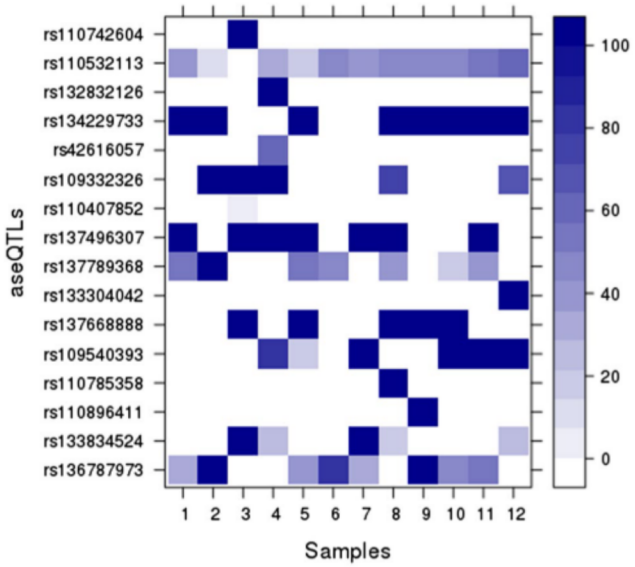

Figure 4. Percentage of aseQTLs methylation in 12 Nelore animals. The color intensity increases proportionally according to the methylation percentage for each SNP per sample. Heatmap plotted with Lattice R package (http://lattice.r-forge.r-project.org/).

are expressed in bovine muscle[30]. Considering only aseQTLs that change muscle-expressed TFBS, we found 215 aseQTLs, distributed in the pink ring of Fig. 1. From these, 150 presented putative TFBS with the reference allele and 117 with the alternative allele. The SMAD4 binding site was found affected by 37 aseQTLs; AHR by 31 aseQTLs; VDR and SP3 by 26 aseQTLs; STAT3 by 25 aseQTLs; PPARA by 24 aseQTLs; DR1, TAL1, and ZNF219 by 20 aseQTLs; STAT6, SP1, and MZF1 TFs were affected by 16, 12, and 10 aseQTLs, respectively.

aseQTLs within miRNA binding sites. We used a subset of 163 aseQTLs identified outside intergenic regions and that overlapped with our reference population's QTL data for miRNA binding site prediction. We identified 153 miRNA binding sites corresponding to 163 aseQTLs flanking sequences. Considering miRNA binding sites affected by aseQTL, we obtained 1448 aseQTLs/miRNA binding site pairs. In silico analysis of these pairs predicted 693 miRNA binding sites with the alternative aseQTL allele, whereas 755 binding sites were predicted in the reference aseQTL allele (green ring of Fig. 1). The pairs were formed with 162 aseQTLs and 149 miRNAs, being 114 miRNAs expressed in the muscle of our population[31]. These muscle-expressed miRNAs were denoted without an asterisk in Supplementary Table 10, which contains the results of miRNA binding sites prediction.

We analyzed whether the ASE genes were previously identified as targets for the miRNAs affected by the associated aseQTLs with the MiRWalk tool[32]. Twenty genes associated with aseQTLs that potentially change miRNA binding sites' affinity were used in the MiRWalk analysis. These genes were targets for 583 miRNAs. From these, 31 miRNAs had their binding sites affected by the aseQTLs predicted to regulate 14 ASE genes, which were themselves identified as targets for the corresponding miRNAs.