RESEARCH PAPER

Gene co-expression and alternative splicing analysis of key metabolic tissues to unravel the regulatory signatures of fatty acid composition in cattle

Hui-Zeng Sun^{a,b}, Zhi Zhu^{b,c}, Mi Zhou^b, Jian Wang^{b,d}, Michael E. R. Dugan^d, and Le Luo Guan^b

^aInstitute of Dairy Science, Ministry of Education Key Laboratory of Molecular Animal Nutrition, College of Animal Sciences, Zhejiang University, Hangzhou, China; ^bDepartment of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, Canada; ^cDepartment of Animal Science, Southwest University, Chongqing, P.R. China; ^dShaanxi Key Laboratory of Agricultural Molecular Biology, College of Animal Science and Technology, Northwest A&F University, Yangling, Shaanxi, China

ABSTRACT

Increasing the healthy/unhealthy fatty acid (FA) ratio in meat is one of the urgent tasks required to address consumer concerns. However, the regulatory mechanisms ultimately resulting in FA profiles vary among animals and remain largely unknown. In this study, using ~1.2 Tb high-quality RNA-Seq-based transcriptomic data of 188 samples from four key metabolic tissues (rumen, liver, muscle, and backfat) together with the contents of 49 FAs in backfat, the molecular regulatory mechanisms of these tissues contributing to FA formation in cattle were explored. Using this large dataset, the alternative splicing (AS) events, one of the transcriptional regulatory mechanisms in four tissues were identified. The highly conserved and absent AS events were detected in rumen tissue, which may contribute to its functional differences compared with the other three tissues. In addition, the healthy/unhealthy FA ratio related AS events, differentially expressed (DE) genes, co-expressed genes, and their functions in four tissues were analysed. Eight key genes were identified from the integrated analysis of DE, co-expressed, and AS genes between animals with high and low healthy/unhealthy FA ratios. This study provides an applicable pipeline for AS events based on comprehensive RNA-Seq analysis and improves our understanding of the regulatory mechanism of FAs in beef cattle.

KEYWORDS

Beef cattle; alternative splicing events; gene co-expression; four tissue transcriptomics; fatty acid composition

Introduction

Beef is an important fatty acid (FA) source in human diets and plays a critical role in human nutrition and health. Based on the health risks associated with consuming certain types of FA by humans, some beef FAs can be defined as healthy or unhealthy FAs [1]. Among which, the healthy FAs of t11-18:1, C18:3 n-3, c9,t11-18:2, and the unhealthy FAs of C14:0, C16:0, t10-18:1 are most widely studied [2]. Development of strategies to increase healthy FAs in beef, however, remains challenging due to large variations among animals in FA formation and a lack of comprehensive understanding of modes of action regulating FA synthesis at the molecular level.

In cattle, FA synthesis and metabolism in cattle are very complex and involve in many biological processes in multiple tissues [3]. However, most ruminant studies to date have only focused on single tissue/organ, which lacks a holistic view to identify potential interactive mechanisms across different tissues. Generally, rumen, liver, muscle, and backfat are the major tissues associated with hydrolysis, hydrogenation, synthesis, metabolism, and modification of FA in cattle, which jointly determine the proportion of different FAs in fat and muscle tissues [4]. Recently, RNA-Sequencing (RNA-Seq) based transcriptomics has been gradually applied to decipher gene expression patterns in backfat [5,6], muscle [7,8], and liver [9,10] tissues related to FA formation in

different cattle species and breeds. In addition, it has been shown that genes interact with each other and can jointly affect many biological processes [11], and alternative splicing (AS) regulates transcription processes in almost every aspect of eukaryotic biology, which enables cells to generate different protein isoforms from a single gene [12]. To date, gene co-expression analysis and alternative splicing (AS) events investigation based on substantial information in RNA-Seq data is very limited in bovine research, and it is unknown if AS is one of the major regulators for FA formation in cattle.

In this study, we performed comprehensive analysis of co-expressed genes and AS events in rumen, liver, muscle, and backfat tissues from 48 beef steers using the 188 RNA-Seq datasets (Fig. 1). We also measured the relative concentrations of 49 FAs in backfat tissue and used them to define the phenotype in this study. It has been reported that the FA composition in backfat can reflect that in intramuscular fat to some extent [13]. For example, the conjugated linoleic acid (one of the beneficial FAs) content in *longissimus lumborum* muscle was similar to that in backfat in Canadian beef [14], suggesting that profiles of some FAs are correlated in these two tissues. We analysed the FA profiles, differentially expressed (DE) genes, co-expressed genes, and AS events in four tissues from high and low healthy/unhealthy FA ratio animals (Fig. 1). The potential systematic FA synthesis/deposition regulatory mechanisms including key genes, transcription factors, AS genes, and biological

Figure 1. The workflow of comprehensive RNA-Seq analysis applied in this study.

functions were identified. The data generated here can be re-used to investigate more scientific questions regarding ruminant biology. The findings of this study provide feasible workflow (Fig. 1) of comprehensive RNA-Seq analysis for AS event identification and extend our understanding of regulatory mechanism of FA composition in beef cattle, which will benefit cattle breeding, meat science, human nutrition, and health in the future.

