Identification of single nucleotide polymorphisms (SNPs) potentially associated with residual feed intake in Qinchuan beef cattle by hypothalamus and duodenum RNA-Seq data

Zonghua Su[1], Chenglong Li[1], Chaoyun Yang[1], YanLing Ding[1], Xiaonan Zhou[1], Junjie Xu[1], Chang Qu[1], Yuangang Shi[1,†], Cong-Jun Li[2] and Xiaolong Kang[1]

[1] Key Laboratory of Ruminant Molecular and Cellular Breeding, College of Animal Science and Technology, Ningxia University, Yinchuan, China
[2] Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service, United States Department of Agricultural, Beltsville, MD, United States
[†] Deceased author.

ABSTRACT

The regulation of residual feed intake (RFI) in beef cattle involves brain-gut mechanisms due to the interaction between neural signals in the brain and hunger or satiety in the gut. RNA-Seq data contain an extensive resource of untapped SNPs. Therefore, hypothalamic and duodenal tissues from ten extreme RFI individuals were collected, and transcriptome sequenced in this study. All the alignment data were combined according to RFI, and the SNPs in the same group were identified. A total of 270,410 SNPs were found in the high RFI group, and 255,120 SNPs were found in the low RFI group. Most SNPs were detected in the intronic region, followed by the intergenic region, and the exon region accounts for 1.11% and 1.38% in the high and low RFI groups, respectively. Prediction of high-impact SNPs and annotation of the genes in which they are located yielded 83 and 97 genes in the high-RFI and low-RFI groups, respectively. GO enrichment analysis of these genes revealed multiple NADH/NADPH-related pathways, with *ND4*, *ND5*, and *ND6* significantly enriched as core subunits of NADH dehydrogenase (complex I), and is closely related to mitochondrial function. KEGG enrichment analysis of *ND4*, *ND5*, and *ND6* genes was enriched in the thermogenic pathway. Multiple genes, such as *ATP1A2*, *SLC9A4*, and *PLA2G5*, were reported to be associated with RFI energy metabolism in the concurrent enrichment analysis. Protein-protein interaction analysis identified multiple potential candidate genes related to energy metabolism that were hypothesized to be potentially associated with the RFI phenotype. The results of this study will help to increase our understanding of identifying SNPs with significant genetic effects and their potential biological functions.

Subjects Agricultural Science, Genetics, Genomics, Zoology
Keywords RNA-Seq, Single nucleotide polymorphisms, Beef cattle, Hypothalamus, Duodenum

Submitted 13 December 2024
Accepted 14 March 2025
Published 15 April 2025

Corresponding authors
Cong-Jun Li, congjun.li@usda.gov
Xiaolong Kang, kangxl9527@126.com

Academic editor
Jonathan Thomas

Additional Information and Declarations can be found on page 18

DOI **10.7717/peerj.19270**

Copyright
2025 Su et al.

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

How to cite this article Su Z, Li C, Yang C, Ding Y, Zhou X, Xu J, Qu C, Shi Y, Li C, Kang X. 2025. Identification of single nucleotide polymorphisms (SNPs) potentially associated with residual feed intake in Qinchuan beef cattle by hypothalamus and duodenum RNA-Seq data. **PeerJ 13:e19270 DOI 10.7717/peerj.19270**

INTRODUCTION

Feed cost accounts more than 70% of the total input cost in cattle production, making feed utilization a crucial metric for evaluating production expenditures (*Patience, Rossoni-Serao & Gutierrez, 2015*). Efficient feed utilization can reduce herd maintenance costs by 9–10%, lower feed intake by 10–12%, and decrease methane emissions by 15–20% (*Moore, Mujibi & Sherman, 2009*). Thus, optimizing feed utilization and reducing production costs are essential for livestock development. Residual feed intake (RFI) represents the disparity between the average daily feed intake (ADFI) and the average expected feed intake (AEFI) needed to maintain production levels (*Koch et al., 1963*). RFI provides a precise measure of feed utilization efficiency in livestock, isolating the effects of animal growth traits and rates (*Richardson & Herd, 2004*).

RFI is a promising candidate for genetic improvement due to its moderate heritability (0.28–0.58) (*Moore, Mujibi & Sherman, 2009*) and significant genetic variability (*Archer & Bergh, 2000*; *Herd & Bishop, 2000*). Our research found that RFI is related to gut microbiota (*Zhou et al., 2023*), circRNA-miRNA interaction (*Zhao et al., 2023*), and gene expression (*Yang et al., 2023*, *2021*, *2022*). These studies support a comprehensive analysis of RFI and show that its influencing factors are numerous and complex.

The hypothalamus and duodenum are critical organs in animal feed intake, energy metabolism, and digestion. The hypothalamic arcuate nucleus regulates appetite, where neuropeptide Y (NPY) and agouti-related peptide (AGRP) promote feeding. Conversely, α-MSH (α-melanocyte-stimulating hormone) induces satiety (*Perkins et al., 2014*). The duodenum, a key organ for nutrient absorption, facilitates various metabolic functions such as glucose, fat, vitamin B, calcium, zinc, and iron (*Anand et al., 2021*; *Cooke & Clark, 1976*; *Reeves & Chaney, 2004*). The interplay between the central nervous and digestive systems is evident in the microbiota-gut-brain axis (MGBA). The nervous system influences gut function through neurotransmitters and gut hormones, while gut microbes play crucial roles in host nutrient metabolism (*The 1000 Genomes Project Consortium, 2015*; *Olivier, 2003*). Thus, the close association of the hypothalamus and duodenum with feeding efficiency underscores their importance in studies investigating RFI in beef cattle.

Single nucleotide polymorphism (SNP) is a genetic variation resulting from a single nucleotide change in the DNA sequence. These variants can influence phenotypes and disease susceptibilities (*Kim & Misra, 2007*). Due to the low cost and high availability of RNA-seq data, coding region variants from RNA-Seq data are widely studied for their potential contribution to phenotype (*Karczewski et al., 2020*). Transcriptome data offer gene expression levels that can be utilized to investigate cis-regulation based on the expression of genes with SNP sites (*Jehl et al., 2021*). A wealth of research has been dedicated to extracting SNPs from transcriptomic data, yielding significant advancements across various fields. For instance, transcriptome sequencing of cow's milk has facilitated the discovery of SNPs, providing a robust foundation for marker-trait association studies (*Canovas et al., 2010*). In aquaculture, RNA-seq analysis has identified SNPs potentially linked to the immune response and the growth performance of *Penaeus vannamei* (*Santos, Andrade & Freitas, 2018*). In crop science, the development of genome-wide SNP markers

for barley has been achieved through reference-based RNA-Seq analysis (*Tanaka et al., 2019*). In animal husbandry, RNA-Seq SNP data has revealed potential causal mutations relevant to pig production traits and the intricacies of RNA editing (*Martinez-Montes et al., 2017*). SNP analysis of transcriptomic data from 20 human and bovine tissues revealed that cis-regulatory elements of gene expression are conserved between humans and cattle (*Yao et al., 2022*). Differential expression of an intronic SNP in *FABP4* was found to correlate with lipid transport and intracellular homeostatic regulation in studies of bovine rumen acidosis (*Zhao et al., 2017*).

This study characterized the SNPs from the hypothalamus and duodenum tissues of the same cattle with high and low residual feed intake based on RNA-seq data. The objective was to identify SNPs related to beef RFI and conduct subsequent bioinformatics analysis to detect the functional SNPs/genes associated with feed utilization performance in beef cattle and expand our understanding of the role of genetic variants in RFI phenotypes from expressed regions of the genome.

MATERIALS AND METHODS

Experimental animals and data collection

Based on our previous study (*Yang et al., 2021*), 30 Qinchuan bulls with similar age (15 ± 1 months) and weight (280.6 ± 30.9 kg) were selected from a farm in Ningxia, China. The study subjects were given a standardized feeding regimen throughout the experimental period, and free access to water and food was ensured. Body weight measurements were taken monthly throughout the 81-day experimental period, then daily feed intake, average daily gain (ADG), and the midpoint metabolic body weight (MMBW0.75) was calculated based on feed intake (FI). To classify the cattle into high and low RFI groups, we used multiple linear regression of FI on the midpoint MMBW^0.75 and ADG to estimate individual RFI (*Yang et al., 2021*).

