


RESEARCH ARTICLE

Open Access



Expression quantitative trait loci in sheep liver and muscle contribute to variations in meat traits

Zehu Yuan^{1,2,3}, Bolormaa Sunduimijid², Ruidong Xiang^{2,4}, Ralph Behrendt⁵, Matthew I. Knight⁵, Brett A. Mason², Coralie M. Reich², Claire Prowse-Wilkins², Christy J. Vander Jagt², Amanda J. Chamberlain², Iona M. MacLeod², Fadi Li¹, Xiangpeng Yue^{1*} and Hans D. Daetwyler^{2,6*} 

Abstract

Background: Variants that regulate transcription, such as expression quantitative trait loci (eQTL), have shown enrichment in genome-wide association studies (GWAS) for mammalian complex traits. However, no study has reported eQTL in sheep, although it is an important agricultural species for which many GWAS of complex meat traits have been conducted. Using RNA sequence data produced from liver and muscle from 149 sheep and imputed whole-genome single nucleotide polymorphisms (SNPs), our aim was to dissect the genetic architecture of the transcriptome by associating sheep genotypes with three major molecular phenotypes including gene expression (geQTL), exon expression (eeQTL) and RNA splicing (sQTL). We also examined these three types of eQTL for their enrichment in GWAS of multi-meat traits and fatty acid profiles.

Results: Whereas a relatively small number of molecular phenotypes were significantly heritable ($h^2 > 0$, $P < 0.05$), their mean heritability ranged from 0.67 to 0.73 for liver and from 0.71 to 0.77 for muscle. Association analysis between molecular phenotypes and SNPs within ± 1 Mb identified many significant cis-eQTL (false discovery rate, FDR < 0.01). The median distance between the eQTL and transcription start sites (TSS) ranged from 68 to 153 kb across the three eQTL types. The number of common variants between geQTL, eeQTL and sQTL within each tissue, and the number of common variants between liver and muscle within each eQTL type were all significantly ($P < 0.05$) larger than expected by chance. The identified eQTL were significantly ($P < 0.05$) enriched in GWAS hits associated with 56 carcass traits and fatty acid profiles. For example, several geQTL in muscle mapped to the *FAM184B* gene, hundreds of sQTL in liver and muscle mapped to the *CAST* gene, and hundreds of sQTL in liver mapped to the *C6* gene. These three genes are associated with body composition or fatty acid profiles.

Conclusions: We detected a large number of significant eQTL and found that the overlap of variants between eQTL types and tissues was prevalent. Many eQTL were also QTL for meat traits. Our study fills a gap in the knowledge on the regulatory variants and their role in complex traits for the sheep model.

*Correspondence: lexp@lzu.edu.cn; hans.daetwyler@agriculture.vic.gov.au

¹ State Key Laboratory of Grassland Agro-ecosystems; Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs; Grassland Agriculture Engineering Center, Ministry of Education; College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou 730020, People's Republic of China

² Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC 3083, Australia

Full list of author information is available at the end of the article

Background

Gene expression varies between tissues and individuals [1, 2], and their phenotypes can be shaped by gene expression [3]. Gene and exon expression levels can be directly quantified by counting the RNA sequence (RNA-seq) reads that map to the gene and its exons [4, 5]. RNA-splicing can be estimated by isoform ratios,



exon inclusion levels [6] and intron excision ratios [7]. There are statistical challenges to estimate isoform abundance [8] and technical effects [7] to estimate exon inclusion levels when using conventional short-read data. Therefore, calculating the excised intron ratios could be an accurate method to quantify RNA-splicing, because excised introns are inferred directly from short reads that span exon–exon junctions [9]. Thus, the three types of molecular phenotypes, i.e. gene expression, exon expression and intron excision ratio, can be accurately quantified via RNA-seq.

Different DNA mutations among individuals can result in varied molecular phenotypes in a population. A mutation that regulates nearby (± 1 Mb) molecular phenotypes from RNA-seq data (i.e. gene expression, exon expression and intron excision ratio) is defined as a cis expression quantitative trait locus or eQTL (eQTL are defined here as gene expression QTL (geQTL), exon expression QTL (eeQTL), and splicing QTL (sQTL)). In beef cattle, geQTL have been investigated in liver and muscle [10–12]. Some of the geQTL that have been identified in humans, cattle, and pigs overlap with (or colocalize at) single nucleotide polymorphisms (SNPs) that are linked with diseases and complex traits [3, 13–16]. For example, ten QTL linked with pork meat quality colocalized with 12 eQTL in the *longissimus dorsi* muscle [16]. Therefore, eQTL analysis may improve our knowledge on the genetic architecture of complex traits. In dairy cattle, Xiang et al. [12] showed that eeQTL overlapped with geQTL and sQTL. Consequently, the combined analysis of geQTL, eeQTL and sQTL may increase the chance to identify loci that regulate gene expression [12, 17]. However, to our knowledge, eQTL have not been analyzed in sheep, although it is one of the major farm species for human consumption.

Sheep and lamb meat consumption continues to grow worldwide due to the animal's adaptability to be farmed in different climatic conditions and to its unique culinary value [18]. Meat production and quality traits are polygenic and are influenced by many QTL, each making a small contribution to a trait [19, 20]. The polygenicity of sheep meat traits is supported by recent genome-wide association studies (GWAS) with SNPs [21–27]. In particular, Bolormaa et al. [28] used high-density SNPs (~ 500 K) to identify pleiotropic SNPs that are associated with 56 carcass composition traits, and Rovadoscki et al. [29] used 50 K SNPs to identify SNPs that are linked with fatty acid profiles [29]. The results from GWAS have laid the foundation for studying sheep meat characteristics that are regulated by genetic variants. However, the biological mechanisms that underlie how these QTL affect meat traits are not well known. In

this study, we use eQTL data to begin to explore these mechanisms.

Using imputed whole-genome SNPs and RNA-seq data, our aim was to quantify the heritability of three molecular phenotypes and identify cis eQTL in sheep liver and muscle. Then, we investigated if the same SNPs were significantly associated with multiple eQTL phenotypes and/or if the eQTL were significant in both tissues. We also investigated if eQTL were enriched in GWAS QTL of multiple sheep meat traits [28, 29]. Our results contribute to a better understanding of how genomic differences in sheep lead to variation in complex meat traits.

Methods

An overview of the analysis is shown in Additional file 1: Figure S1.

Sample collection and RNA extraction

In total, 149 crossbred wether lambs (ewes: Merino \times Border Leicester and Maternal/Coopworth Composites, nine sires: Polled Dorset and White Suffolk) between 7 and 8 months old were randomly selected for RNA-seq analysis from 436 male lambs that were balanced across dam and lamb nutritional treatments, birth types, breeds and sires [30]. Lambs were slaughtered in their experimental blocks on three slaughter dates [30]. Liver and *longissimus dorsi* muscle were sampled within 10 min from slaughter at the abattoir and were flash-frozen in liquid nitrogen and stored at -80°C until use. Frozen liver and muscle tissues were ground using the Geno/Grinder[®] 2010 (SPEX[™], Metuchen, NJ, US). Ground tissue was homogenized in Trizol[®] (Life Technologies[™]) and RNA extracted using the Trizol[®] plus RNA extraction kit (Life Technologies[™]) according to the manufacturer's instructions. RNA quality and integrity were evaluated by the 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany). All samples with an RNA integrity number larger than 7 and a 28S/18S ratio higher than 1.0 were used for library preparation using the Sure-Select Strand Specific RNA Library Prep Kit (Agilent Technologies).

RNA-seq and data quality control

Libraries were randomly pooled and sequenced on a HiSeq2000 genome analyzer (Illumina Inc) in a paired end 100 cycle run to produce 20 million paired-reads per library. CASAVA v1.8 (Illumina Inc) was used to call fastq files. QuadTrim (<https://bitbucket.org/arobinson/quadttrim>) was used to trim adapter and low-quality bases (quality score < 20) from each end of the reads, and then to discard low-quality reads (failed chastity filter, or mean quality score < 20 , or $N_s > 3$, or final length < 50). Finally,

only paired-reads that passed the quality control criteria were kept for downstream analysis.

Quantification of molecular phenotypes

Clean paired-reads were aligned to the sheep reference genome Oar_v3.1 (ftp://ftp.ensembl.org/pub/release-91/fasta/ovis_aries/dna/) using STAR [31] along with the annotation file (Ovis_aries.Oar_v3.1.91.gtf.gz, containing 27,054 genes). The parameters used for alignment are documented in Additional file 2: Table S1. Parameters were chosen to maximize uniquely mapped reads [32]. Each Sequence Alignment Map (SAM) file was sorted and transformed into a Binary Alignment Map (BAM) file by SAMtools [33]. The proportion of reads that were uniquely mapped was summarized from STAR alignment results (see Additional file 4: Table S2). Gene saturation was assessed as the number of detected genes per percent of total reads at increasing read depths, with reads sampled from each BAM file. Saturation of splice junctions and 3'/5' bias were investigated using RseQC [34].