Results

Fatty acid profiling

Forty-nine FAs within nine classifications were identified in the backfat tissue (Table S1). Healthy FAs analysed in this study referred to t11-18:1, C18:3 n-3 and c9,t11-18:2, and unhealthy FAs included C14:0, C16:0, and t10-18:1. The concentration distribution of FA showed similar patterns among different animals, in which the c9-18:1 (~38.0%), C16:0 (~27.2%), C18:0

(~11.6%), c9-16:1 (~5.2%), and C14:0 (~3.8%) were the most abundant FAs and accounted for >80% of total FA in backfat tissue (Fig. 2A). Cis-monounsaturated FAs (MUFAs) and straight chain saturated FAs were the most abundant FA types in backfat tissue, which had relative concentrations of $49.01 \pm 3.65\%$ and $44.47 \pm 3.36\%$, respectively (Fig. 2A). The healthy and unhealthy FA profiles had a wide range among different animals. The proportion of t11-18:1 ranged from 0.169% to 1.027% (Coefficient of Variation (CV) = 42.04), C18:3 n-3 ranged from 0.059% to 0.197% (CV = 24.82), c9,t11-18:2 ranged from 0.009% to 0.552% (CV = 61.20), C14:0 ranged from 2.675% to 5.427% (CV = 16.83), C16:0 ranged from 21.320% to 31.561% (CV = 9.19), t10-18:1 ranged from 0.194% to 3.473% (CV = 51.60) (Fig. 2A). These led to a range of healthy/unhealthy FA ratios from 0.0089 to 0.0557.

The PCA analysis showed the profile of 49 FAs was not separated by breed (Fig. 2B). The healthy/unhealthy FA ratio was also not affected by breed (ANOVA $P = 0.11$). The relative concentrations of t11-18:1, C18:3 n-3, c9,t11-18:2 in backfat in

Figure 2. The relative concentration of 49 fatty acids in backfat tissue in 48 animals and their difference among three breeds and between different FA ratio groups. A. Interleaved box of the relative concentration of 49 FAs in backfat tissue in 48 beef steers. The mean $\pm$ standard deviation value of each FA was labelled. B. The principle component analysis plot of the relative concentration of 49 FAs in 48 animals. C. The relative concentrations of three healthy FAs (t11-18:1, C18:3 n-3, c9, t11-18:2) and three unhealthy FAs (C14:0, C16:0, t10-18:1) between high and low FA ratio groups. ** represents FDR < 0.01, *** represents FDR < 0.001. The light pink solid round, light blue solid square, and orange solid triangle in figs a and b represent Kinsella, Angus, and Charolais breeds, respectively. FA, fatty acid.

the high ratio group were all significantly higher than those in low ratio group (FDR < 0.001, Fig. 2C). For unhealthy FAs, the proportion of C16:0 and t10-18:1 in high ratio group were significantly lower than low ratio group (FDR < 0.001) while no significant difference of C14:0 proportion was observed between the low and high ratio groups (FDR = 0.059, Fig. 2C).

Different alternative splicing events across four tissues

The PCA plot showed that the expression profiles of splicing sites in different samples were clustered by tissue, and there was no separation among the three breeds (Fig. 3A). Based on our results, rumen tissue had the most significantly different expression profiles of splicing sites when compared with the other three tissues (Fig. 3A). We found 10,399, 10,003, 9,871, 10,012 AS genes in the rumen, liver, muscle, and backfat tissues, respectively (Table S2). Among which, 8,682 AS genes were shared in four tissues and 438 (4.23%), 275 (2.75%), 333 (3.37%), 137 (1.37%) tissue-specific AS genes were found in the rumen, liver, muscle, and backfat tissues, respectively (Table S2).

Using the cut-off of FDR < 0.05 and $|\Delta\text{PSI}| > 0.1$, 669 to 1,531 significantly different AS events were identified in pairwise comparisons among the four tissues (Fig. 3B). Among them, rumen tissue showed 669, 1,358, and 1,531 significantly different AS events when compared with liver, muscle, and backfat tissues, respectively (Fig. 3B). Further analysis identified 2 rumen highly conserved AS events (related to *SLK* and *CD44* genes) with PSI > 0.85 in rumen tissues and PSI < 0.1 in

the other two tissues (Fig. 3C). Another 7 AS events (related to *CLSTN1*, *EPB41*, *CD46*, and *TPM4* genes) were detected in the other two tissues, but absent in the rumen with PSI < 0.1 in rumen and PSI > 0.85 in the other two tissues (Fig. 3D). The detailed information of these 9 AS events were included in Table S3.

Healthy/unhealthy FA-ratio-related co-expressed genes, DE genes, AS events analysis

Yellow ($R = -0.34$, $P = 0.03$) and black ($R = -0.31$, $P = 0.05$) gene modules were significantly correlated with FA ratio in rumen tissue, which contained 915 and 308 co-expressed genes, respectively (Fig. 4A). In the liver, three gene modules were significantly correlated with FA ratio including lightgreen ($R = -0.28$, $P = 0.05$), pink ($R = -0.31$, $P = 0.03$) and saddlebrown ($R = 0.32$, $P = 0.03$) modules, which had 118, 248, and 37 co-expressed genes (Fig. 4A). In the muscle, two gene modules including pink ($R = 0.48$, $P = 0.001$) and cyan ($R = 0.5$, $P = 0.0005$) with 146 and 49 co-expressed genes were significantly positively correlated with FA ratio. Another two gene modules including salmon ($R = -0.32$, $P = 0.03$) and turquoise ($R = -0.37$, $P = 0.01$) modules were negatively correlated with the healthy/unhealthy FA ratio (Fig. 4A). Both of the gene modules in the backfat, magenta ($R = 0.48$, $P = 0.0006$) and red ($R = 0.32$, $P = 0.03$) with 240 and 468 genes, respectively, were positively related with the healthy/unhealthy FA ratio and c9,t11-18:2 (Fig. 4A). The volcano plot showed 153 (60

Figure 3. The splicing sites expression profiles and alternative splicing events analysis across rumen, liver, muscle, and backfat tissues.