RNA extraction and sequencing

Based on the results of the RFI calculation, five individuals with extremely low RFI (LRFI, high efficiency) and high RFI (HRFI, low efficiency) phenotypes were selected for slaughter after a 16-h fasting period. All experimental procedures involving animals were conducted by the Guidelines for Ethical Review of Laboratory Animal Welfare of Ningxia University (NXUC20211015). The hypothalamus (including the arcuate strong nucleus, parabrachial nucleus, supraoptic nucleus, dorsal/ventral medial nucleus, and other brain tissues) and the descending duodenum (mucosa, submucosa, and external muscular propria) were collected post-slaughter. Our tissue samples include hypothalamus and duodenum tissues from five high RFI and five low RFI cattle, totaling 20 samples. These tissue samples were washed with PBS to remove blood and other impurities, then cut into small pieces to increase surface area, and placed into sterile tubes for further processing. Total RNA was extracted from 500 mg of tissue samples using TRIzol method (TaKaRa Bio, Beijing, China), following the manufacturer's instructions. After extraction, the RNA was further purified using column purification to enhance purity and remove residual contaminants.

DNase treatment was performed to eliminate genomic DNA contamination. The quality and integrity of the extracted RNA were assessed using 1% agarose gel electrophoresis, Nanodrop, and Agilent 2100 to ensure a sample concentration of ≥500 ng/μL, 28S:18S > 1.0, and RIN ≥ 7. For library construction, the RNA was first reverse transcribed into cDNA. Adapter ligation was then performed to facilitate the sequencing process, followed by amplification to enrich the library. The library's initial quantification was carried out using Qubit 2.0, and the insert size was verified using an Agilent 2100. The effective concentration of the library (effective concentration > 2 nM) was accurately determined using qRT-PCR. The hypothalamus tissue samples underwent whole transcriptome sequencing to capture both coding and non-coding RNA species, providing a comprehensive view of the transcriptome. The duodenum tissue samples were subjected to standard transcriptome sequencing. Finally, pair-end sequencing data (raw data) with 150 bp read length were generated using the Illumina HiSeq 4000 platform. High RFI sequencing data of hypothalamus and duodenum were named Q_H1~Q_H5 and S_H1~S_H5, respectively, while low RFI sequencing data were named Q_L1~Q_L5 and S_L1~S_L5, respectively.

Quality control, mapping and transcript assembly

The statistical power of this experimental design, calculated in RNASeqPower (https://bioconductor.org/packages/release/bioc/html/RNASeqPower.html) is 0.86 (The corresponding code can be found in Script S1). The quality of raw data was assessed using the fastQC v.0.11.9. Subsequently, the Trimmomatic v.0.39 was used to perform quality control on the data. This included removing adapter sequences, trimming bases with Phred scores below 30 at the beginning and end of reads, applying a sliding window approach with a window size of 5 bp to remove bases with an average Phred score below 20, and discarding reads shorter than 75 bp. The cleaned data were then reevaluated using the fastQC software to ensure they met the requirements for subsequent analysis. The clean data were aligned to the bovine reference genome (ARS-UCD1.2; INSDC Assembly) using the STAR v.2.7.3a with the following parameters: –outSAMtype BAM Unsorted SortedByCoordinate, -outFilterMismatchNmax 999, –outFilterMismatchNoverReadLmax 0.04, –outFilterMultimapNmax 1. The resulting alignment files were further processed using the AddOrReplaceReadGroups tool of the PICARD v.2.27.4. This added the sample ReadGroups (RG) information to each alignment file. Additionally, the MarkDuplicates tool was applied to remove duplicate amplifications resulting from the PCR process during library construction.

Merging of sample data

For increasing the number of reads per variant locus, enhancing the depth coverage of reads across the entire transcriptome, as well as the depth coverage and quality of variant calls (*Lam et al., 2020*), in this study, the data from hypothalamic and duodenal tissues were merged into two BAM files based on phenotype (high RFI group and low RFI group). This merging process, performed using the "merge" command of the samtools v.1.16.1,

aimed to balance sequencing depth between samples and minimize the impact on SNP analysis results. Both the high RFI and low RFI groups in the subsequent analysis referred to the combined group data (Fig. 1).

SNPs recognition, filtering and annotation

BCFtools v.1.16 was utilized to execute the variant calling on the combined data of the high RFI group and low RFI group respectively, enabling identification of SNP sites and generating BCF files containing variant information. The "norm" parameter of BCFtools was then employed to normalize the variant information, thereby eliminating ambiguity caused by varying methods. Subsequently, the low-quality SNPs data underwent further filtering to reduce the likelihood of false positives and alleviate computational resource requirements for subsequent analysis. The software BCFtools and VCFtools v.0.1.16) were employed for variant filtering, employing the following criteria: (1) Removal of SNPs within a 5 bp range near indels; (2) setting a minimum coverage (DP) of 10; (3) enforcing a minimum allele frequency not less than 0.2 and a secondary allele depth not less than 2; (4) filtering loci with quality scores below 30. Finally, the functional annotation of SNPs was performed using the SnpEff v.5.1d with the built-in ARS-UCD1.2.105 database. The thresholds of above software are referenced from previous study (*Lam et al., 2020*).

Identifying and annotating high and low RFI group-specific SNPs

Using the SnpEff software, the VCF files underwent annotation, allowing for the identification of SNPs specific to the high and low RFI groups. SnpSift v.5.1d was then employed to screen for SNP loci with significant functional and modifier-type impacts. This enabled the selection of candidate genes associated with these SNP loci.

Gene function enrichment analysis and protein interaction network analysis

We employ clusterProfiler to conduct GO (Gene Ontology) and KEGG (Kyoto Encyclopedia of Genes and Genomes) enrichment analysis. The filter parameter is set as p value < 0.05. The GO enrichment can further clarify the main biological functions of the genes where the specific SNPs are located. The KEGG pathway enrichment can be used to understand the signal pathway regulated by the genes. Using the STRING database (https://cn.string-db.org/), we perform protein interaction analysis on the relevant genes to select core genes that have interaction effects.

Statistical analysis

The data from the experiment were analyzed and visualized using R v.4.3.0 (*R Core Team, 2023*) and Prism v.10.1.1 software. Statistical significance between the treatment and control groups was assessed using non-parametric tests or t-tests. A p-value of less than 0.05 was considered indicative of a significant difference. The corresponding code scripts can be found in Script S1.