Read counts of each gene and each exon were calculated by FeatureCounts [5]. In each tissue, genes and exons with a count per million (CPM) higher than 1 in more than 30 individuals (>20%) were retained, and $\log_{10}(x+1)$ transformed. Intron excision ratios were quantified by LeafCutter [7] as the change in intron usage in the intron cluster of which it is a member, where several intron clusters are often inferred within a gene. Intron excision ratios higher than 0 in at least 30 animals (>20%) and mean ratio values higher than 0.01 across all individuals were retained and $\log_{10}(x+1)$ transformed, intron-wise quantile normalized and individual-wise Z-score standardized [12].

Estimation of the heritability of molecular phenotypes

In total, 13,243 genes, 63,872 exons and 91,699 intron excision events in liver and 12,989 genes, 60,230 exons and 87,257 intron excision events in muscle were used to estimate heritabilities. Each molecular phenotype was adjusted for fixed effects, including slaughter day, feed lot pen number and replicate, dam breed, sire breed, birth type and dam gestational body condition score, using the `lm()` function in R (<https://www.r-project.org/>). Heritabilities (h^2) of molecular phenotypes, which are defined here as the proportion of the phenotypic variance explained by all the SNPs, were estimated using the ASReml[®] software [35] by fitting the following linear mixed model:

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{a} + \mathbf{e},$$

where \mathbf{y} is an $n \times 1$ vector of the molecular phenotypes of all individuals ($n=149$), $\mathbf{1}_n$ is a vector of 1s, μ is the

overall mean, \mathbf{a} is an $n \times 1$ vector of additive genetic effects, following a normal distribution $\mathbf{N} \sim (0, \mathbf{G}\sigma_a^2)$, where \mathbf{G} is the genomic relationship matrix (GRM) [36] calculated from the ~500 K SNP panel for 149 individuals, σ_a^2 is additive genetic variance, \mathbf{e} is a $n \times 1$ vector of random residuals, following a normal distribution $\mathbf{N} \sim (0, \mathbf{I}\sigma_e^2)$. The heritability of molecular phenotypes was calculated as:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2}.$$

The differences in heritability between the three molecular phenotypes within a tissue were tested by the `wilcox.test()` function in R.

Imputed whole-genome SNPs

All 149 individuals were genotyped with the high-density Ovine SNP Beadchip (HD, ~500 K SNP) and imputed to whole-genome sequence genotypes as part of a study by Bolormaa et al. [37]. Briefly, 935 animals with whole-genome sequence variants (117 pure Merino, 726 European breeds, 92 other breeds) were selected as the reference population (<https://www.ebi.ac.uk/eva/?eva-study=PRJEB31241>) to impute the target population (~47,000). Both reference and target population HD genotypes were phased using the Eagle software with default parameters and the target population genotypes were imputed to whole-genome sequence using the Minimac3 software with default parameters [38, 39]. Finally, SNPs with an R^2 (imputation quality index reported from Minimac3) lower than 0.4 were removed to reduce the impact of poorly imputed SNPs. On average, the empirical imputation accuracy across all target breeds was 0.97. Imputed variants in the 149 individuals with a minor allele frequency (MAF) lower than 0.05 were removed to avoid spurious associations in the subsequent analyses.

Cis eQTL detection

Molecular phenotypes were tested for association with all the SNPs within ± 1 Mb of each molecular feature (to reduce computational burden). In total, 12,373 genes, 56,233 exons, and 79,146 intron excision events in liver and 12,151 genes, 53,660 exons and 76,824 intron excision events in muscle located on autosomes were used for eQTL mapping. Association testing between each molecular phenotype and SNPs was implemented one SNP at a time using the Wombat software [40] by fitting the following linear mixed model:

$$\mathbf{y} = \mathbf{1}_n\mu + \mathbf{Z}\beta + \mathbf{a} + \mathbf{e},$$

where \mathbf{y} is a $n \times 1$ vector of the molecular phenotypes of all individuals ($n=149$ in the current study), \mathbf{Z} is a $n \times 1$ vector of SNP genotypes coded as 0, 1 or 2, β is marker effect, \mathbf{a} is a $n \times 1$ vector of polygenic effects, following a normal distribution $N \sim (0, \mathbf{G}\sigma_a^2)$, \mathbf{e} is a $n \times 1$ vector of random residuals, following a normal distribution $N \sim (0, \mathbf{I}\sigma_e^2)$. For each molecular phenotype, variance components estimated from ASreml[®] were regarded as prior information for Wombat to implement association testing. When a SNP was significantly associated with multiple molecular phenotypes, only the most significant one was retained. Finally, SNPs with a false discovery rate (FDR) lower than 0.01 were regarded as significant cis eQTL.

Transcription start site (TSS) coordinates of all genes were obtained from Ensembl via BioMart (<http://www.ensembl.org>). The absolute distance between a cis eQTL and the TSS of the eQTL anchored gene was calculated. The eQTL functional catalogues were determined using the NGS-SNP pipeline [41] with the Variant Effect Predictor software [42].

Relationships between eQTL type and their effect across tissues

It is possible that the three eQTL types are not independent. geQTL are associated with a change in total gene expression, which can affect all the isoforms or only one of them. In addition, eeQTL are associated with a change in exon expression, which can affect all the exons in a gene, and therefore total gene expression, or only one exon and therefore a single isoform. Thus, it is likely that some geQTL are also eeQTL or sQTL or both. For this reason, we tested the overlap between the three eQTL types, i.e. we tested whether a single variant was significantly associated with two molecular phenotypes. In addition, it is possible that the same eQTL has an effect in both tissue types. Thus, we tested the overlap between tissues for each eQTL type, i.e. we tested whether a single variant was significantly associated with a particular molecular phenotype in both tissues.

The significance (p-values) of overlaps between two sets (e.g. set A and set B) were calculated with the hypergeometric enrichment test R function `phyper`, which requires the number of elements in the background set W ($N, A \subseteq W, B \subseteq W$), the number of elements in set A (n), the number of elements in set B (M), and the number of elements in the intersection of A and B ($m, A \cap B$). Equivalently, one can use the `newGeneOverlap` and `testGeneOverlap` functions in the `GeneOverlap` R package [43]. When testing for significance of overlaps between different eQTL types within and across tissues, N is the number of SNPs used for eQTL identification ($N=20,824,844$ in this study). For overlaps of geQTL

between different tissues, N is the number of expressed genes in both liver and muscle. Similarly, to test the overlap of eeQTL and sQTL between tissues, N is the number of expressed exons and intron excision events in both liver and muscle, respectively.

Enrichment of eQTL in GWAS hit regions

Because the GWAS of Bolormaa et al. [28] and Rovadoscki et al. [25] included many detailed post-slaughter phenotypes in sheep, we used their results to investigate the overlap with the eQTL identified in our study. However, based on the difference in SNP densities in these studies (Bolormaa et al. [28] (~500 K) and Rovadoscki et al. [29]) (~50 K) and in our analysis (~20 million), it was unlikely that eQTL would overlap exactly with GWAS SNPs. Thus, we investigated the overlap of eQTL in a defined interval of 25 kb up- and down-stream of significant (FDR < 0.01) pleiotropic SNPs (all 932 significant pleiotropic SNPs from the multi-trait meta-analysis in Bolormaa et al. [28] and significant regions in Rovadoscki et al. [29]). The OvineSNP50 BeadChip mean gap size between probes was 50.9 kb (median = 42.6 kb), thus the use of 25-kb intervals should capture most of the eQTL in these regions. The hypergeometric test (Eq. 1) was used to investigate whether identified eQTL were significantly enriched in GWAS hit regions, where N is the number of all SNPs used in the eQTL analysis ($N=20,824,844$ in this study); n is the number of significant eQTL; M is the number of SNPs in the current study that are located in GWAS hit regions and m is the number of significant eQTL in M .

Results

Data quality

In total, 20,824,844 SNPs passed quality control and were used for eQTL association testing. The distribution of these SNPs is shown in Additional file 3: Figure S2A. After filtering the low-quality raw reads, 8379 million clean read pairs from 298 samples were retained, i.e. an average of 27.76 million read pairs per sample (see Additional file 4: Table S2). Roughly, 7107 million clean read pairs were uniquely mapped to the sheep genome, i.e. an average of 23.85 million read pairs per sample (see Additional file 4: Table S2). The gene body plots of all expressed genes indicated little 5' bias (see Additional file 3: Figure S2B). In the RNA-seq technology, saturation is reached when an increment in the number of reads does not result in the detection of additional expressed genes or in the calling of more features, e.g., splice junctions [44]. Our gene saturation analysis showed that the number of detected genes increased with increasing percent of total reads and that, when the percent of reads reached 75% of the total reads, few additional genes were

detected [see Additional file 3: Figure S2C]. This result indicates that most of the expressed genes were detected by the RNA-seq study. The splice junction saturation analysis revealed that when the percent of reads reached 100%, the number of all detected splice junctions (including annotated and novel splice junctions) did not reach an asymptote (see Additional file 3: Figure S2D). Almost all annotated splice junctions were rediscovered (see Additional file 3: Figure S2E), but novel splice junctions did not reach an asymptote (see Additional file 3: Figure S2F). This result suggests that some novel splice junctions were still missing at this level of coverage.