A. The principal component analysis score map of the expression of splicing sites identified in four tissues. Each dot represents one tissue sample from a single animal. Rumen, liver, muscle, and backfat tissues were represented using green, red, blue, and brown colour respectively. Solid square, round, and triangle represent Kinsella, Angus, and Charolais breeds, respectively. Circles indicate significant clusters. B. The number of significantly different alternative splicing (AS) events between any two tissues. C-D. The venn diagram of highly conserved and absent AS events in rumen tissue. PSI, per cent spliced in.

up- and 93 down-regulated), 120 (56 up- and 64 down-regulated), 102 (71 up- and 31 down-regulated) and 138 (56 up- and 82 down-regulated) genes were significantly differentially expressed between high and low FA ratio groups in rumen, liver, muscle, and backfat tissues, respectively ($|\log_2FC| > 1$ & $P < 0.05$, Fig. 4B). A total of 3 (1 mutually exclusive exons (MXE) and 2 skipped exon (SE)), 4 (1 alternative 5' splicing sites (A5SS) and 3MXE), 5 (2 Alternative 3' splicing sites (A3SS), 1 A5SS, 1 SE, and 1 retained intron (RI)), and 7 (1 A3SS, 2 MXE, 3 SE, and 1 RI) predicted AS events were significant different ($FDR < 0.05$) between high and low FA ratio groups in the rumen, liver, muscle, and backfat tissues, respectively (Fig. 4C).

FA-ratio-related genes and function analysis, key genes identification, and transcription factor prediction

We found 1,341 FA-ratio-related genes in the rumen from the sum of DE genes ($n = 153$), module genes ($n = 1,223$), and AS genes ($n = 3$) after removing replicates. Similarly, 498, 4,450, 792 FA-ratio-related genes were found in liver, muscle, and backfat tissues, respectively, which included 120, 102, 138 DE genes, 403, 4,643, 708 module genes, as well as 4, 5, 7 AS genes in the three

tissues, respectively. The healthy/unhealthy FA-ratio-related genes in rumen, liver, muscle, and backfat tissues were involved in 942, 231, 1,424, 387 GO terms (Table S4), among which 330, 19, 511, 57 GO terms were significantly enriched ($FDR < 0.01$).

Sixty-six healthy/unhealthy FA-ratio-related genes were identified in at least three different tissues (Fig. 5A). Among them, eight genes were commonly characterized in all the four tissues (Fig. 5A). These eight shared key genes included CDK5 regulatory subunit associated protein 1 like 1 (*CDKAL1*), heparan sulphate proteoglycan 2 (*HSPG2*), replication timing regulatory factor 1 (*RIF1*), protoporphyrinogen oxidase (*PPOX*), dishevelled associated activator of morphogenesis 1 (*DAAM1*), ubiquitin specific peptidase 33 (*USP33*), zinc finger protein 782 (*ZNF782*), centrosomal protein 290 (*CEP290*).

We further identified nine TFs based on the eight key genes, which contained NKX2-1 ($NES = 6.14$), PRRX2 ($NES = 6.04$), FOXO1 ($NES = 5.94$), RBPJ ($NES = 5.56$), ZNF143 ($NES = 5.46$), TBX21 ($NES = 5.33$), HBP1 ($NES = 5.31$), GATA1 ($NES = 5.10$), FOXO4 ($NES = 5.07$) (Fig. 5B). The network of TFs and key genes interaction as well as key gene-key gene relationship were constructed (Fig. 5B). NKX2-1 showed the most regulatory functions and regulated five key genes (*HSPG2*, *USP33*, *CDKAL1*, *ZNF782*, and *DAAM1*) (Fig. 5B). *DAAM1* was the

Figure 4. Fatty acid ratio related co-expressed genes, differential expressed genes, and alternative splicing events in rumen, liver, muscle, and backfat tissues. A. The healthy/unhealthy fatty acid (FA) ratio significantly correlated co-expression gene modules in four tissues. The number of genes in each module was also indicated. B. The volcano plot of differential expressed genes between high and low FA ratio groups in four tissues. C. The number of different types of predicted alternative splicing events in high and low FA ratio groups in four tissues.

only gene regulated by all nine TFs (Fig. 5B). *CEP20* highly correlated with *CDKAL1* ($R = 0.70$), *RIF1* ($R = 0.89$), *DAAM1* ($R = 0.80$), *USP33* ($R = 0.61$), and *ZNF782* ($R = 0.83$).