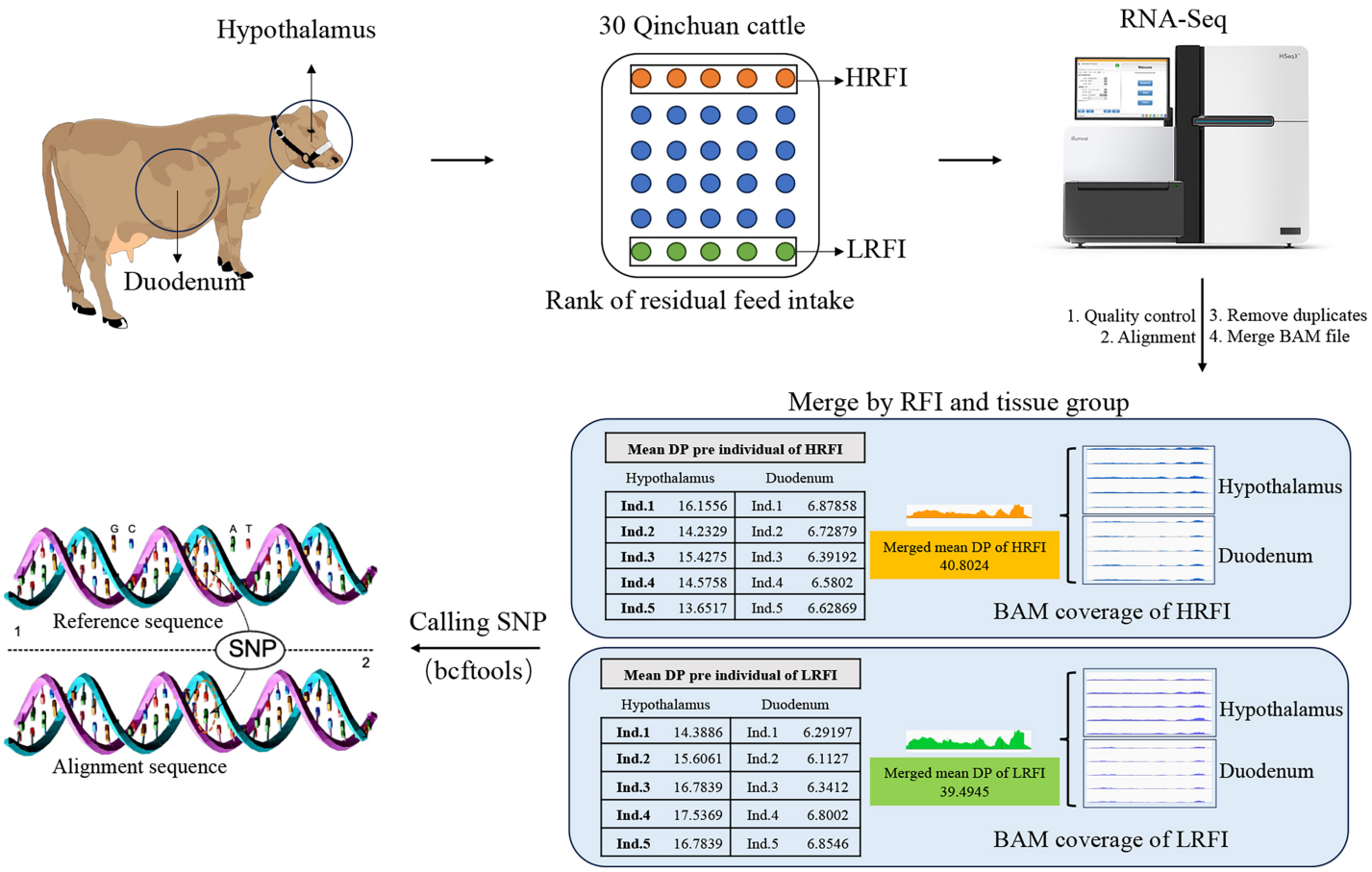

Figure 1 Sample collection and bioinformatics analysis. Hypothalamic and duodenal tissues from extreme RFI individuals were collected from 30 Qinchuan cows. After transcriptome sequencing, quality control, alignment, and deduplication, the alignment files were merged. The data from the two tissues were merged according to the RFI group, resulting in two alignment files. The merger greatly increased the depth (DP) of the reads, and the average reads DP of the two groups was basically consistent. Finally, the BCFtools software was used to identify SNPs in the merged data.
Full-size DOI: 10.7717/peerj.19270/fig-1

RESULTS

RNA-seq sequencing data quality and comparisons

From the RNA-seq data of the hypothalamus and duodenum, we obtained a total of 1,019 million and 275 million paired sequencing reads, respectively. After quality control, all 20 samples in this study had a Q30 score (error rate $p < 0.001$) above 96%. The GC% content was approximately 50% (Table S1). The clean data was aligned to the bovine reference genome, where the percentage of reads aligning to the reference genome was above 91%, and the percentage of reads with a unique alignment position ranged from 81.07% to 93.81% (Table S2). Analysis of the transcript expression for each sample indicated relatively consistent transcript abundances (Fig. 2). These results demonstrate the high quality of the obtained data, reducing the impact of sequencing errors on subsequent analysis.

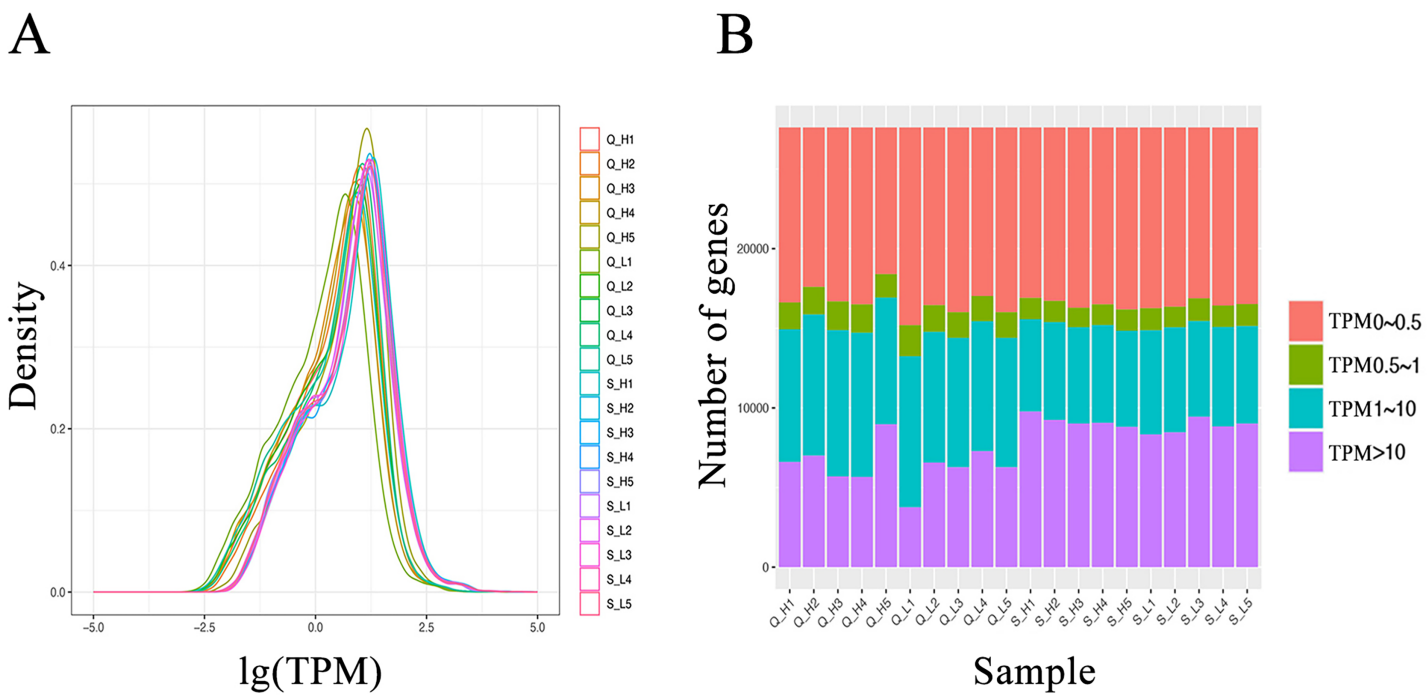

Figure 2 Transcript expression (TPM) analysis of each sample. (A) Transcript expression density in each sample. (B) Transcript expression in each sample.
Full-size DOI: 10.7717/peerj.19270/fig-2

Table 1 Statistics on the number of homozygote and heterozygote SNPs in different RFI groups.

RFI group	SNP statistics	Homozygote sites	Heterozygote sites
HRFI	Number of SNPs	11,991	258,419
	Probability	4.43%	95.57%
LRFI	Number of SNPs	14,007	241,113
	Probability	5.49%	94.51%

SNPs screening and analysis

SNPs in the high and low RFI groups were identified by combining RFI phenotype with RNA-Seq data from hypothalamic and duodenal tissues. The numbers of homozygous and heterozygous mutations in the combined samples were counted (Table 1), revealing that there were 270,410 and 255,120 specific SNPs in the high and low RFI groups, respectively. Among these SNPs, 11,991 (4.43%) and 14,007 (5.49%) were homozygous mutations, with heterozygous mutations far outnumbering homozygous mutations. Statistical analysis based on the different types of mutations in the SNPs (Fig. 3) showed that the total number of transition types (A–G, C–T) was higher than the total number of transversion types (A–T, C–G, A–C, G–T). Among the six types of single nucleotide variations, A–G and C–T had the highest occurrence rates, while the occurrence rates of the other four types of transitions were relatively lower. In the high RFI group, A–G accounted for 36.5% and C–T accounted for 35.4%. In the low RFI group, A–G accounted for 35.7% and C-T accounted

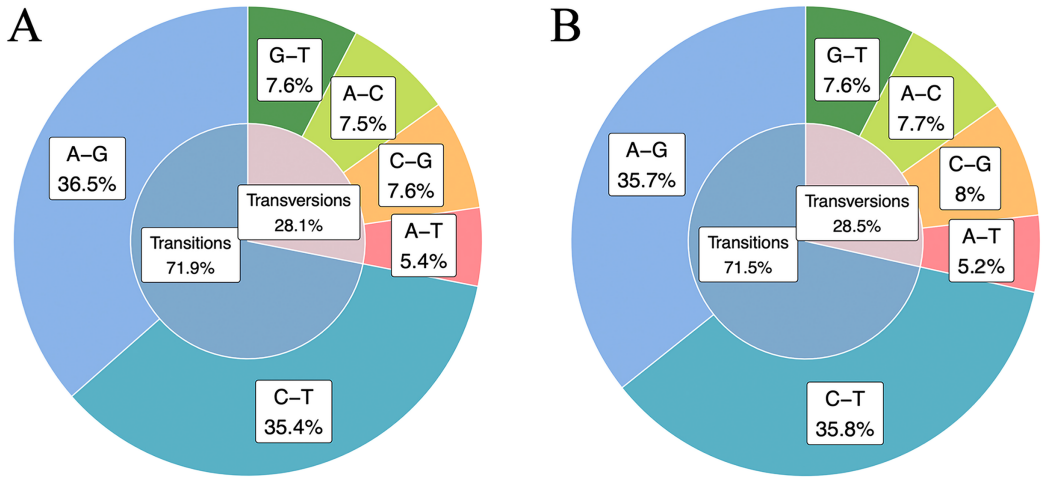

Figure 3 High RFI group and low RFI group combined data SNPs type statistics. (A) Statistics on SNPs types in the high RFI group. (B) Statistics on SNPs types in the low RFI group.
Full-size DOI: 10.7717/peerj.19270/fig-3

for 35.8%. The Ts/Tv ratios in the high and low RFI groups were 2.55 and 2.50, respectively. The differences in occurrence rates of the six types of single nucleotide variations between the high and low RFI groups were small, with a mean occurrence rate of transitions being 71.72% and transversions being 28.29%.