Heritability of molecular phenotypes

Molecular phenotypes with a heritability (h^2) higher than 0 (Z test, $P < 0.05$) were considered heritable. In liver, the mean heritabilities across the 290 gene expression (2.19%), 7014 exon expression (10.98%) and 2389 intron excision ratios (2.61%) of heritable molecular phenotypes were 0.73, 0.67 and 0.71, respectively (Table 1), and in muscle, the mean heritabilities across the 487 of gene expression (3.75%), 12,599 exon expression (20.92%) and 4198 intron excision ratios (4.81%) of heritable molecular phenotypes were 0.77, 0.73 and 0.71, respectively. Both in liver and muscle, the level of heritable exon expression was higher than that of gene expression and intron excision ratio.

In liver, 640,976 (FDR < 0.01) geQTL, 376,181 eeQTL and 678,657 sQTL were identified (Table 2) and (see Additional file 5: Figure S3A). The geQTL, eeQTL and sQTL in liver were associated with 3631, 2265 and 2633 genes, respectively (Table 2). In muscle, 356,380 geQTL, 223,900 eeQTL and 383,044 sQTL were identified (Table 2) and (see Additional file 5: Figure S3A). The geQTL, eeQTL and sQTL in muscle were associated with 2396, 1564 and 2047 genes, respectively

Table 2 Summary of detected expression quantitative trait loci (eQTL), including gene expression QTL, (geQTL), exon expression QTL, (eeQTL), and splicing QTL, (sQTL) including number of QTL detected, number of genes in which those QTL were located and number of eQTL per gene

Tissue	Cis eQTL type	eQTL number	Gene number	eQTL per gene
Liver	geQTL	640,976	3631	176.53
	eeQTL	376,181	2265	166.08
	sQTL	678,657	2633	257.75
Muscle	geQTL	356,380	2396	148.74
	eeQTL	223,900	1564	143.16
	sQTL	383,044	2047	187.12

(Table 2). The number of significant eQTL was larger than the number of tagged genes, which indicated that many SNPs were in high linkage disequilibrium (LD) or that there were multiple variants associated with the same gene (Table 2).

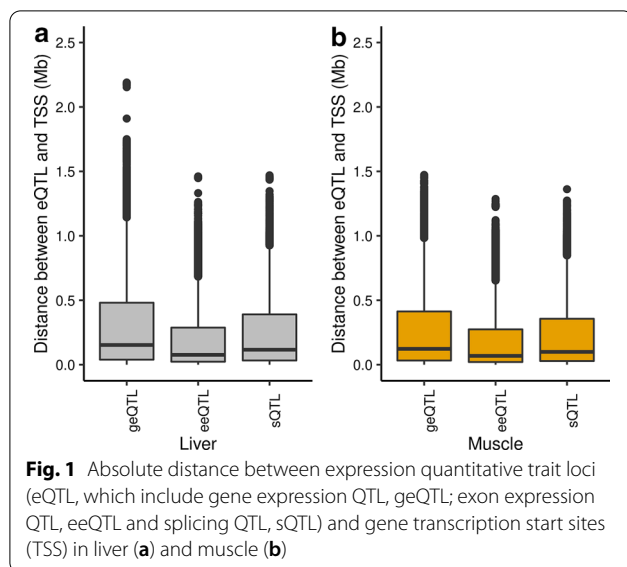
In liver, the median distances between the geQTL, eeQTL, and sQTL and TSS were 153, 76 and 116 kb, respectively (Fig. 1a), and in muscle, they were 123, 68 and 99 kb, respectively (Fig. 1b). Both in liver and muscle, eeQTL were closer to TSS than geQTL or sQTL.

Compared to the total number of SNPs in the genome, a decreased proportion of intergenic SNPs and an increased proportion of intronic, missense, synonymous, gene-end, UTR and splicing SNPs were found for the three eQTL types (Table 3). Whereas we expected sQTL to be enriched in the annotated ‘Splice’ category, followed by eeQTL, and geQTL [12], we found that that all three eQTL types shared a similar proportion of ‘Splice’ variants both in liver and muscle. However, the absolute numbers of ‘Splice’ sQTL, eeQTL and geQTL are in the expected order.

Table 1 Summarized information for molecular phenotype heritability

Molecular phenotype	Total number	Number of heritable phenotypes	Heritable ratio	Mean heritability (standard deviation)
Liver gene expression	13,243	290	0.0219	0.73 (0.14)
Liver exon expression	63,872	7014	0.1098	0.67 (0.13)
Liver intron excision	91,699	2389	0.0261	0.72 (0.14)
Muscle gene expression	12,989	487	0.0375	0.77 (0.12)
Muscle exon expression	60,230	12,599	0.2092	0.73 (0.13)
Muscle intron excision	87,257	4198	0.0481	0.71 (0.13)

Details include the total number of molecular phenotypes (total number), the number of molecular phenotypes that were heritable (number of heritable phenotypes, heritability > 0, $P < 0.05$), the ratio of heritable molecular phenotypes to total number (heritable ratio) and the mean heritability of the heritable molecular phenotypes (mean heritability) and standard deviations



Relationship between eQTL types and their effect across tissues

The amount of overlap of SNPs between the three eQTL types within each tissue was significantly ($P < 0.05$) larger than expected by chance (Fig. 2). The largest number of shared SNPs was observed between geQTL and eeQTL (Fig. 2), followed by that between geQTL and sQTL, and between eeQTL and sQTL.

A significant ($P < 0.05$) overlap of geQTL, eeQTL and sQTL between liver and muscle was observed (see Additional file 6: Figure S4a). The largest number of overlapping SNPs between liver and muscle was observed for sQTL, followed by eeQTL, and geQTL (see Additional file 6: Figure S4a). Genes, exons and intron excision events tagged by eQTL were shared at a significant level

between liver and muscle (see Additional file 6: Figure S4B). Interestingly, when eQTL were significant in both liver and muscle, their effect (t-value) was usually in the same direction (see Additional file 6: Figure S4c), with eeQTL being the most consistent.

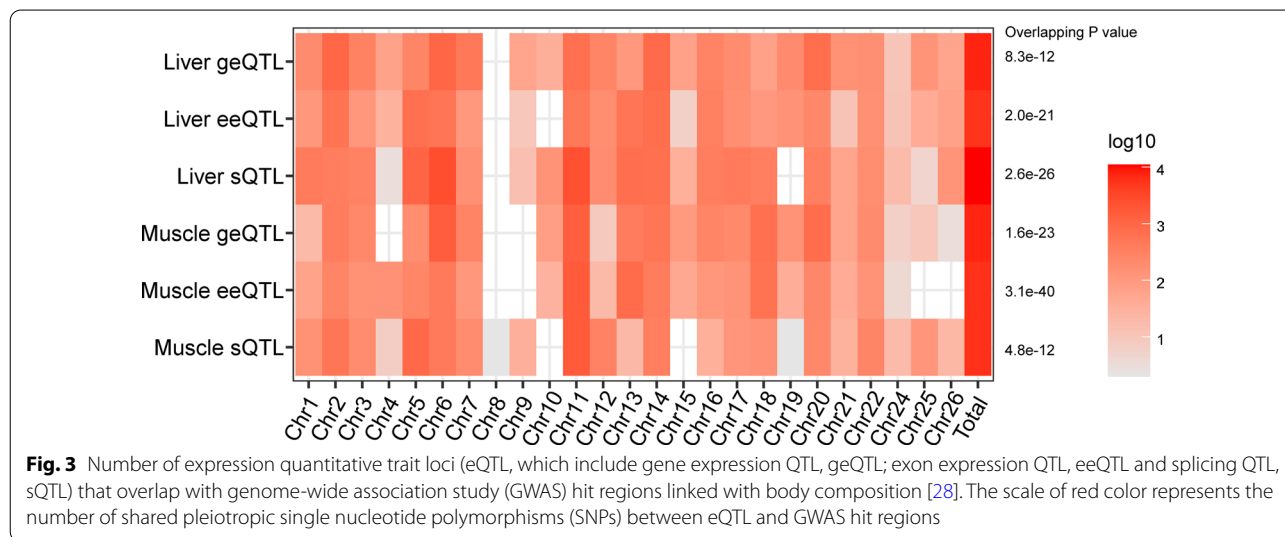
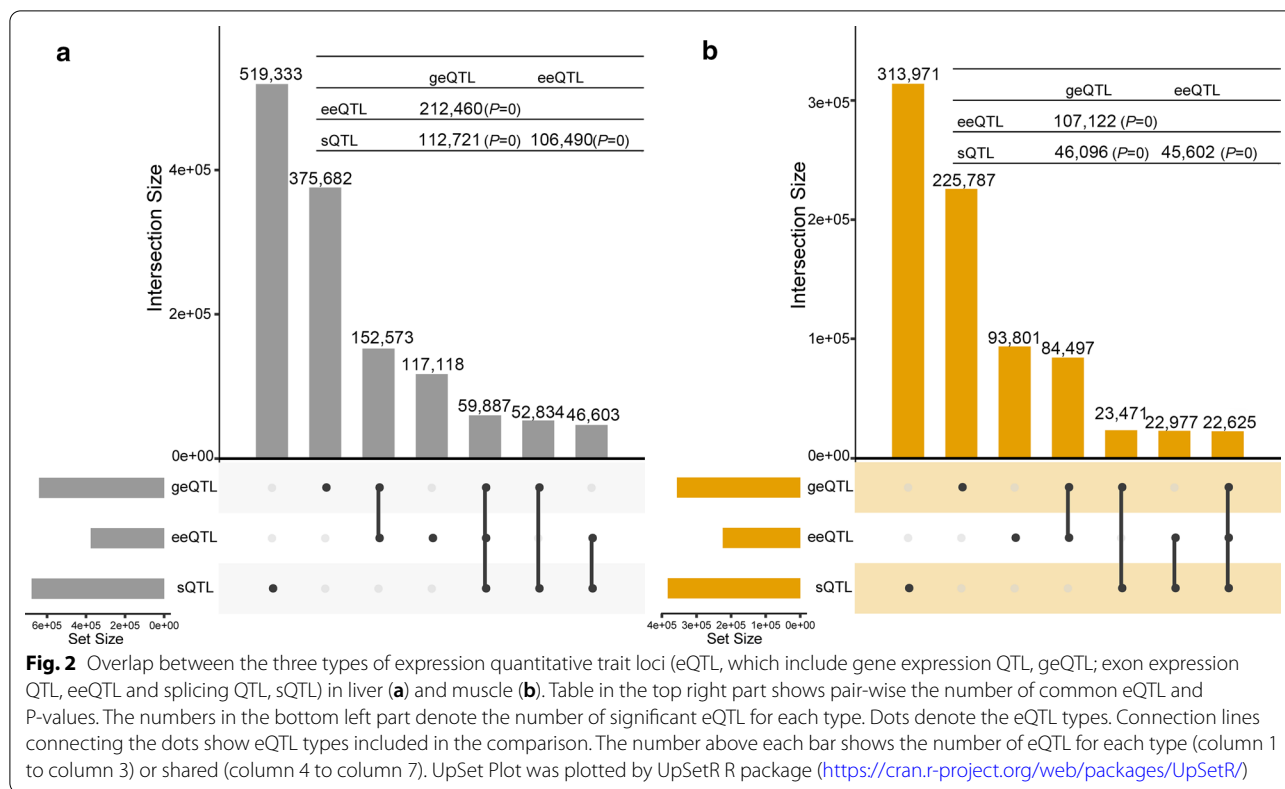
We used the results from two published GWAS [28] and [25] for meat traits to evaluate the significant regions and eQTL in our study. First, 1130 pleiotropic SNP regions (from a multi-trait meta-analysis, $FDR < 0.01$) [28] were used to assess the overlap between these and eQTL. A summary of the results is shown in Fig. 3 and the detailed results are documented in Additional file 7: Table S3. All three types of eQTL in liver and muscle were significantly enriched in GWAS hit regions ($P < 0.05$, Fig. 3). In total, 43.45% (491/1130), 26.99% (305/1130) and 52.12% (589/1130) of the GWAS hit regions identified by Bolormaa et al. [28] were covered by geQTL, eeQTL and sQTL in liver, respectively, and 43.98% (497/1130), 26.02% (294/1130) and 30.62% (346/1130) of the GWAS hit regions were covered by geQTL, eeQTL and sQTL in muscle, respectively.