Discussion

In our previous paper, the comparative gene expression profiles among rumen, liver, muscle, and backfat tissues and their roles in cattle feed efficiency were presented [15]. This study

continued to further analyse the splicing sites and AS events across the four tissues. In addition, using the massive RNA-Seq dataset in cattle, we systematically investigated the potential regulatory signature of healthy/unhealthy FA ratios through analysing FA profiles, DE genes, co-expression genes, AS events, transcription factors, and biological functions across four tissues. The results considered the transcription changes in different aspects and in multiple tissues, which will provide a novel and

Figure 5. The key genes related to healthy/unhealthy fatty acid ratio in rumen, liver, muscle, and backfat tissues. A. The venn diagram of genes identified from fatty acid (FA)-ratio-related differential expressed genes, co-expressed genes and alternatively spliced genes in four tissues. B. The network of eight key genes and predicted transcription factors (TFs). NES, normalized enrichment score.

comprehensive view to help understand the healthy/unhealthy FA ratio and its related mechanism.

AS is the main reason for transcript diversity in eukaryotic genomes, which leads to large complexity in encoded protein structure and function [12]. Our results showed ~33.6% of genes at tissue level (32.9–34.7%) were alternatively spliced in cattle, which were greater than the previous reported 21% [16] or 26% [17] AS rates in bovine genome-wide analyses. However, the tissue specificity was not considered in previous studies. Our results showed among all the alternatively spliced genes (PSI > 0.1), ~86% were shared in four tissues and 2 ~ 4% were tissue-specific, suggesting that tissue difference should be paid attention to in certain AS events analysis. The AS complexity across different tissues has been assessed in humans [12,18], mice, chickens, frogs, and lizards [19], in which reported that tissue type was the primary source of variability underlying AS patterns. A recent study compared the AS patterns between five high-altitude vertebrates (Tibetan chicken, pig, sheep, goat, and yak) and their low-altitude relatives (low-altitude chicken, pig, sheep, goat, and cattle) showed that species, followed by tissues dominated the complexity of AS patterns, both of which contributed more variance than altitude [20]. In our study, tissues accounted for the largest variance on AS patterns while breeds did not contribute to such variation. This is reasonable since the three different breeds (Angus, Charolais, and Kinsella) included in this study were all meat-producing beef cattle breeds and raised under the same diet and management conditions, which should have fewer variations in AS than when comparing different ruminant species (e.g., yak vs cattle). This also supported our down-stream analysis of different AS events between high- and low-healthy/unhealthy FA ratio groups using animal population with combined breeds.

Even though the AS patterns in bovine subcutaneous fat [4,21,22], testis [23], mammary gland [24], muscle [25], liver [26] and blood [27] have been investigated using global transcriptomics, the exploration of AS patterns of rumen tissue has not been performed to date. Interestingly, rumen tissue had the most different splicing site expression profiles when compared to liver, muscle, and backfat tissues. In our previous work, we found that the gene expression profile of muscle tissue was most distinctive from others, while those of rumen and backfat tissues were similar [15]. Splicing sites are the intron-exon junctions in the precursor mRNA of eukaryotes, which can be recognized by spliceosome in the splicing process, affecting the diversity and efficiency of AS events and ultimately altering protein types [28]. Our results suggest that the functional uniqueness of rumen tissue may be attributed to the protein diversity rather than gene expression levels. More importantly, the highly conserved and absent AS events were identified in rumen tissue, which could generate different protein isoforms that play different roles between rumen and other tissues. It is hypothesized that the AS of *SLK* and *CD44* may give rise to multiple protein isoforms in rumen, liver, muscle, and backfat tissues in cattle, in which liver, muscle, and backfat tissues shared at least one isoform, and rumen is predominant with a most distinctive isoform. For example, the gene *CD44* has nine different splice variants in cattle [29]. The highly conserved AS in rumen was the skipped exon splicing type, which started and ended skip at 65,718,576 and 65,718,771 bp, respectively (Table S3). The *CD44* plays

essential roles in cellular fundamental processes (e.g., cell adhesion, migration, and cell–cell interaction) and is expressed in a wide variety of cell types [30]. Among them, the large *CD44E* isoform is expressed in some epithelium cells and shows distinct adhesion potentials [31]. It is reported that *CD44* splice isoforms play vital roles in epithelial-mesenchymal transition, which has been proven to be a major epithelial variant in mice and humans [32]. The identified highly conserved skipped exon splicing type of *CD44* suggests that in cattle, this protein isoform may affect cell adhesion, tissue tension, and tight junctions, which partially explains the unique epithelial structure of the rumen and its functionality of nutrient absorption and interactions with the rumen microbiome. In addition, the unique and absent AS events provide important markers for rumen development and function.

The previous FA-related AS events analyses in cattle were only performed in the backfat tissue [4,21,22]. However, FA synthesis and metabolism in cattle are affected by many complex biological processes in multiple tissues [9]. Using muscle transcriptomics, *HSPG2* was found to be more highly expressed in early stages of muscle development and FA deposition in Qinchuan cattle [33]. *HSPG2* was also down-regulated in the muscle in Chinese Simmental cattle with higher C18:3 n-3 and lower C14:0 FA [5]. This suggests that *HSPG2* may negatively regulate FA deposition and the healthy/unhealthy FA ratio in different cattle populations. *CDKAL1* and *CEP290* were expressed more highly in backfat tissue in cattle during feed restriction and involved in the process of FA mobilization and degradation [34]. In addition, *CDKAL1* and *HSPG2* were reported as co-expressed genes in the intramuscular fat correlated gene module in Nelore cattle [6], which showed similar results in the healthy/unhealthy FA-ratio-related gene modules in our study. When comparing the backfat transcriptome between high and low FA content cattle, the AS events in *CDKAL1*, *HSPG2*, *RIF1*, *PPOX*, *USP33*, and *CEP290* genes were detected [21], which were consistent with our results. The above limited findings in cattle research partially supported the function of these six key genes (from eight driver genes identified in this study) in FA synthesis and metabolism.