Distribution statistics of SNPs

The study investigated the distribution and variation of SNPs across different chromosomes, providing insights into the genetic diversity among genes. Analysis of the combined high and low RFI groups revealed no significant difference ($p > 0.05$) in distribution between the two groups. Chromosome 1 exhibited the highest number of SNPs, while the mitochondria (MT) showed the lowest distribution (Fig. 4A). To account for differences in chromosome length, the ratio of SNPs number to chromosome length was calculated, revealing that the MT had the highest variant rate among both groups, indicating a higher density of SNPs per unit length (Fig. 4B). The chromosomal distribution of SNPs within the high and low RFI groups is delineated in Figs. 4C and 4D.

Comprehensive statistical analysis of SNP loci distribution across the genome was conducted for both high and low RFI datasets, emphasizing different genomic functional regions, such as downstream, exon, intergenic, intron, untranslated region, *etc.* (Table 2). The statistical analysis of the SNP locations in the genome for the high RFI and low RFI groups shows ten different distributions (DOWNSTREAM, EXON, INTERGENIC, INTRON, SPLICE_SITE_ACCEPTOR, SPLICE_SITE_DONOR, SPLICE_REGION, UPSTREAM, UTR_3_PRIME, UTR_5_PRIME) (Table 2). A single SNP may be located in multiple transcript regions. The analysis found that the INTRON region had the most SNP locations in both groups, with 429,995 areas annotated in the HRFI group and 435,881 areas in the LRFI group, significantly more than other functional regions. The next most abundant functional elements are INTERGENIC and DOWNSTREAM, while the remaining functional regions are less common. SPLICE_SITE_ACCEPTOR has the fewest

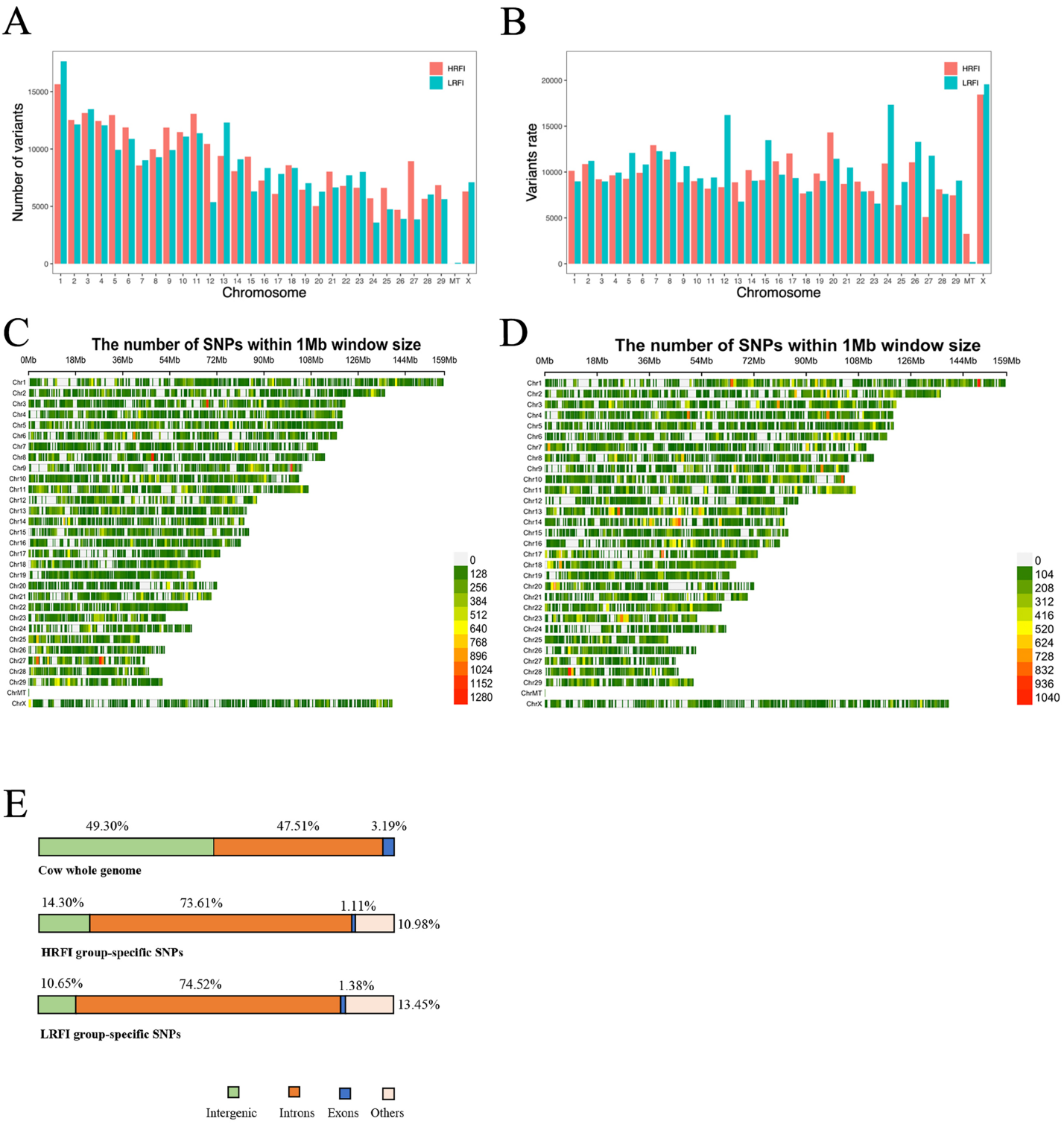

Figure 4 SNPs distribution statistics on chromosomes. (A) Statistics on the number of SNPs on different chromosomes. (B) SNPs number and length ratio statistics on different chromosomes. (C) Distribution of SNPs on chromosome locations in the high RFI group. (D) Distribution of SNPs on chromosome locations in the low RFI group. (E) Percentage of the genome comprising each type of feature (top) and the proportion of SNPs detected by HRFI group-specific SNPs (middle) and LRFI group-specific SNPs (bottom) across these genomic features.

Full-size DOI: 10.7717/peerj.19270/fig-4

Table 2 Genomic functional annotation of SNPs.

	HRFI group		LRFI group	
	Count	Percent	Count	Percent
DOWNSTREAM	35,350	6.05%	43,281	7.40%
EXON	6,504	1.11%	8,070	1.38%
INTERGENIC	83,530	14.30%	62,282	10.65%
INTRON	429,995	73.61%	435,881	74.52%
SPLICE_SITE_ACCEPTOR	37	0.01%	73	0.01%
SPLICE_SITE_DONOR	80	0.01%	169	0.03%
SPLICE_SITE_REGION	309	0.05%	366	0.06%
UPSTREAM	22,703	3.89%	27,289	4.67%
UTR_3_PRIME	4,371	0.75%	5,920	1.01%
UTR_5_PRIME	1,251	0.21%	1,621	0.28%

functional regions, with 37 in the high RFI group and 73 in the low RFI group. We hope to find most of the SNPs in the exonic regions, but coding regions generally experience higher selective pressure compared to non-coding regions (*Zhao et al., 2003*). The annotation of SNPs in the high and low RFI groups accounts for 1.11% and 1.38% in the exonic regions, respectively. At the same time, this also explains our detection results: the higher distribution of SNPs in intron regions is partly due to the fact that unspliced transcripts are also detected during sequencing, and partly because intron regions constitute 47.51% of the whole genome, which is significantly higher than the length of exonic regions (Fig. 4E). SNPs located in intergenic regions may be found in new genes or gene portions that have not been annotated yet.