Second, we used 27 significant regions linked with fatty acid profiles [29] to investigate their overlap with eQTL from our study. Generally, liver eQTL were more enriched in genomic regions linked with fatty acid profile than muscle eQTL (Table 4). The identified geQTL in liver were significantly ($P < 0.05$) enriched in genomic regions that were associated with saturated fatty acids (SFA), polyunsaturated fatty acids (PUFA), the ratio of linoleic acid ($\omega 6$) to alpha-linolenic acid ($\omega 3$) and the ratio of PUFA to SFA, the identified eeQTL in liver were significantly ($P < 0.05$) enriched in genomic regions that are associated with PUFA, and the sQTL were significantly ($P < 0.05$) enriched in genomic regions that are associated with SFA, monounsaturated fatty acids

Table 3 Proportional functional annotation of expression quantitative trait loci (eQTL, including gene expression QTL, geQTL; exon expression QTL, eeQTL and splicing QTL, sQTL) and all single nucleotide polymorphisms (SNPs) used for eQTL detection

Functional category	All	Liver			Muscle		
		geQTL	eeQTL	sQTL	geQTL	eeQTL	sQTL
Splice	0.0007	0.0016	0.0021	0.0019	0.0018	0.0026	0.0022
UTR	0.0022	0.0057	0.0076	0.0060	0.0069	0.0081	0.0072
Gene_end	0.0606	0.1219	0.1371	0.1226	0.1316	0.1477	0.1340
Synonymous_variant	0.0032	0.0075	0.0102	0.0085	0.0085	0.0124	0.0096
Missense_variant	0.0018	0.0047	0.0057	0.0052	0.0047	0.0060	0.0056
Intron_variant	0.3032	0.4327	0.4540	0.4434	0.4356	0.4700	0.4658
Intergenic	0.6271	0.4249	0.3822	0.4113	0.4095	0.3520	0.3746
Others	0.0010	0.0011	0.0010	0.0012	0.0015	0.0013	0.0011

All eQTL were annotated as intergenic, intronic, missense, synonymous, up or downstream (Gene_end), in untranslated regions (UTR, 3' or 5'), having some splicing function (Splice) or other (all other categories)



(MUFA), and ω_6/ω_3 ratios. In muscle, geQTL were significantly ($P < 0.05$) enriched in genomic regions that are associated with SFA and MUFA. Both the identified eeQTL and sQTL in muscle were significantly ($P < 0.05$) enriched in genomic regions that are associated with SFA and PUFA.

Examples of QTL that may be eQTL

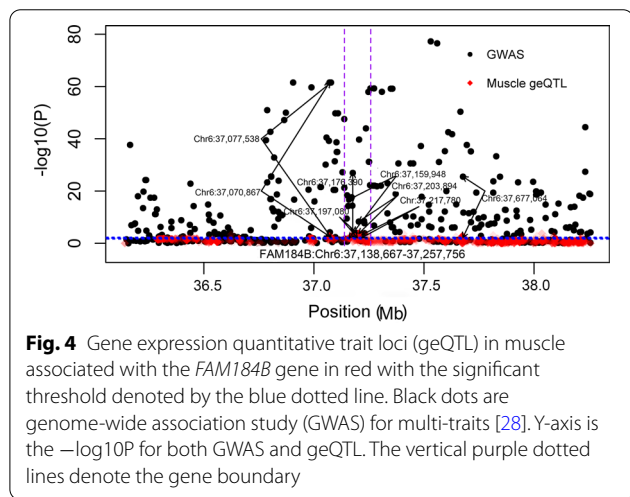
The *FAM184B* gene (*family with sequence similarity 184 member B*; Chr6:37138667–37257756) is located in a candidate QTL region (close to Chr6:37.53 Mb) for meat traits in sheep, which was identified by using the 50 K [27] and HD SNPs chips [28]. Eight SNPs (Chr6:37070867,

Table 4 Expression quantitative trait loci (eQTL, which include gene expression QTL (geQTL), exon expression QTL (eeQTL), and splicing QTL (sQTL)) enriched in genome-wide association study (GWAS) hit regions linked with fatty acid profiles [29]

Fatty acid	Liver geQTL	Liver eeQTL	Liver sQTL	Muscle geQTL	Muscle eeQTL	Muscle sQTL
Saturated fatty acid (SFA)						
Myristic acid (C14:0)	612 (1.2e-33)	229 (0.15)	707 (9e-51)	1255 (0e+00)	1176 (0.00)	233 (0.15)
Palmitic acid (C16:0)	406 (2.6e-26)	0 (1.00)	0 (1.00)	0 (1.00)	0 (1.00)	340 (1.2e-48)
Stearic acid (C18:0)	296 (6.7e-03)	5 (1.00)	570 (5.4e-59)	997 (0.00)	43 (1.00)	150 (0.6)
Total SFA	617 (5.8e-08)	3 (1.00)	1978 (0.00)	55 (1.00)	43 (1.00)	5 (1.00)
Monounsaturated fatty acids (MUFA)						
Palmitoleic acid (C16:1)	764 (1.7e-13)	9 (1.00)	2031 (0.00)	947 (8.8e-178)	61 (1.00)	147 (1.00)
Oleic acid (C18:1)	94 (1.00)	5 (1.00)	226 (1.00)	61 (1.00)	36 (1.00)	56 (1.00)
Total MUFA	221 (1.00)	3 (1.00)	1505 (7.8e-152)	32 (1.00)	36 (1.00)	0 (1.00)
Polyunsaturated fatty acids (PUFA)						
Linoleic acid (C18:2, ω6)	396 (1.1e-51)	0 (1.00)	0 (1.00)	0 (1.00)	0 (1.00)	0 (1.00)
alpha-Linolenic acid (C18:3, ω3)	1172 (2.1e-116)	696 (9.6e-71)	292 (1.00)	818 (4.6e-128)	329 (2.1e-18)	183 (1.00)
Conjugated linoleic acid (CLA, c9t11)	277 (2e-17)	96 (0.45)	142 (0.99)	88 (0.58)	6 (1.00)	201 (3e-21)
Total ω3	0 (1.00)	197 (1.5e-03)	127 (1.00)	0 (1.00)	0 (1.00)	65 (1.00)
Total ω6	396 (1.6e-37)	0 (1.00)	0 (1.00)	0 (1.00)	6 (1.00)	0 (1.00)
Total PUFA	396 (6.7e-10)	0 (1.00)	0 (1.00)	1 (1.00)	6 (1.00)	0 (1.00)
ω6/ω3 and PUFA/SFA ratios						
ω6/ω3	418 (1.00)	13 (1.00)	1481 (4.4e-128)	208 (1.00)	36 (1.00)	0 (1.00)
PUFA/SFA	396 (3.6e-154)	0 (1.00)	0 (1.00)	0 (1.00)	0 (1.00)	0 (1.00)
Total	6849			4610		