Regarding functional aspects, GO terms enrichment analysis identified terms consistent with the known physiology of FA metabolism in four tissues, especially in liver. We found half of enriched GO terms (9/19) in liver were involved in lipid metabolic processes, which indicates liver function largely contributes to variation in cattle FA. The TFs predicted based on eight key genes in this study were also proven to be involved in the regulation of FA metabolism in cattle. For example, *PRRX2* was associated with the proliferation and differentiation of adipocytes in cattle [35]. It has also been reported that *PRRX2* is significantly correlated with beef flavour, explains 67% of its variability [36], and can be attributed to FA composition and lipid metabolic process (e.g., lipid oxidation). Many studies showed *FOXO1* plays a vital role during adipocyte development in Lulu cattle [37], Qinchuan cattle [38], and Nanyang cattle [39,40]. *FOXO1* was also identified as an important TF related to the healthy/unhealthy FA ratio in this study, which suggests that *FOXO1* may act as a negative regulator of transcription in preadipocyte differentiation and then affects FA profiles. Our study was able to identify the gene function and a couple of

candidate regulators related to the healthy/unhealthy FA ratio using four tissue transcriptomes in cattle. These findings provide some novel insights into the molecular mechanisms of FA metabolism and deposition in cattle that can be incorporated in future research.

In summary, this paper has provided new biological insights derived from co-expression and AS analysis of the dataset. The data serves as one of the most comprehensive transcriptomic related datasets available for cattle studies and can be easily obtained from public database for further analysis, which provides valuable sharable phenotypic and sequence data for future animal nutrition, meat science, and bioinformatics research. Using the dataset, we found rumen tissue presented the most different splicing site expression profiles with several unique and absent AS events. The healthy/unhealthy FA ratios were related to eight key genes identified through integrating the co-expressed, DE, and AS genes from the four tissues. These findings not only provided a new understanding of the regulatory processes underlying ruminant adipogenesis, but also contribute important information for cattle breeding and human health-related nutrition.

Materials and methods

Animals, sample collection, and analysis workflow

All animal experimentation was in compliance with the Canadian Council of Animal Care guidelines (1993) [41] and approved by the University of Alberta Livestock Animal Care and Use Committee (Protocol No.: AUP00000882). The workflow of this study is described in Fig. 1. One hundred and forty-three beef steers with different feed efficiency (residual feed intake range from -1.67 to 2.17) were selected from a large cohort consisting of three different breeds (Angus, Charolais, and Kinsella) raised under the same diet and management conditions at the Roy Berg Kinsella Ranch, University of Alberta (Alberta, Canada). Rumen epithelium, liver, muscle, and backfat samples were collected after slaughter and snap-frozen in liquid nitrogen within 30 minutes. All the tissue samples were subsequently stored at -80°C until FA measurement and RNA extraction.

Fatty acid measurement

Fifty mg backfat were weighed, freeze-dried and methylated with sodium methoxide [14]. One ml of 1 mg c10-17:1 methyl ester/ml hexane (Nu-Check Prep Inc., Elysian, MN, USA) was used as an internal standard and added to samples before methylation [42]. Fatty acid methyl esters (FAMES) were analysed using a CP-3800 gas chromatograph (Varian Inc., Walnut Creek, CA, USA). Most FAMES were analysed as described by Kramer et al. [43] using a CP-Sil88 column (100 m, 25 μm ID, 0.2 μm film thickness, Agilent Technologies, Santa Clara, CA), except t7,c9-18:2 and c9,t11-18:2, which were analysed according to Turner et al. [44] using an SLB IL 111 column (30 m, 0.25 mm ID, 0.2 mm film thickness, Supelco Inc., Bellefonte, PA). The concentrations of FA were expressed as a percentage of FAME quantified. The healthy/unhealthy FA ratio was calculated as the

total proportion of healthy FA/total proportion of unhealthy FA. The cut-off for high and low FA ratio were defined as >0.03 or <0.03 , respectively. The 0.03 was the median value of FA ratio in 143 animals (Table S5). Statistical significance of FA between each two breeds was defined as $\text{FDR} < 0.05$. More details can be found in Supplementary Methods.

RNA sequencing data generation and processing

Forty-eight beef steers were then selected ($n = 16$ in each breed) based on backfat fatty acid profiles from 143 animals. For all the 48 animals, we sequenced 188 tissue samples using the Illumina HiSeq 4000 system (Illumina, San Diego, CA, USA) (3 muscle and 1 rumen samples failed at library construction and sequencing steps, respectively). The procedures for RNA sequencing data generation and processing were modified based on our previous reports [15]. Briefly, the major steps include: a) extracting total RNA using a mirVana total RNA Isolation Kit (Ambion, Carlsbad, CA, USA) for rumen, liver, and muscle tissue samples and a RNeasy Lipid Tissue Mini Kit (Qiagen, Hilden, Germany) for backfat tissues; b) constructing cDNA library using a Truseq Stranded Total RNA Sample Preparation kit (Illumina, San Diego, CA, USA); c) 100 bp paired-end sequencing using the Illumina HiSeq 4000 system (Illumina, San Diego, CA, USA); d) trimming sequencing reads using trimmomatic tool (version 0.35) [45] with the cut-off of read length no less than 75 and quality score greater than 25; e) aligning clean reads to the bovine genome (ARS-UCD1.2, release 100) using the HISAT2 (version 2.1.0) [46].