Influence prediction and amino acid change

By using the SnpEff software to evaluate the potential effects of SNP mutations on codons, it was found that over 98% of the SNPs in both groups were classified as modifiers, having minimal effect on genes and proteins. However, 159 SNPs in the high RFI group and 293 SNPs in the low RFI group were predicted to have a high effect (Figs. 5A, 5B; Tables S3, S4), warranting further investigation. This situation is as we expected, most SNPs are located in intron regions and intergenic areas, making it difficult to directly affect protein coding. Therefore, high-impact SNPs will be relatively fewer. Additionally, analyzing the overall levels of each amino acid can provide important insights into evolutionary pressures and adaptation mechanisms. This approach helps identify patterns and frequencies of amino acid substitutions in different biological contexts, enhancing our understanding of how these changes affect protein stability, function, and interactions. To analyze the potential impact of SNPs on genes and proteins, the effect of intergroup-specific SNPs (Tables S5, S6) on codons and subsequent amino acids was assessed. The analysis revealed that the amino acids most affected in both the high and low RFI groups were alanine-threonine, alanine-valine, and isoleucine-valine (Figs. 5C, 5D). By identifying SNP loci that significantly impacted both the high and low RFI groups and

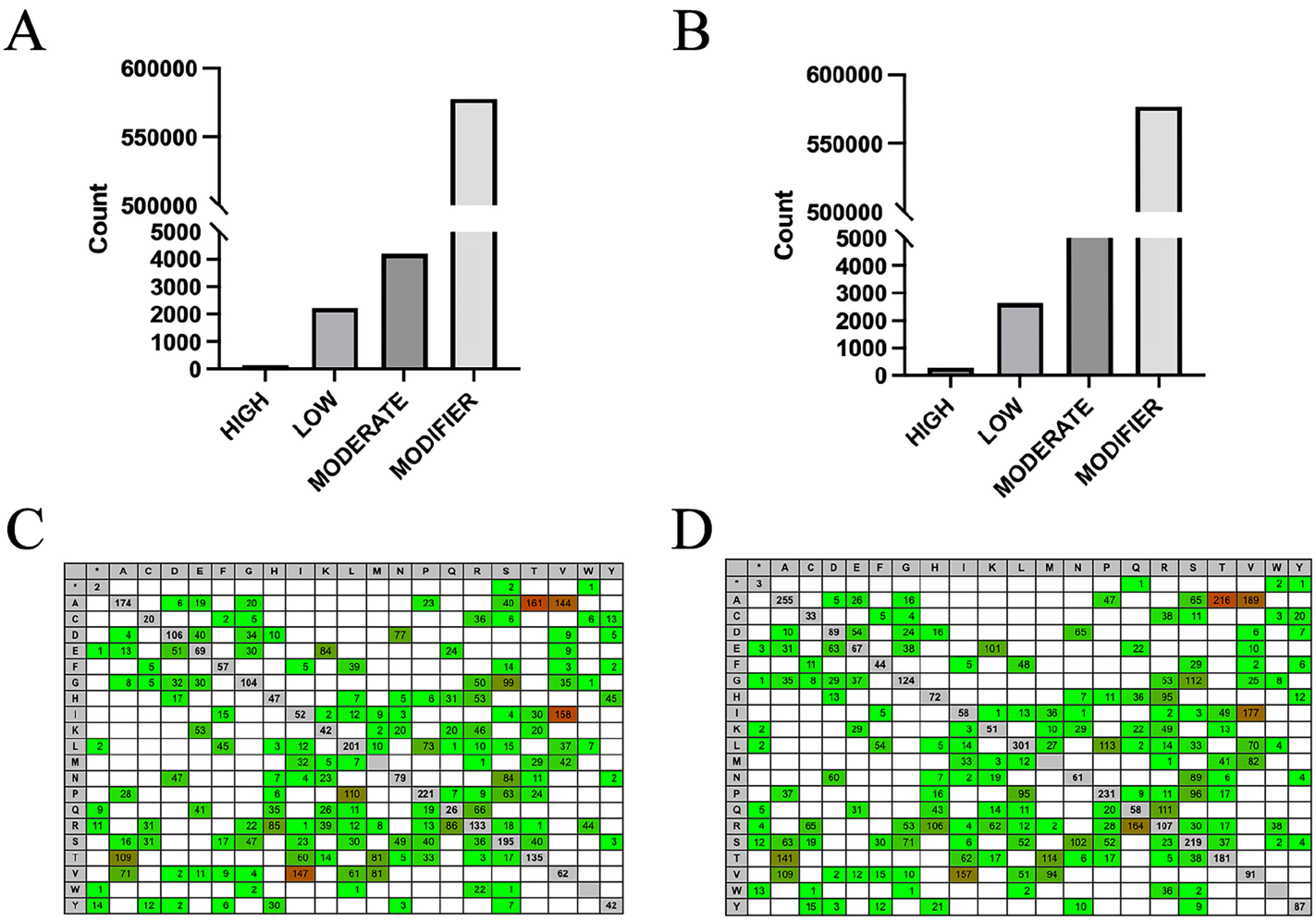

Figure 5 Influence prediction and amino acid change. (A) High RFI group specific SNPs influence prediction statistics. (B) Low RFI group specific SNPs influence prediction statistics. (C) The amino acid changes caused by SNPs in the high RFI group were reference amino acids horizontally and altered amino acids vertically. Red background colors indicate that more changes happened (heat-map). (D) The amino acid changes caused by SNPs in the low RFI group.
Full-size DOI: 10.7717/peerj.19270/fig-5

mapping them to the corresponding genes using the SNPs annotation files, a total of 83 genes were identified in the high RFI group and 97 genes in the low RFI group. Interestingly, one gene, JSP.1, belonging to the MHC class I family, was common in both groups, and played a key role in regulating animal health within the immune system (*Hewitt, 2003*).

Genes function annotation of high-impact SNP loci

GO functional annotation and enrichment analysis were conducted for the aforementioned genes (Figs. 6A, 6B). The results revealed that the enriched genes in the high and low RFI groups were primarily associated with protein binding and enzyme binding processes. Notably, a significant number of genes related to NADH activity were found in the low RFI (high feed efficiency) group. These genes were associated with oxidoreductase activity, acting on NADH or NADPH; NADH dehydrogenase activity;