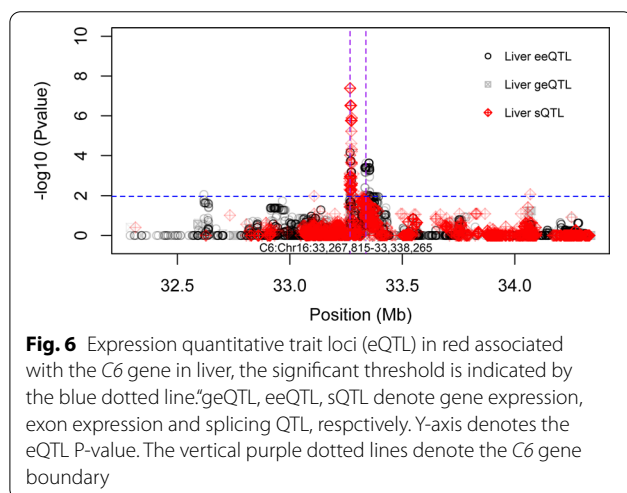
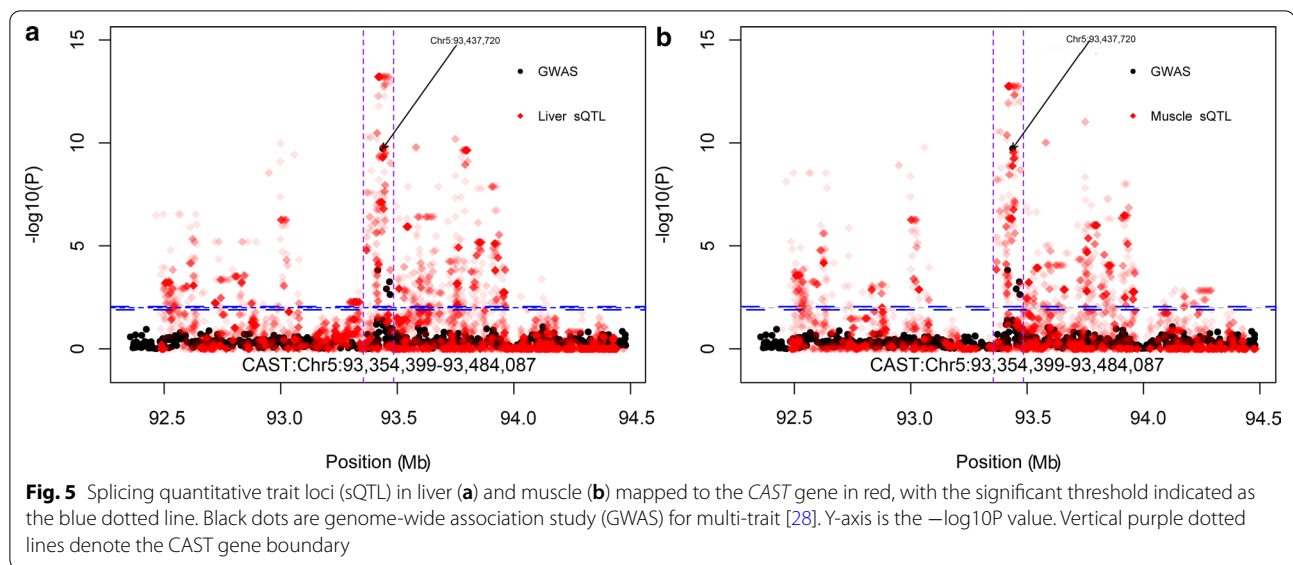
The number in each cell denotes the number of eQTL located in genomic regions linked with a fatty acid trait

The number in parentheses denotes the P-value for overlap between the two datasets



Chr6:37077538, Chr6:37159948, Chr6:37176390, Chr6:37197080, Chr6:37203894, Chr6:37217780 and Chr6:37677064) that are located in close proximity to *FAM184B* and that have been identified as significant pleiotropic SNPs linked with meat traits [28], were also geQTL in muscle for *FAM184B* (Fig. 4) in our study.

The most significant SNP (Chr5:93437720) associated with tenderness [28] was located in the *CAST* gene (*calpastatin*; Chr5:93354399–93484087) (Fig. 5). We found 1985 sQTL in liver associated with three intron excision events in *CAST* (Chr5:93439378–93444596 and Chr5: 93394527–93427992), and 1800 sQTL in muscle associated with intron excision events (Chr5:93435918–93437744, Chr5:93439378–93444596 and Chr5:93482729–93483763) in *CAST*. Some sQTL in liver and muscle were associated with the same intron excision region (Chr5:93439378–93444596), which was between exon 9 (Chr5:93439292–93439378) and exon 11 (Chr5:93444596–93444694) of *CAST* (Fig. 5). A putative QTL region (Chr16:33207525–33550464) associated with alpha-linolenic acid (ω3) [29] included the *C6* (*complement C6*, Chr16:33267815–33338265) gene. In total, 146 eeQTL and 128 sQTL in liver mapped to the *C6* gene QTL region (Fig. 6). The 146 eeQTL were associated with the expression of two exons in the *C6* gene (exon1, Chr16:33267815–33267879; and exon 18, Chr16:33338090–33338265). The 128 sQTL were associated with four intron excision regions (Chr16:33266238–33267815, Chr16:33266238–33272429, Chr16:33266695–33272429 and Chr16:



33267879–33272429), for which, the closely located exon (*C6*: Chr16:33267815–33267879, exon1) was regulated by an eeQTL.

Discussion

Our study is the first that analyses the associations between sequence variants and cis molecular phenotype variation (± 1 Mb) at transcriptomic levels (including gene expression, exon expression and splicing) in sheep liver and muscle.

We estimated the heritability of global gene expression, exon expression, and intron excision events in sheep, using SNPs. Molecular phenotypes were considered heritable when they were significantly different from zero. In humans, the median SNP heritability of gene expression

has been reported to range from 0.1 to 0.34 [45–51] for 31 to 85% of the heritable molecular phenotypes across tissues (including lymphoblastoid cell lines, lymphocytes, blood and adipose tissue) [45–47, 49, 51]. In pig, the reported mean SNP heritabilities of molecular phenotypes range from 0.18 to 0.51 [16, 52]. In the current study, the number of heritable molecular phenotypes was smaller than in most previous publications. However, on average, the non-zero heritability estimates of molecular phenotypes are high, and our estimation of the heritability of molecular phenotypes is higher than that reported in humans and pigs. In part this is due to our relatively small sample size, which will cause heritabilities to be overestimated. However, the transcriptome is also the first level at which DNA mutations have an impact, once properly captured, the molecular heritability is expected to be higher than the heritability of complex traits, which is impacted by DNA mutations via several omics-layers [53]. In addition, measuring gene expression based on RNA-seq data is quite precise, compared to gross measurements of phenotypic traits (e.g., body weight). This can reduce the error variance in some molecular phenotypes, which in part increases the signal-to-noise ratio in the estimation of molecular heritability.

We found that a relatively low proportion of molecular phenotypes reached a heritability that was significantly different from zero, which could be due to several reasons. First, whereas our study is the largest sheep eQTL study to date, its sample size (i.e. 149 lambs) was modest, which can lead to large variance errors for the heritability estimates. As such, many of the molecular phenotypes with a moderate heritability estimate would have not been deemed significant due to the large standard error.

With a larger dataset, it is likely that the number of significant molecular phenotypes would increase and the mean heritability estimates would decrease as more molecular phenotypes with a moderate heritability reached significance. Second, gene expression is transient and often time-point- and tissue-specific, and so lower expression levels may lead to lower heritability estimates. We studied only two tissues in 149 sheep, and it would be interesting to study the molecular heritability of other sheep tissues in the future. Third, our dataset allowed us to test only SNP or additive heritability of molecular phenotypes, although regulatory relationships between SNPs and adjacent genes can be largely non-additive. Future studies using much larger sample sizes should test for non-additive effects to recapture the ‘missing’ heritability in gene expression.

This study only considered cis eQTL to reduce the number of tests [54]. A larger number of tests for trans eQTL would have required a more stringent threshold and likely would have reduced the power to detect cis eQTL. In addition, cis eQTL may be easier to detect because their effects are larger than those of trans eQTL [54]. The majority of the identified sheep eQTL were within 250 kb from TSS, which is consistent with the results from human studies [55]. The identified eeQTL in sheep were closer to TSS than the sQTL and geQTL, which disagrees with a study in cattle that reported that sQTL were closer to TSS than other types of eQTL [12] and that used a different and more stringent definition for sQTL. Indeed, they report only intron excision events that had an adjacent exon inclusion event, which is an exon usage phenotype, i.e. exon read count/gene read count [12]. In our study, all sQTL were intron excision events only. In human and cattle, sQTL have been reported to be enriched in intron variants [12, 55, 56] compared to all the SNPs detected. Consistently, in our study, eeQTL in liver and muscle had a higher proportion of intron SNPs than geQTL. However, in sheep liver and muscle, the identified sQTL did not have a higher proportion of intron SNPs than eeQTL. This inconsistent result could be explained by differences in the definition of significant sQTL, in computational methods, and in sample sizes, tissues and species between these studies.