The mapped reads aligned to each bovine gene were counted using HTSeq-count tool (version 0.9.0) [47] and were further normalized into transcripts per kilobase million (TPM). The expressed genes were defined as genes with TPM value greater than 1 [48]. Differentially expressed genes between high and low FA ratio groups were identified using DESeq2 [49] with the cut-off $\text{FDR} < 0.05$.

Splicing sites expression and alternative splicing events

The splice-site coordinates of each junction were analysed through matching the Ensembl and UCSC bovine mRNA annotated exons and introns, individual exon splicing sites, and exon-junctions using rMATs software (version 4.1.0) [50]. Five types of different alternative splicing events were generated: SE (Skipped exon), A5SS (Alternative 5' splicing sites), A3SS (Alternative 3' splicing sites), MXE (Mutually exclusive exons), and RI (Retained intron). The expression profiles of splicing sites among different tissues were displayed in principal component analysis (PCA) plot of FA profiles using the 'ggplot2' and 'factoextra' packages in R software (version 3.5.2).

The percent spliced in (PSI) value was introduced to indicate the possibility of one AS event occurred (range from 0 to 1, higher value represent stronger intensity), which is calculated as the ratio between reads including and excluding exons [51]. $\text{dPSI} = \text{PSI mean value in one treatment} - \text{PSI mean value in another treatment}$. The significantly different AS events were analysed using student t-test in R software [52], statistical

significance was defined as $FDR < 0.05$ and $|dPSI| > 0.1$. Tissue-specific AS events were defined as one AS event occurred in one tissue ($PSI > 0.1$) but not occurred in other tissues ($PSI < 0.1$). We featured the highly conserved AS events in certain tissue as PSI value > 0.85 in that tissue and < 0.1 in at least two tissues. Similarly, the highly conserved absent AS events were characterized as PSI value < 0.1 in that tissue and > 0.85 in at least the other two tissues. AS genes were defined as genes contain at least one AS event with PSI values > 0.1 in more than 50% of samples. The AS genes identified in only one tissue were regarded as tissue-specific AS genes.

Weighted gene co-expression network analysis (WGCNA)

The co-expressed gene modules were identified based on all expressed genes, and further correlated with FA ratio, healthy FAs (t11-18:1, C18:3 n-3, c9,t11-18:2) and unhealthy FAs (C14:0, C16:0, t10-18:1) traits using WGCNA package in R software. WGCNA analysis was conducted to identify module-traits relationships using 47, 48, 45, and 48 samples in rumen, liver, muscle, and backfat tissues, respectively. The co-expressed gene module-traits relationship was analysed separately for each tissue. Gene modules consisting of a different number of genes were displayed with different colours. The module-trait relationship with P value < 0.05 & $|R| > 0.25$ was regarded as significant. To ensure achieving the optimal scale-free topology property and R^2 value of power-law module > 0.9 , the parameters of 'estimated soft-thresholding', 'signed topological overlap matrix type', 'minimum module size = 30', 'merged cut height = 0.25' were set in this study. The significantly correlated modules in each tissue were extracted and combined into one figure.

Healthy/unhealthy FA-ratio-related genes function analysis, key genes identification, and transcription factor prediction

Healthy/unhealthy FA-ratio-related genes in each tissue were obtained from the sum of DE genes between high and low FA ratio groups, the genes in FA ratio significantly correlated modules from WGCNA analysis, and AS genes between high and low FA ratio groups. The duplicates among DE genes, module genes, and AS genes were removed in this step. Functional analysis of the FA-ratio-related genes in each tissue was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID, <https://david.ncifcrf.gov>). The cut-off of significant enriched gene ontology (GO) terms was $FDR < 0.01$.

The overlapped FA-ratio-related genes among four tissues were identified as key genes and further used to predict transcriptional factors (TFs) using iRegulon tool (version 1.3). The putative regulatory region was selected as 20 kb centred around transcription start site [53]. Only TFs with normalized enrichment score (NES) > 5 were displayed in the network to show the regulatory relationship with key genes using Cytoscape (version 3.7.1). The co-expressed gene-gene associations of the key genes in four tissues were also shown in the network. For the co-expression association, only the edges in the network with correlation coefficient $R > 0.5$ were displayed.

Acknowledgments

We thank Dr. C. Fitzsimmons, Mr. B. Irving and staff at Roy Berg Kinsella Ranch for animal trial. We appreciate our lab members in the Department of Agricultural, Food and Nutritional Science, University of Alberta, for their help on sample collection.

Funding

This work was supported by Alberta Livestock and Meat Agency Ltd. (No. 2015P008R), Ministry of Alberta Agriculture and Forestry (No. AF2018F095R) and Canada NSERC Discovery grant.

Authors' contributions

HS conducted bioinformatics analysis, submitted sequencing data to the public database and drafted the manuscript. ZZ and JW participated in the sequence alignment and generated gene raw counts. MZ participated in the design of the study. MERD helped in fatty acids measurement and assisted in editing manuscript. LLG conceived of the study, participated in its design and coordination and helped to revise the manuscript. All authors read and approved the final manuscript.