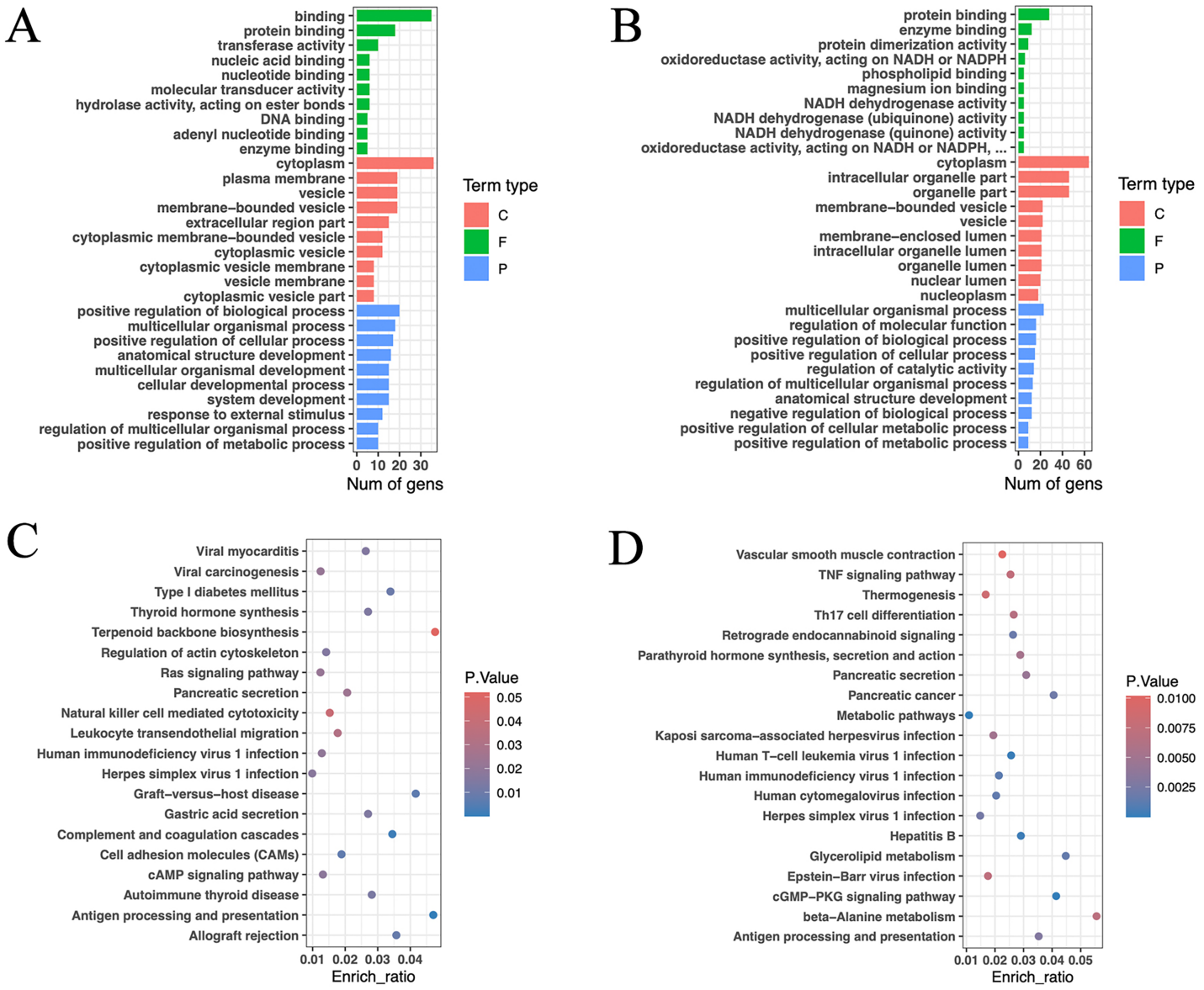

Figure 6 Gene function annotation of high-impact SNP loci. (A) Gene GO enrichment analysis ($p < 0.05$) of HRFI group-specific high-impact SNP sites, and select the top 10 for each term type based on p-value. (C: cellular component; F: molecular function; P: biological process). (B) Gene GO enrichment analysis ($p < 0.05$) of LRFI group-specific high-impact SNP sites, and select the top 10 for each term type based on p value. (C) Gene KEGG enrichment analysis ($p < 0.05$) of HRFI group-specific high-impact SNP sites, and select the top 20 based on p value. (D) Gene KEGG enrichment analysis ($p < 0.05$) of LRFI group-specific high-impact SNP sites, and select the top 20 based on p value.

Full-size DOI: 10.7717/peerj.19270/fig-6

NADH dehydrogenase (ubiquinone) activity; NADH dehydrogenase (quinone) activity; and oxidoreductase activity, acting on NADH or NADPH, quinone, or similar compounds as acceptors. NADH and its phosphorylation product NADPH play pivotal roles as coenzymes in various metabolic activities, including cell signaling, protein modification, energy metabolism, mitochondrial function, calcium homeostasis, antioxidative stress, biosynthesis, and cell death (*Berger, Ramirez-Hernandez & Ziegler, 2004*; *Patterson et al., 2005*; *Xiao et al., 2018*; *Ying, 2006*, *2007*, *2008*). In particular, the enrichment genes *ND4*,

ND5, and *ND6* are core subunits of the mitochondrial respiratory chain NADH dehydrogenase (complex I). They facilitate the transfer of electrons from NADH through the respiratory chain, utilizing ubiquinone as an electron acceptor, and are crucial for the catalysis and assembly of complex I (*UniProt, 2023*).

The KEGG enrichment analysis outcomes indicated enrichment of pathways related to thyroid hormone synthesis, pancreatic secretion, gastric acid secretion, cAMP signaling pathway, thermogenesis, parathyroid hormone synthesis and secretion, glycerolipid metabolism, TNF signaling pathway and beta-alanine metabolism in the high and low RFI groups (Figs. 6C, 6D). Notably, the thermogenesis pathway exhibited enrichment of *ND4*, *ND5*, and *ND6* genes. Additionally, *ATP1A2*, *SLC9A4*, and *PLA2G5* were identified as genes associated with energy metabolism (*Lingrel, 1992*; *Sakuta et al., 2020*; *Sun et al., 2004*).

Protein-protein interaction analysis of high-impact SNP loci

Protein-protein interaction analysis is a method used to study the interactions between proteins, which can be employed to uncover the relationships and networks among proteins, consequently explaining functional interactions and illustrating the intricate interconnections between proteins. Our results revealed distinct patterns of core genes and interaction relationships between the two groups. In the high RFI group, we identified 29 core genes and 23 interaction relationships, while in the low RFI group, we found 42 core genes and 41 interaction relationships (Fig. 7). Several genes, such as *HSP90AA1*, *EIF2AK3*, *PAK1*, *MAP3K7*, *PGM2L1*, *DNM1L* and *CYB5R3*, were found to be related to energy metabolism, fat deposition and muscle development (*Badri et al., 2018*; *Charoensook et al., 2012*; *Chen et al., 2019*; *Chiang & Jin, 2014*; *Hogarth et al., 2018*; *Liu et al., 2024*; *Lopez-Bellon et al., 2022*; *Zhang, O'Keefe & Jonason, 2017*).

Analysis of candidate gene SNP loci

Based on the results of GO and KEGG analysis, we focused on phenotype-related terms. In the high RFI group we screened GO terms: positive regulation of metabolic process and multicellular organismal development; KEGG terms: thyroid hormone synthesis, gastric acid secretion, cAMP signaling pathway, pancreas signaling pathway. In the low RFI group we screened GO terms: oxidoreductase activity, acting on NADH or NADPH; KEGG terms: thermogenesis, metabolic pathways, TNF signaling pathway. Finally, 18 genes were identified in the high RFI group and 21 genes in the LRFI group (Tables S3, S4, S7). Also combining protein interaction analysis and existing studies, we finally focused on 14 genes. In these genes, we combined the prediction results of SNP impact, showing high-impact SNP sites in the genes. We found that these SNPs are mostly located in the exon regions, and most are A-G mutation types. This type of mutation might change the coded amino acids, and affect the structure and function of the protein (Table 3). However, their specific function and roles would be detected in the future studies for clarifying their variant effect to phenotype.

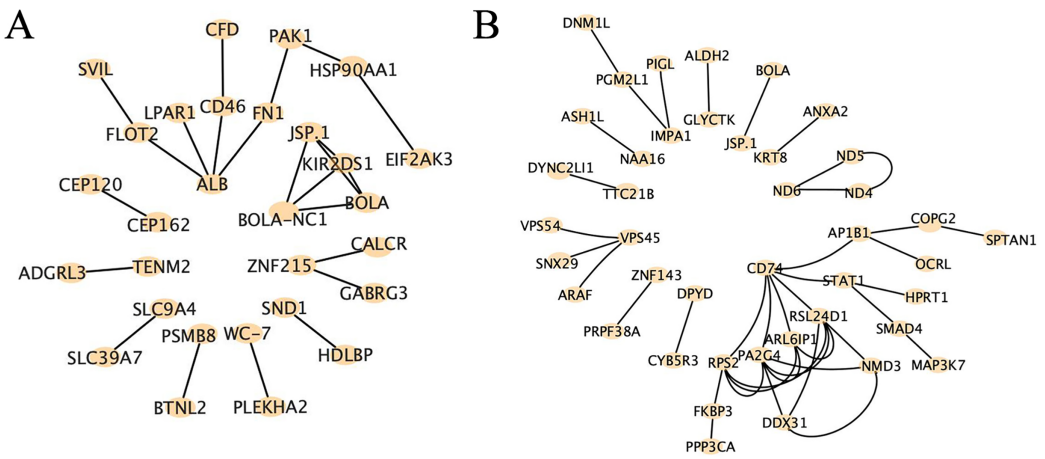

Figure 7 Protein-protein interaction analysis of high-impact SNP loci. (A) Protein-protein interaction analysis of high-impact SNP sites in HRFI group, using string database, composed of 29 nodes and 23 edges. (B) Protein-protein interaction analysis of high-impact SNP sites in LRFI group, using string database, composed of 42 nodes and 41 edges. Full-size DOI: 10.7717/peerj.19270/fig-7

Table 3 Location of SNPs from candidate genes.

Gene	Chromosome	Site	Reference nucleotide	Mutated nucleotide	Group	References
ATP1A2	3	9,540,257	C	G	High RFI	*Lingrel (1992)*
SLC9A4	11	7,280,369	G	A	High RFI	*Sakuta et al. (2020)*
PLA2G5	2	132,645,789	G	A	High RFI	*Sun et al. (2004)*
HSP90AA1	21	66,941,687	C	T	High RFI	*Badri et al. (2018)*, *Charoensook et al. (2012)*
EIF2AK3	11	47,484,569	G	A	High RFI	*Chen et al. (2019)*
PAK1	29	18,525,245	A	T	High RFI	*Chiang & Jin (2014)*
MAP3K7	9	59,698,567	A	T	Low RFI	*Zhang, O'Keefe & Jonason (2017)*
PGM2L1	15	53,718,210	A	G	Low RFI	*Liu et al. (2024)*
DNM1L	5	77,174,829	A	C	Low RFI	*Hogarth et al. (2018)*
CYB5R3	5	113,485,291	T	G	Low RFI	*Lopez-Bellon et al. (2022)*
SMAD4	24	50,564,735	G	A	Low RFI	*Li et al. (2020)*
ND4	MT	11,329	A	G	Low RFI	*Yang et al. (2023)*
ND5	MT	12,672	G	A	Low RFI	*Yang et al. (2023)*
ND6	MT	14,066	C	T	Low RFI	*Yang et al. (2023)*

DISCUSSION

Enhancing animal feed efficienvy and reducing production costs are key objectives in livestock production, as they are essential for animal growth and development. Besides development high-quality feeds through selective breeding, investigating genetic factors affecting feed efficiency in beef cattle is a significant research focus. The hypothalamus acts as a central control regulator of feeding, interacting with neuron groups to produce signals that stimulate or suppress appetite, thereby influencing food intake (*Perkins et al., 2014*; *Sartin, Whitlock & Daniel, 2011*). The duodenum, the initial segment of the small intestine,

is crucial for digestion and absorbing nutrients, particularly carbohydrates and micronutrients (*Cooke & Clark, 1976*; *Reeves & Chaney, 2004*). The brain-gut axis, comprising the central, enteric, and autonomic nervous systems, facilitates the complex communication between the gut and the brain *via* neurohumoral pathways (*Margolis, Cryan & Mayer, 2021*). Axes like the hypothalamic-pituitary-adrenal (HPA) axis, part of the brain-gut axis, are identified as key contributors to the variability in RFI (*DiGiacomo et al., 2018*).