In humans, significant commonalities of sQTL between tissues and of geQTL between tissues were reported [57, 58]. Recent studies have suggested that, in cattle, three types of eQTL (sQTL, eeQTL and geQTL) were shared between tissues [12] and that, in pig, geQTL were extensively shared between liver and muscle [59]. Our results support the extensive sharing of all eQTL types between liver and muscle, but disagree with other studies [47, 60]. The ratio of eQTL overlap between tissues depends on the similarity of biological functions, with more similar

tissues sharing more eQTL [61]. Our results support the significant overlap of eQTL between liver and muscle [see Additional file 5: Figure S3], which could be due to the similarity of biological functions that regulate meat traits between liver and muscle.

A recent eQTL study in humans, which detected geQTL, eeQTL, and sQTL from blood samples, showed significant enrichment for GWAS SNPs for disease or complex traits [62]. Similarly, we found that the three types of eQTL in sheep liver and muscle were significantly enriched in GWAS hit regions that were previously identified by Bolormaa et al. [28]. Whereas two published cattle studies showed a relatively limited overlap between eQTL and single-trait GWAS SNPs [11, 54], we found a relatively high overlap between eQTL and multi-trait QTL. Several reasons might have contributed to our results. First, the pleiotropic SNPs from multi-trait GWAS (multi-trait meta-analysis) may be more informative than SNPs from single-trait GWAS [28, 63] and this is supported by the relatively few eQTL that we observed in several genomic regions linked with single traits (e.g. fatty acid profiles). In addition, we used GWAS hit regions instead of exact overlaps (SNP to SNP), which could increase the probability of overlap because eQTL may be in LD with QTL. This is in line with a cattle study published in 2019, which used eQTL SNPs to build functional GRM that explained a large amount of the genetic variance in 34 complex traits [64]. Moreover, the relationship between the phenotypes included in the GWAS and the tissues used for the eQTL analysis may affect the overlap between GWAS and eQTL. For example, a recent study in cattle suggested that 10 of 163 SNPs associated with stature were identified as geQTL in white blood cells [13]. However, SNPs associated with milk production and fertility did not overlap with geQTL identified in white blood cells [54]. These results show that we still have a very limited understanding of how eQTL and trait QTL relate, and thus, future research in this area is required.

We highlight a few examples of QTL that may also be eQTL. In sheep, *FAM184B* was located in the most significant GWAS region that is associated with body weight [27] and body composition traits [28]. In cattle, the expression level of *FAM184B* was suggested to be regulated by cis eQTL [54]. In our study, five significant pleiotropic SNPs located near *FAM184B* [28] also affected the total expression level of *FAM184B* significantly and confirmed the previously reported association.

Our second example highlights *CAST*. In sheep, the most significant SNP (Chr5:93437720) associated with shear force (i.e. meat tenderness) was within the *CAST* gene [28]. Twenty-one isoforms of the human *CAST* gene are included in RefSeq (<https://www.ncbi.nlm.nih.gov/gene/831#>), and similarly, several isoforms of

bovine *CAST* [65] are found in RefSeq (<https://www.ncbi.nlm.nih.gov/gene/281039>). This evidence lends support to our finding of an intron excised region (i.e. exon Chr5:93439378-93444596) detected in both sheep liver and muscle that indicates alternative splicing (skipping exon 10), but this mechanism needs to be validated using molecular techniques. The expression level of exon 10 in *CAST* was low and was removed from the association analysis in our study, which provides additional support for this exon-skip alternative splicing. In humans, SNPs have been identified that regulate alternative splicing in *CAST* [66]. We found that many sQTL in sheep liver and muscle mapped to this intron excised region (Chr5:93439378-93444596) within *CAST*, but unlike a study in cattle [11], we found no significant geQTL for *CAST*. Taken together, these results suggest that SNPs that are linked with tenderness (shear force) in sheep might regulate the expression of *CAST* by excluding the 10th exon from the transcript, while not altering the total level of expression. Recent research in humans suggested that a splice variant (rs7724759) in *CAST* affected exon abundance instead of gene abundance [67], which supports our finding.

The $\omega 3$ and $\omega 6$ polyunsaturated fatty acid composition of sheep meat and lamb is important for human health [68] and depends on both the dietary intake of the animal and the cellular metabolism, which is often controlled by genetic polymorphisms [69]. One example of a gene that controls $\omega 3$ and $\omega 6$ metabolism in the cell is the *C6* gene. This gene is located in a putative QTL region associated with alpha-linolenic acid (C18:3 $\omega 3$) and total $\omega 3$ [29]. *C6* has been reported to be specifically highly expressed in sheep liver [70], which supports its potential role in the fatty acid profile of meat, because the liver plays an essential role in fatty acid synthesis. A protective effect of *C6* deficiency on the development of diet-induced atherosclerosis has been observed when *C6*-deficient rabbits were fed a cholesterol-rich diet for 14 weeks [71], which suggests that it might play a role in lipid metabolism. In our study, the 143 significant sQTL in which the intron excision event had a flanking exon were identified as eeQTL, which supports the reliability of intron excision region detection. Since many significant liver sQTL for *C6* were located in a QTL region that is linked with $\omega 3$ polyunsaturated fatty acid, it could be an eQTL for *C6* splicing level. Therefore, we cautiously speculate that SNPs, which regulate the content of $\omega 3$ polyunsaturated fatty acid in meat, may do so by regulating *C6* alternative splicing in liver, which might affect cellular lipid metabolism.

As the first detailed analysis of sheep cis eQTL, our study has its limitations. Many factors affect the performance of fine mapping, e.g., the local LD structure and

sample size [72]. The strong LD between closely located SNPs results in multiple SNPs appearing as significantly linked with a molecular phenotype when analyzing one SNP at a time. We used 149 lambs from nine sires, which means that the local LD may be relatively strong and leads to broad eQTL peaks. In our study, the sample size was relatively small, which reduces the power to detect eQTL [54], whereas in the current human eQTL studies sample sizes reach tens of thousands of individuals [73].

Conclusions

A relatively small number of molecular phenotypes had a SNP heritability significantly different from zero, but many significant cis eQTL were detected. These were often associated with several eQTL types and were significant in both liver and muscle tissue. The identified eQTL were significantly enriched in previously reported GWAS regions for meat traits, for example several geQTL in muscle mapped to *FAM184B*, hundreds of sQTL in liver and muscle mapped to *CAST*, and hundreds of sQTL in liver mapped to *C6*.

Supplementary information

The online version contains supplementary material available at <https://doi.org/10.1186/s12711-021-00602-9>.

Additional file 1: Figure S1. Overview of the analysis. In total, 298 RNA-seq data in liver and muscle from 149 crossbred male wether lambs were aligned to the sheep reference genome Oar_v3.1 (ftp://ftp.ensembl.org/pub/release-91/fasta/ovis_aries/dna/) using STAR along with the annotation file (Ovis_aries.Oar_v3.1.91.gtf.gz, containing 27,054 genes). Gene and exon expression levels were quantified by counting the reads of the gene and exon using FeatureCount. RNA-splicing was estimated by calculating intron excision ratio using Leafcutter. Heritability (h^2) of the three molecular phenotypes (gene expression, exon expression and intron excision ratio) were estimated using ASReml[®]. Wombat software was used to identify cis expression quantitative trait loci (eQTL, which include gene expression QTL (geQTL); exon expression QTL (eeQTL) and splicing QTL (sQTL)) within 1 Mb of the gene, exon or intron excision event. We investigated the overlap between eQTL and two genome-wide association studies (GWAS), the characteristics of eQTL and the relationship between different tissues, and between different eQTL types.

Additional file 2: Table S1. STAR parameters used for alignment.

Additional file 3: Figure S2. Data information. a: Distribution of imputed whole-genome single nucleotide polymorphisms (SNPs) in the sheep genome. Horizontal axis is the size of the chromosome, vertical axis is the chromosome number (from chromosome 1 to 26), the color scale denotes the number of SNPs within 1 Mb-windows. b: 3'/5' bias. The plots of coverage for all expressed genes in liver (grey) and muscle (yellow) indicated little 5' bias. Error bars represent the standard error of coverage for the 149 samples. c: Gene saturation in liver (grey) and muscle (yellow). Gene and splice junction saturation is reached when an increment in the number of reads does not result in additional expressed genes being detected or in more features, e.g., splice junctions, called. Error bars represent the standard error for the number of detected genes. d: Saturation of total splice junctions in liver (grey) and muscle (yellow). Error bars represent the standard error for the detected splice junctions. e: Saturation of annotated splice junctions in liver (grey) and muscle (yellow). f: Saturation of novel

splice junctions in liver (grey) and muscle (yellow). Error bars represent the standard error for the detected splice junctions.