Data availability

RNA-Seq raw sequencing data (.fastq), processing data and phenotypic data has been submitted to the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) with the accession number GSE116775. The data analysis code has been submitted to the Github (<https://github.com/huizeng001/Cattleomics>).

Disclosure statement

The authors have declared no competing interests.

ORCID

Hui-Zeng Sun <http://orcid.org/0000-0001-5380-6030>

References

- [1] Calder PC. Functional roles of fatty acids and their effects on human health. *J Parenter Enteral Nutr.* **2015**;39:18S–32S.
- [2] Dewhurst RJ, Scollan ND, Lee MR, et al. Forage breeding and management to increase the beneficial fatty acid content of ruminant products. *P Nutr Soc.* **2003**;62(2):329–336.
- [3] Shingfield KJ, Bonnet M, Scollan ND. Recent developments in altering the fatty acid composition of ruminant-derived foods. *Animal.* **2013**;7(s1):132–162.
- [4] Brzozowska AM, Oprządek J. Metabolism of fatty acids in tissues and organs of the ruminants – a review. *Anim Sci Pap Rep.* **2016**;34:3.
- [5] Wærp HK, Waters SM, McCabe MS, et al. RNA-seq analysis of bovine adipose tissue in heifers fed diets differing in energy and protein content. *PloS One.* **2018**;13(9):e0201284.
- [6] Huang W, Guo Y, Du W, et al. Global transcriptome analysis identifies differentially expressed genes related to lipid metabolism in Wagyu and Holstein cattle. *Sci Rep.* **2017**;7(1):5278.
- [7] Zhang H, Xia H, Jiang H, et al. Longissimus dorsi muscle transcriptomic analysis of Yunling and Chinese simmental cattle differing in intramuscular fat content and fatty acid composition. *Genome.* **2018**;61(8):549–558.
- [8] Oliveira GB, Regitano LC, Cesar AS, et al. Integrative analysis of microRNAs and mRNAs revealed regulation of composition and metabolism in Nelore cattle. *BMC Genomics.* **2018**;19(1):126.