RNA-Seq, a second-generation transcriptome sequencing technology, offers various approaches approaches for high-throughput functional genomics, including gene expression profiling (*Song et al., 2019*), genome annotation (*Li et al., 2011*), non-coding RNA discovery (*Jiang et al., 2022*), and gene mutation analysis (*Lopez-Maestre et al., 2016*). These approaches collectively elucidate the intricate complexities biological systems. Residual feed intake, a quantitative trait, is governed by multiple genes and influenced by diverse physiological metabolic processes (*Arthur et al., 2001*). Previous research has effectively identified genes linked to feed utilization efficiency through SNP screening. For instance, *Higgins et al. (2018)* found a strong association between the variant rs43555985 and RFI (P = 8.28E−06). *Bolormaa et al. (2011)* identified 111 and 75 significantly associated SNPs with RFI ($p < 0.001$) using the 10 K and 50 K SNP microarray data, respectively. *Lima et al. (2016)* used comprehensive GWA, AWM, and RNA-Seq analyses to identify the *PRUNE2* gene as a potential candidate affecting feed efficiency. Various tools have been developed for SNP detection from RNA-seq data and for determining concordance of SNP and genotype detection between RNA-seq and DNA-seq (*Dobin et al., 2013*; *Liu, Shen & Bao, 2022*; *Luo et al., 2019*; *Quinn et al., 2013*; *Tang et al., 2014*; *Van der Auwera et al., 2013*). In our study, we collected hypothalamic and duodenal tissues from beef cattle with high and low RFI. Using high-throughput transcriptome sequencing, we obtained data from 20 samples and merged the tissue data by RFI groups to identify SNPs, aiming to improve the accuracy of SNP functional annotation.

SNPs represent DNA sequence variations due to single nucleotide changes. Analysis SNPs data from 20 samples revealed 270,410 unique SNPs in the high RFI group and 255,120 in the low RFI group. The high RFI group had 11,991 homozygous SNPs, while the low RFI group had 14,007. Over 70% of SNPs in both groups were located in the intron region, followed by the intergenic region, likely due to unspliced transcripts (premature transcripts) and unannotated regions (*Jehl et al., 2021*). However, only 1.11% and 1.38% of SNPs in high and low RFI groups were in the intron region. This distribution is anticipated, as intronic regions typically face higher selection pressures compared to other non-coding regions (*Zhao et al., 2003*). Furthermore, the transition-to-transversion ratio (Ts/Tv) was of 2.55 in the high RFI group and 2.50 in the low RFI group, reflecting a higher incidence of transition mutations, consistent with previous findings (*Nandanpawar et al., 2023*; *Raizada & Souframanien, 2019*; *Van Deventer et al., 2020*). The consistency of Ts/Tv values supports the reliability of SNP identification in this study (*Arabnejad et al., 2018*). Using SnpEff software for SNP annotations, functionally significant SNPs unique to the high and low RFI groups were identified, along with their corresponding genes. A total of

83 genes were found in the high RFI group, while 97 genes were identified in the low RFI group.

GO and KEGG pathway enrichment analyses were conducted independently for each gene set. At the molecular function level, the enriched GO terms in both groups were primarily associated with protein binding and enzyme binding. In the low RFI group, specific enrichment was noted for NADH-related terms, including oxidoreductase activity acting on NADH or NADPH, NADH dehydrogenase activity, NADH dehydrogenase (ubiquinone) activity, NADH dehydrogenase (quinone) activity, and oxidoreductase activity acting on NADH or NADPH, quinone or similar compound as acceptor. NADH dehydrogenase, also known as NADH: ubiquinone oxidoreductase or complex I, facilitating electron transfer from NADH to coenzyme Q, crucial for energy metabolism in the mitochondrial inner membran (*Nakamaru-Ogiso et al., 2010*). The genes *ND4*, *ND5*, and *ND6* are core subunits of mitochondrial respiratory chain NADH dehydrogenase (complex I), essential for its catalytic function and assembly (*UniProt, 2023*). Previous genomic analyses have suggested that the *ND (2,3,4,4L,5,6)* gene cluster may significantly impact feed efficiency changes (*Yang et al., 2023*). Mitochondria generate approximately 90% of cellular energy and are abundant in metabolically active cells, such as liver, kidney, muscle, and brain cells. Studies in poultry and livestock have shown a close relationship between feed efficiency and mitochondrial function and biochemistry. Research indicated that animals with low RFI exhibit increased rates of mitochondrial respiration (*Kolath et al., 2006*), enhanced coupling of the electron transport chain (*Bottje & Carstens, 2009*), higher activity of respiratory chain complexes I-V (*Iqbal et al., 2005*), and lower heat production per kilogram of metabolic body weight (MBW) (*Nkrumah et al., 2006*). Moreover, the electron transport chain is also recognized as the site of reactive oxygen species (ROS) production, and elevated ROS levels pose a significant threat to the antioxidant defense system by increasing the susceptibility of various cellular components to oxidative damage (*Nolfi-Donegan, Braganza & Shiva, 2020*). Animals with higher feed efficiency tend to exhibit lower oxidative stress phenomena (*Bottje & Carstens, 2009*; *Iqbal et al., 2005*, *2004*). KEGG pathway enrichment analysis also identified several pathways associated with energy metabolism. In the high RFI group, enriched pathways included thyroid hormone synthesis, pancreatic secretion, gastric acid secretion, and cAMP signaling pathway were enriched. Enriched pathways in the low RFI group included thermogenesis, parathyroid hormone synthesis, secretion and action, pancreatic secretion, triglyceride metabolism, and alanine metabolism.

One of the enriched genes, *ATP1A2*, is involved in ATP hydrolysis and facilitates sodium and potassium ion exchange across the plasma membrane, establishing an electrochemical gradient for the active transport of nutrients (*Lingrel, 1992*). Additionally, *SLC9A4* functions act as a sodium ion sensor, regulating water intake behavior (*Sakuta et al., 2020*). *PLA2G5* is speculated to play a role in the biosynthesis of N-acyl ethanolamines, compounds that regulate energy metabolism (*Sun et al., 2004*). Through protein-protein interaction analysis, we have discovered that genes such as *HSP90AA1*, *EIF2AK3*, *PAK1*, *SMAD4*, *MAP3K7*, *PGM2L1*, *DNM1L*, and *CYB5R3* were found to be related to metabolism, which might be related to cattle RFI variants. For example, genetic

variations in *HSP90AA1* are associated with thermoregulatory traits in cattle (*Badri et al., 2018*; *Charoensook et al., 2012*). Activation of *EIF2AK3* has been shown to promote metabolic dysfunctions (*Chen et al., 2019*). *PAK1* is involved in the regulation of glucose uptake (*Chiang & Jin, 2014*). *SMAD4* is linked to aerobic glycolysis and obesity (*Li et al., 2020*). *MAP3K7* induces adipocyte differentiation *via PPARγ* signaling (*Zhang, O'Keefe & Jonason, 2017*). *PGM2L1* is suggested to be related to meat quality and muscle development in sheep (*Liu et al., 2024*). Variations in *DNM1L* can lead to mitochondrial fragmentation, decreased membrane potential, reduced oxidative capacity, and increased levels of reactive oxygen species (ROS) (*Hogarth et al., 2018*). *CYB5R3* works with coenzyme Q, participating in the cross-membrane redox system to protect cells against oxidative stress (*Lopez-Bellon et al., 2022*). These genes will be targeted in future cellular and molecular experiments to validate their associations with the RFI trait.