Additional file 4: Table S2. RNA-seq library information. Mean, minimum, maximum and median values of the number of read pairs that pass trimming and filtering (clean reads), and are uniquely mapped reads, and the proportion of clean reads that map uniquely to the genome in liver and muscle samples. The distribution of the proportion of clean reads that are uniquely mapped in liver and muscle samples is also shown.

Additional file 5: Figure S3. Circle Manhattan plot of expression quantitative trait loci (eQTL, which include gene expression QTL (geQTL); exon expression QTL (eeQTL) and splicing QTL (sQTL)) in liver (a) and muscle (b). From inside to outside, the circle Manhattan plot denotes geQTL, eeQTL and sQTL, respectively. Red dash line in each Manhattan plot represents the threshold ($FDR < 0.01$). Circle Manhattan plots were plotted using the CMAplot R package (<https://github.com/YinLiLin/R-CMAplot>).

Additional file 6: Figure S4. Overlap between liver and muscle for the three types of expression quantitative trait loci (eQTL, which include gene expression QTL (geQTL); exon expression QTL (eeQTL) and splicing QTL (sQTL)). a: Venn diagrams showing the expression quantitative trait loci detected in liver and in muscle, and in both. b: Gene expression, exon expression, and intron excision events with eQTL detected in liver and in muscle, and in both. c: Correlation of eQTL effects (t-value of eQTL) for which the eQTL were significant both in liver and muscle.

Additional file 7: Table S3. Significant expression quantitative trait loci (eQTL, which include gene expression QTL (geQTL), exon expression QTL (eeQTL), and splicing QTL (sQTL)) in liver and muscle overlapping with genome-wide association study (GWAS) hit regions linked with body composition [28]. This file contains six columns. Column 1: tissue type, namely, muscle or liver; column 2, eQTL type, namely, geQTL, eeQTL or sQTL; column 3: eQTL ID; column 4: eQTL FDR-value; column 5: GWAS SNP ID; column 6: GWAS FDR-values.

Acknowledgements

The authors thank the many Agriculture Victoria staff involved in animal husbandry and sample collection. The HD BeadChip was developed under the auspices of the International Sheep Genomics Consortium, in close collaboration between AgResearch New Zealand, CSIRO Livestock Industries, Baylor College of Medicine, and Agriculture Victoria, with work underwritten by FarnIQ (<http://www.farniq.co.nz>), a joint New Zealand government and industry Primary Growth Partnership programme. The authors thank James Kijas, Rudiger Brauning, Alan McCulloch, and Sean McWilliams for their contribution to SheepGenomesDB (<http://sheepgenomedb.org>) and all institutions that have made their sheep WGS data public. We are grateful to Prof. Paul Stothard and team at the University of Alberta for providing much of the functional annotation of sequence variants.

Authors' contributions

ZY, HDD, AJC, MIK, RB conceived and designed the study; MIK, RB, BAM, CMR, CPW, CJV, AJC, IMM, HDD sampled tissues; BAM, CMR, CPW generated RNA sequence and genotype data; ZY, BS, RX, CJV, IM analyzed data; HDD, AJC, IMM, FL, XY supervised the team; ZY, RX, AJC, CJV, IMM, and HDD wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was funded by the National Natural Science Foundation of China (31872319) and Agriculture Victoria's Growing Food and Fibre project. China Scholarship Council (CSC) funded ZY for his study in AgriBio, Centre for AgriBioscience. The original experiment from which the samples were collected was funded by Agriculture Victoria and Meat and Livestock Australia.

Availability of data and materials

Whole-genome sequence genotypes used for imputation are available at the European Variant Archive (<https://www.ebi.ac.uk/eva/?eva-study=PRJEB31241>). Raw RNA sequence read data are available at NCBI PRJNA689847. Processed datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. The datasets supporting the conclusions of this article are included in the article.

Ethics approval and consent to participate

All animal procedures were approved by DJPR Agriculture Research and Extension Animal Ethics Committee (AEC Approval: 2012–02) and conducted in accordance with Australian code for the care and use of animals for scientific purposes [74].

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹ State Key Laboratory of Grassland Agro-ecosystems; Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs; Grassland Agriculture Engineering Center, Ministry of Education; College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou 730020, People's Republic of China. ² Agriculture Victoria, AgriBio, Centre for AgriBioscience, Bundoora, VIC 3083, Australia. ³ Institutes of Agricultural Science and Technology Development (Joint International Research Laboratory of Agriculture & Agri-Product Safety), Yangzhou University, Yangzhou, 225000, People's Republic of China. ⁴ Faculty of Veterinary & Agricultural Science, University of Melbourne, Parkville, VIC 3010, Australia. ⁵ Agriculture Victoria, Hamilton Centre, Hamilton, VIC 3300, Australia. ⁶ School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3083, Australia.

Received: 20 January 2020 Accepted: 8 January 2021

Published online: 18 January 2021

References

1. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science*. 2015;348:648–60.
2. Mele M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. The human transcriptome across tissues and individuals. *Science*. 2015;348:660–5.
3. Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. *Nat Rev Genet*. 2015;16:197–212.
4. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31:166–9.
5. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923–30.
6. Anders S, Reyes A, Huber W. Detecting differential usage of exons from RNA-seq data. *Genome Res*. 2012;22:2008–17.
7. Li YI, Knowles DA, Humphrey J, Barbeira AN, Dickinson SP, Im HK, et al. Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet*. 2018;50:151–8.
8. Lacroix V, Sammeth M, Guigo R, Bergeron A. Exact transcriptome reconstruction from short sequence reads, algorithms in bioinformatics. In: Crandall KA, Lagergren J, editors. *Algorithms in bioinformatics*. WABI 2008. Lecture notes in computer science. Berlin: Springer; 2008. p. 50–63.
9. Vaquero-Garcia J, Barrera A, Gazzara MR, Gonzalez-Vallinas J, Lahens NF, Hogenesch JB, et al. A new view of transcriptome complexity and regulation through the lens of local splicing variations. *Elife*. 2016;5:e11752.
10. Higgins MG, Fitzsimons C, McClure MC, McKenna C, Conroy S, Kenny DA, et al. GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and GFRA2 expression in beef cattle. *Sci Rep*. 2018;8:14301.
11. Khansefid M, Pryce JE, Bolormaa S, Chen Y, Millen CA, Chamberlain AJ, et al. Comparing allele specific expression and local expression quantitative trait loci and the influence of gene expression on complex trait variation in cattle. *BMC Genomics*. 2018;19:793.
12. Xiang R, Hayes BJ, Vander Jagt CJ, MacLeod IM, Khansefid M, Bowman PJ, et al. Genome variants associated with RNA splicing variations in bovine are extensively shared between tissues. *BMC Genomics*. 2018;19:521.
13. Bouwman AC, Daetwyler HD, Chamberlain AJ, Ponce CH, Sargolzaei M, Schenkel FS, Sahana G, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet*. 2018;50:362–7.