- [9] Wang X, Zhang Y, Zhang X, *et al.* The comprehensive liver transcriptome of two cattle breeds with different intramuscular fat content. *Biochem Biophys Res Commun.* **2017**;490(3):1018–1025.
- [10] Liang R, Han B, Li Q, *et al.* Using RNA sequencing to identify putative competing endogenous RNAs (ceRNAs) potentially regulating fat metabolism in bovine liver. *Sci Rep.* **2017**;7(1):6396.
- [11] Miklos GLG, Rubin GM. The role of the genome project in determining gene function: insights from model organisms. *Cell.* **1996**;86(4):521–529.
- [12] Pan Q, Shai O, Lee LJ, *et al.* Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nature Genet.* **2008**;40(12):1413.
- [13] Abe T, Saburi J, Hasebe H, *et al.* Novel mutations of the FASN gene and their effect on fatty acid composition in Japanese Black beef. *Biochem Genet.* **2009**;47:397–411.
- [14] Aldai N, Dugan ME, Rolland D, *et al.* Survey of the fatty acid composition of Canadian beef: backfat and longissimus lumborum muscle. *Can J Anim Sci.* **2009**;89(3):315–329.
- [15] Sun H-Z, Zhao K, Zhou M, *et al.* Landscape of multi-tissue global gene expression reveals the regulatory signatures of feed efficiency in beef cattle. *Bioinformatics.* **2019**;35(10):1712–1719.
- [16] Chacko E, Ranganathan S. Genome-wide analysis of alternative splicing in cow: implications in bovine as a model for human diseases. In: *BMC genomics*, **2009**. S11.
- [17] Kim N, Alekseyenko AV, Roy M, *et al.* The ASAP II database: analysis and comparative genomics of alternative splicing in 15 animal species. *Nucleic Acids Res.* **2006**;35(suppl_1):D93–D98.
- [18] Wang ET, Sandberg R, Luo S, *et al.* Alternative isoform regulation in human tissue transcriptomes. *Nature.* **2008**;456(7221):470.
- [19] Barbosa-Morais NL, Irimia M, Pan Q, *et al.* The evolutionary landscape of alternative splicing in vertebrate species. *Science.* **2012**;338(6114):1587–1593.
- [20] Tang Q, Gu Y, Zhou X, *et al.* Comparative transcriptomics of 5 high-altitude vertebrates and their low-altitude relatives. *GigaScience.* **2017**;6(12):gix105.
- [21] Mei C, Li S, Abbas SH, *et al.* Performance measurement and comparative transcriptome analysis revealed the efforts on hybrid improvement of Qinchuan cattle. *Anim Biotechnol.* **2019**;30(1):13–20.
- [22] Cai H, Li M, Sun X, *et al.* Global transcriptome analysis during adipogenic differentiation and involvement of transthyretin gene in adipogenesis in cattle. *Front Genet.* **2018**;9.
- [23] Mipam TD, Ai Y, Chai Z, *et al.* Comparative transcriptome analysis revealed lower genetic variations of genes in cattleyak testis. *Anim Sci Pap Rep.* **2019**;37:1.
- [24] Wang XG, Ju ZH, Hou MH, *et al.* Deciphering transcriptome and complex alternative splicing transcripts in mammary gland tissues from cows naturally infected with *Staphylococcus aureus* mastitis. *PLoS One.* **2016**;11(7):e0159719.
- [25] Sun X, Li M, Sun Y, *et al.* The developmental transcriptome landscape of bovine skeletal muscle defined by Ribo-Zero ribonucleic acid sequencing. *J Anim Sci.* **2015**;93(12):5648–5658.
- [26] Pegolo S, Cecchinato A, Mach N, *et al.* Transcriptomic changes in liver of young bulls caused by diets low in mineral and protein contents and supplemented with n-3 fatty acids and conjugated linoleic acid. *PLoS One.* **2016**;11(12):e0167747.
- [27] Bai X, Zheng Z, Liu B, *et al.* Whole blood transcriptional profiling comparison between different milk yield of Chinese Holstein cows using RNA-seq data. *BMC Genomics.* **2016**;17(7):512.
- [28] Amit M, Donyo M, Hollander D, *et al.* Differential GC content between exons and introns establishes distinct strategies of splice-site recognition. *Cell Rep.* **2012**;1(5):543–556.
- [29] Shamimuzzaman M, Tourneau JJJ, Unni DR, *et al.* Bovine genome database: new annotation tools for a new reference genome. *Nucleic Acids Res.* **2020**;48(1):676–681.
- [30] Goodison S, Urquidí V, Tarin D. CD44 cell adhesion molecules. *Mol Pathol.* **1999**;52(4):189–196.
- [31] Stamenkovic I, Aruffo A, Seed B. The hematopoietic and epithelial forms of CD44 are distinct polypeptides with different adhesion potentials for hyaluronate-bearing cells. *Embo J.* **1991**;10:343–348.
- [32] Brown RL, Reinke LM, Damerow MS, *et al.* CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. *J Clin Invest.* **2011**;121(3):1064–1074.
- [33] He H, Chen S, Liang W, *et al.* Genome-wide proteomics analysis on longissimus muscles in Qinchuan beef cattle. *Anim Genet.* **2017**;48(2):131–140.
- [34] Cunningham HC, Cammack KM, Hales KE, *et al.* Differential transcript abundance in adipose tissue of mature beef cows during feed restriction and realimentation. *PLoS One.* **2018**;13(3):e0194104.
- [35] Wei S, Duarte MS, Zan L, *et al.* Cellular and molecular implications of mature adipocyte dedifferentiation. *J Genomics.* **2013**;1:5.
- [36] Bernard C, Cassar-Malek I, Le Cunff M, *et al.* New indicators of beef sensory quality revealed by expression of specific genes. *J Agri Food Chem.* **2007**;55(13):5229–5237.
- [37] Liu X, Liu G, Tan X, *et al.* Gene expression profiling of SIRT1, FoxO1, and PPAR γ in backfat tissues and subcutaneous adipocytes of Lulu bulls. *Meat Sci.* **2014**;96(2):704–711.
- [38] Sun Y, Xue J, Guo W, *et al.* Haplotypes of bovine FoxO1 gene sequence variants and association with growth traits in Qinchuan cattle. *J Genet.* **2016**;93(2):8–14.
- [39] Sun X, Li M, Hao D, *et al.* Two novel polymorphisms of bovine SIRT2 gene are associated with higher body weight in Nanyang cattle. *Mol Biol Rep.* **2015**;42(3):729–736.
- [40] Li M, Sun X, Hua L, *et al.* SIRT1 gene polymorphisms are associated with growth traits in Nanyang cattle. *Mol Cell Prob.* **2013**;27(5–6):215–220.
- [41] Olfert ED, Cross BM, McWilliam AA. Guide to the care and use of experimental animals. Vol. 1. Ottawa: Canadian Council on Animal Care; **1993**.
- [42] Vahmani P, Aalhus J, Rolland D, *et al.* Sequential feeding of lipid supplement enriches beef adipose tissues with 18: 3n-3 biohydrogenation intermediates. *Lipids.* **2017**;52(7):641–649.
- [43] Kramer JKG, Hernandez M, Cruz-Hernandez C, *et al.* Combining results of two GC separations partly achieves determination of all cis and trans 16:1, 18:1, 18:2 and 18:3 except CLA isomers of milk fat as demonstrated using ag-ion SPE fractionation. *Lipids.* **2008**;43(3):259–273.
- [44] Turner T, Rolland D, Aldai Elkoru-Irribé N, *et al.* Short communication: rapid separation of cis9, trans11- and trans7, cis9-18: 2 (CLA) isomers from ruminant tissue using a 30 m SLB-IL111 ionic column. **2011**.
- [45] Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics.* **2014**;30(15):2114–2120.
- [46] Kim D, Langmead B, Salzberg S. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods.* **2015**;12:57–360.
- [47] Anders S, Pyl PT, Huber W. HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics.* **2015**;31(2):166–169.
- [48] Mortazavi A, Williams BA, McCue K, *et al.* Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods.* **2008**;5(7):621.
- [49] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**;15:550.
- [50] Shen S, Park JW, Lu Z, *et al.* rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci USA.* **2014**;111(51):5593–5601.
- [51] Schafer S, Miao K, Benson CC, *et al.* Alternative splicing signatures in RNA-seq data: percent spliced in (PSI). *Curr Protoc Hum Genet.* **2015**;87(1):11–16.
- [52] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* **2008**;9(1):559.
- [53] Verfaillie A, Imrichová H, Van de Sande B, *et al.* iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol.* **2014**;10(7):e1003731.