14. Brown AA, Vinuela A, Delaneau O, Spector TD, Small KS, Dermitzakis ET. Predicting causal variants affecting expression by using whole-genome sequencing and RNA-seq from multiple human tissues. *Nat Genet.* 2017;49:1747–51.
15. Gonzalez-Prendes R, Quintanilla R, Canovas A, Manunza A, Figueiredo Cardoso T, Jordana J, et al. Joint QTL mapping and gene expression analysis identify positional candidate genes influencing pork quality traits. *Sci Rep.* 2017;7:39830.
16. Velez-Irizarry D, Casiro S, Daza KR, Bates RO, Raney NE, Steibel JP, et al. Genetic control of longissimus dorsi muscle gene expression variation and joint analysis with phenotypic quantitative trait loci in pigs. *BMC Genomics.* 2019;20:3.
17. Guan L, Yang Q, Gu M, Chen L, Zhang X. Exon expression QTL (eeQTL) analysis highlights distant genomic variations associated with splicing regulation. *Quant Biol.* 2014;2:71–9.
18. Knapik J, Ropka-Molik K, Pieszka M. Genetic and nutritional factors determining the production and quality of sheep meat—a review. *Ann Anim Sci.* 2017;17:23–40.
19. Andersson L, Georges M. Domestic-animal genomics: deciphering the genetics of complex traits. *Nat Rev Genet.* 2004;5:202–12.
20. Gao Y, Zhang R, Hu X, Li N. Application of genomic technologies to the improvement of meat quality of farm animals. *Meat Sci.* 2007;77:36–45.
21. Zhang L, Liu J, Zhao F, Ren H, Xu L, Lu J, et al. Genome-wide association studies for growth and meat production traits in sheep. *PLoS One.* 2013;8:e66569.
22. Matika O, Riggio V, Anselme-Moizan M, Law AS, Pong-Wong R, Archibald AL, et al. Genome-wide association reveals QTL for growth, bone and in vivo carcass traits as assessed by computed tomography in Scottish Blackface lambs. *Genet Sel Evol.* 2016;48:11.
23. Kominakis A, Hager-Theodorides AL, Zoidis E, Saridaki A, Antonakos G, Tsiamis G. Combined GWAS and 'guilt by association'-based prioritization analysis identifies functional candidate genes for body size in sheep. *Genet Sel Evol.* 2017;49:41.
24. Garza Hernandez D, Mucha S, Banos G, Kaseja K, Moore K, Lambe N, et al. Analysis of single nucleotide polymorphisms variation associated with important economic and computed tomography measured traits in Texel sheep. *Animal.* 2018;12:915–22.
25. Pasandideh M, Rahimi-Mianji G, Gholizadeh M. A genome scan for quantitative trait loci affecting average daily gain and Kleiber ratio in Baluchi Sheep. *J Genet.* 2018;97:493–503.
26. Ghasemi M, Zamani P, Vatankhah M, Abdoli R. Genome-wide association study of birth weight in sheep. *Animal.* 2019;13:1797–803.
27. Al-Mamun HA, Kwan P, Clark SA, Ferdosi MH, Tellam R, Gondro C. Genome-wide association study of body weight in Australian Merino sheep reveals an orthologous region on OAR6 to human and bovine genomic regions affecting height and weight. *Genet Sel Evol.* 2015;47:66.
28. Bolormaa S, Hayes BJ, van der Werf JH, Pethick D, Goddard ME, Daetwyler HD. Detailed phenotyping identifies genes with pleiotropic effects on body composition. *BMC Genomics.* 2016;17:224.
29. Rovadoscki GA, Pertile SFN, Alvarenga AB, Cesar ASM, Pertile F, Petrini J, et al. Estimates of genomic heritability and genome-wide association study for fatty acids profile in Santa Ines sheep. *BMC Genomics.* 2018;19:375.
30. Knight MI, Butler KL, Slocombe LL, Linden NP, Raeside MC, Burnett VF, et al. Reducing the level of nutrition of twin-bearing ewes during mid to late pregnancy produces leaner prime lambs at slaughter. *Animal.* 2020;14:864–72.
31. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics.* 2013;29:15–21.
32. Baruzzo G, Hayer KE, Kim EJ, Di Camillo B, FitzGerald GA, Grant GR. Simulation-based comprehensive benchmarking of RNA-seq aligners. *Nat Methods.* 2017;14:135–9.
33. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25:2078–9.
34. Wang L, Wang S, Li W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics.* 2012;28:2184–5.
35. Gilmore AR, Gogel BJ, Cullis BR, Welham SJ, Thompson R. ASReml User Guide Release 4.1 functional specification. Hemel Hempstead: VSN International Ltd; 2014.
36. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. *Nat Genet.* 2010;42:565–9.
37. Bolormaa S, Chamberlain AJ, Khansefid M, Stothard P, Swan AA, Mason B, et al. Accuracy of imputation to whole-genome sequence in sheep. *Genet Sel Evol.* 2019;51:1.
38. Loh PR, Danecek P, Palamara PF, Fuchsberger C, Reshef YA, Finucane HK, et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat Genet.* 2016;48:1443–8.
39. Das S, Forer L, Schonherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet.* 2016;48:1284–7.
40. Meyer K. WOMBAT: a tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). *J Zhejiang Univ Sci B.* 2007;8:815–21.
41. Grant JR, Arantes AS, Liao X, Stothard P. In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. *Bioinformatics.* 2011;27:2300–1.
42. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, et al. The Ensembl Variant Effect Predictor. *Genome Biol.* 2016;17:122.
43. Shen L. GeneOverlap: An R package to test and visualize gene overlap. 2014. <https://www.bioconductor.org/packages/release/bioc/vignettes/GeneOverlap/inst/doc/GeneOverlap.pdf>. Accessed 3 Dec 2020.
44. Tarazona S, Garcia-Alcalde F, Dopazo J, Ferrer A, Conesa A. Differential expression in RNA-seq: a matter of depth. *Genome Res.* 2011;21:2213–23.
45. Monks SA, Leonardson A, Zhu H, Cundiff P, Pietrusiak P, Edwards S, et al. Genetic inheritance of gene expression in human cell lines. *Am J Hum Genet.* 2004;75:1094–105.
46. Goring HH, Curran JE, Johnson MP, Dyer TD, Charlesworth J, Cole SA, et al. Discovery of expression QTLs using large-scale transcriptional profiling in human lymphocytes. *Nat Genet.* 2007;39:1208–16.
47. Price AL, Helgason A, Thorleifsson G, McCarrroll SA, Kong A, Stefansson K. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals. *PLoS Genet.* 2011;7:e1001317.
48. Dixon AL, Liang L, Moffatt MF, Chen W, Heath S, Wong KC, et al. A genome-wide association study of global gene expression. *Nat Genet.* 2007;39:1202–7.
49. Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, et al. Genetics of gene expression and its effect on disease. *Nature.* 2008;452:423–8.
50. Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet.* 2012;44:1084–9.
51. Lloyd-Jones LR, Holloway A, McRae A, Yang J, Small K, Zhao J, et al. The genetic architecture of gene expression in peripheral blood. *Am J Hum Genet.* 2017;100:228–37.
52. Liaubet L, Lobjois V, Faraut T, Tirczacs A, Benne F, Iannuccelli N, et al. Genetic variability of transcript abundance in pig peri-mortem skeletal muscle: eQTL localized genes involved in stress response, cell death, muscle disorders and metabolism. *BMC Genomics.* 2011;12:548.
53. Mortimer SI, Hatcher S, Fogarty NM, van der Werf JHJ, Brown DJ, Swan AA, et al. Genetic correlations between wool traits and meat quality traits in Merino sheep. *J Anim Sci.* 2017;95:4260–73.
54. van den Berg I, Hayes BJ, Chamberlain AJ, Goddard ME. Overlap between eQTL and QTL associated with production traits and fertility in dairy cattle. *BMC Genomics.* 2019;20:291.
55. Takata A, Matsumoto N, Kato T. Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nat Commun.* 2017;8:14519.
56. Li YI, van de Geijn B, Raj A, Knowles DA, Petti AA, Golan D, et al. RNA splicing is a primary link between genetic variation and disease. *Science.* 2016;352:600–4.
57. Bahcall OG. Human genetics: GTEx pilot quantifies eQTL variation across tissues and individuals. *Nat Rev Genet.* 2015;16:375.
58. Flutre T, Wen X, Pritchard J, Stephens M. A statistical framework for joint eQTL analysis in multiple tissues. *PLoS Genet.* 2013;9:e1003486.
59. Gonzalez-Prendes R, Marmol-Sanchez E, Quintanilla R, Castello A, Zidi A, Ramayo-Caldas Y, et al. About the existence of common determinants of gene expression in the porcine liver and skeletal muscle. *BMC Genomics.* 2019;20:518.

60. Powell JE, Henders AK, McRae AF, Wright MJ, Martin NG, Dermitzakis ET, et al. Genetic control of gene expression in whole blood and lymphoblastoid cell lines is largely independent. *Genome Res.* 2012;22:456–66.
61. Gaffney DJ. Global properties and functional complexity of human gene regulatory variation. *PLoS Genet.* 2013;9:e1003501.
62. Zhernakova DV, Deelen P, Vermaat M, van Iterson M, van Galen M, Arindrarto W, et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat Genet.* 2017;49:139–45.
63. Bolormaa S, Pryce JE, Reverter A, Zhang Y, Barendse W, Kemper K, et al. A multi-trait, meta-analysis for detecting pleiotropic polymorphisms for stature, fatness and reproduction in beef cattle. *PLoS Genet.* 2014;10:e1004198.
64. Xiang R, van den Berg I, MacLeod IM, Hayes BJ, Prowse-Wilkins CP, Wang M, et al. Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc Natl Acad Sci USA.* 2019;116:19398–408.
65. Natrass GS, Cafe LM, McIntyre BL, Gardner GE, McGilchrist P, Robinson DL, et al. A post-transcriptional mechanism regulates calpastatin expression in bovine skeletal muscle. *J Anim Sci.* 2014;92:443–55.
66. Kwan T, Benovoy D, Dias C, Gurd S, Serre D, Zuzan H, et al. Heritability of alternative splicing in the human genome. *Genome Res.* 2007;17:1210–8.
67. Guelfi S, D'Sa K, Botia J, Vandrovcova J, Reynolds RH, Zhang D, et al. Regulatory sites for splicing in human basal ganglia are enriched for disease-relevant information. *Nat Commun.* 2020;25:1041.
68. Ponnampalam EN, Butler KL, Jacob RH, Pethick DW, Ball AJ, Edwards JE, et al. Health beneficial long chain omega-3 fatty acid levels in Australian lamb managed under extensive finishing systems. *Meat Sci.* 2014;96:1104–10.
69. Simopoulos AP. Genetic variants in the metabolism of omega-6 and omega-3 fatty acids: their role in the determination of nutritional requirements and chronic disease risk. *Exp Biol Med (Maywood).* 2010;235:785–95.
70. Clark EL, Bush SJ, McCulloch MEB, Farquhar IL, Young R, Lefevre L, et al. A high resolution atlas of gene expression in the domestic sheep (*Ovis aries*). *PLoS Genet.* 2017;13:e1006997.
71. Schmiedt W, Kinscherf R, Deigner HP, Kamencic H, Nauen O, Kilo J, et al. Complement C6 deficiency protects against diet-induced atherosclerosis in rabbits. *Arterioscler Thromb Vasc Biol.* 1988;18:1790–5.
72. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet.* 2018;19:491–504.
73. Vösa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *bioRxiv.* 2018. <https://doi.org/10.1101/447367>.
74. Australian code of practice for the care and use of animals for scientific purposes. 7th ed. Canberra: Australian Government; 2004.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