RFI, an essential economic trait in feed efficiency research, necessitates a thorough understanding of the genetic mechanisms associated with SNP loci and their impact on RFI regulation in beef cattle. This understanding is vital for analyzing RFI variation in livestock and improving feed conversion efficiency for sustainable and cost-effective animal husbandry. Additionally, enhancing feed utilization efficiency can reduce methane emissions, improve animal health and production performance, and serve as a foundation for selecting and breeding feed-efficient beef cattle. The use of transcriptome data to identify SNPs in the investigation of RFI offers several distinct advantages. Firstly, transcriptome data provides valuable insights into gene expression, thereby enabling researchers to focus directly on genes and SNPs that are associated with specific physiological processes (*Jehl et al., 2021*). Secondly, the acquisition of transcriptome data is relatively cost-effective, particularly in species for which a reference genome is not available (*Lopez-Maestre et al., 2016*). Furthermore, transcriptome data can be integrated with genomic data, leveraging methodologies such as genome-wide association studies (GWAS) and expression quantitative trait locus (eQTL) analyses to further validate the associations between SNPs and RFI (*Ibragimov et al., 2022*). By optimizing the analytical workflows for RNA-Seq data, the accuracy and reliability of SNP detection can be significantly enhanced, thereby reducing the incidence of false positives. Additionally, transcriptome data can be employed to validate SNPs identified from genomic data, thereby augmenting the reliability of the results (*Ge, Li & Zhang, 2024*; *Lam et al., 2020*).

Although we have obtained many SNPs that are meaningful and may have critical genetic effects. However, given many drawbacks of RNA-Seq, such as the uneven depth distribution of reads across the genome from RNA-Seq data (*Jehl et al., 2021*), the large variation in RNA expression levels in different tissues, cells, and physiological stages (*Sims et al., 2014*), and the fact that the variants detected at the RNA level may not exist at the DAN level (*Jehl et al., 2021*), and the fact that the SNPs detection near exon-exon junctions still needs to remain cautious (*Lagarrigue et al., 2013*; *Peng et al., 2012*). SNPs analysis from RNA-seq data should also continuously improve its identification efficiency, or combine with other methods such as sanger sequencing, flight mass spectrometry, fluorescent probes and so on to improve the recognition rate and accuracy of valid SNPs, so as to effectively improve the efficiency of the identification of SNPs in the coding region, to

increase the reliability of the data results, and to reduce false positives. This study also has certain limitations. SnpEff primarily relies on annotation databases and predefined rules to predict the functional impact of SNPs. While this provides a quick assessment, it lacks the comprehensive validation offered by multiple bioinformatics tools. For instance, tools like SIFT (*Ng & Henikoff, 2003*), PolyPhen2 (*Adzhubei et al., 2010*), and Panther (*Tang & Thomas, 2016*) can provide more detailed functional impact predictions. The use of a single tool may lead to partial and inaccurate predictions. Moreover, directly provide predictions on protein stability changes is significative. Tools such as I-Mutant (*Capriotti, Fariselli & Casadio, 2005*), Mupro (*Laskar et al., 2023*), and CUPSAT (*Parthiban, Gromiha & Schomburg, 2006*) can predict the impact of SNPs on protein stability by calculating changes in free energy (ΔΔG), which is crucial for understanding how mutations affect protein folding and function. Molecular dynamics simulations can offer detailed information on how mutations impact protein structure and dynamics, revealing conformational changes and functional effects (*Elangeeb et al., 2024*; *Kamal et al., 2024*). To achieve more comprehensive and accurate SNP functional predictions, it is recommended to integrate multiple bioinformatics tools and methods. This approach will enhance the accuracy and reliability of the research findings.

CONCLUSIONS

Due to its low cost and effective detection, RNA-Seq data has become a reliable resource for polymorphism detection in non-model animals. In this study, RFI-related SNPs and their annotated genes were obtained by integrating multiple tissue RNA-seq data from extreme RFI individuals to improve SNP identification. Variants calling based on RNA-seq data can effectively improve the identification of phenotype-related SNPs, which is an efficient and feasible approach to get potential functional SNPs. By mining SNPs with high impact on genes, this genes and SNPs related to RFI would be helpful and valuable for molecular validation in subsequent studies.

ACKNOWLEDGEMENTS

We thank all the cattle breeding staff who helped with sample collecting in this experiment and all members of this research group for all valuable discussions and feedback on this article.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the National Natural Science Foundation of China (32160776), a grant from the China Agriculture Research System (CARS-36), the Natural Science Foundation of Ningxia Province, China (Grant No. 2021AAC03027), and West Light Foundation of the Chinese Academy of Sciences (Grant No. XAB2022YW11). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Natural Science Foundation of China: 32160776.
China Agriculture Research System (CARS-36).
Natural Science Foundation of Ningxia Province, China: 2021AAC03027.
and West Light Foundation of the Chinese Academy of Sciences: XAB2022YW11.

Competing Interests

Cong-Jun Li is an Academic Editor for PeerJ.

Author Contributions

- Zonghua Su performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Chenglong Li performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Chaoyun Yang performed the experiments, prepared figures and/or tables, and approved the final draft.
- YanLing Ding performed the experiments, prepared figures and/or tables, and approved the final draft.
- Xiaonan Zhou performed the experiments, prepared figures and/or tables, and approved the final draft.
- Junjie Xu analyzed the data, prepared figures and/or tables, and approved the final draft.
- Chang Qu analyzed the data, prepared figures and/or tables, and approved the final draft.
- Yuangang Shi conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Cong-Jun Li conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.
- Xiaolong Kang conceived and designed the experiments, authored or reviewed drafts of the article, and approved the final draft.

Animal Ethics

The following information was supplied relating to ethical approvals (*i.e.*, approving body and any reference numbers):

All experimental animal studies were reviewed and ap- proved by the Animal Welfare Committee of Ningxia University (permit number NXUC20211015).

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

The duodenal tissue data is available at ENA: PRJNA747740.

Data Availability

The following information was supplied regarding data availability:

The code is available in the Supplemental File.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.19270#supplemental-information.

REFERENCES

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. *Nature Methods* **7(4)**:248–249 DOI 10.1038/nmeth0410-248.

Anand P, Kumar AV, Ravi K, Simmi T. 2021. Differential gene expression in duodenum of colored broiler chicken divergently selected for residual feed intake. *Tropical Animal Health and Production* **53**:1–10 DOI 10.1007/s11250-020-02519-9.

Arabnejad M, Dawkins BA, Bush WS, White BC, Harkness AR, McKinney BA. 2018. Transition-transversion encoding and genetic relationship metric in ReliefF feature selection improves pathway enrichment in GWAS. *BioData Mining* **11(1)**:23 DOI 10.1186/s13040-018-0186-4.

Archer JA, Bergh L. 2000. Duration of performance tests for growth rate, feed intake and feed efficiency in four biological types of beef cattle. *Livestock Production Science* **65(1–2)**:47–55 DOI 10.1016/S0301-6226(99)00181-5.

Arthur PF, Archer JA, Johnston DJ, Herd RM, Richardson EC, Parnell PF. 2001. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. *Journal of Animal Science* **79(11)**:2805–2811 DOI 10.2527/2001.79112805x.

Badri TM, Chen KL, Alsiddig MA, Li L, Cai Y, Wang GL. 2018. Genetic polymorphism in Hsp90AA1 gene is associated with the thermotolerance in Chinese Holstein cows. *Cell Stress and Chaperones* **23(4)**:639–651 DOI 10.1007/s12192-017-0873-y.

Berger F, Ramirez-Hernandez MH, Ziegler M. 2004. The new life of a centenarian: signalling functions of NAD(P). *Trends in Biochemical Sciences* **29(3)**:111–118 DOI 10.1016/j.tibs.2004.01.007.

Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, Arthur PF, Herd RM, Goddard ME. 2011. Genome-wide association studies for feedlot and growth traits in cattle. *Journal of Animal Science* **89(6)**:1684–1697 DOI 10.2527/jas.2010-3079.

Bottje WG, Carstens GE. 2009. Association of mitochondrial function and feed efficiency in poultry and livestock species. *Journal of Animal Science* **87**:E48–E63 DOI 10.2527/jas.2008-1379.

Canovas A, Rincon G, Islas-Trejo A, Wickramasinghe S, Medrano JF. 2010. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mammalian Genome* **21(11–12)**:592–598 DOI 10.1007/s00335-010-9297-z.

Capriotti E, Fariselli P, Casadio R. 2005. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Research* **33**:W306–W310 DOI 10.1093/nar/gki375.

Charoensook R, Gatphayak K, Sharifi AR, Chaisongkram C, Brenig B, Knorr C. 2012. Polymorphisms in the bovine HSP90AB1 gene are associated with heat tolerance in Thai indigenous cattle. *Tropical Animal Health and Production* **44(4)**:921–928 DOI 10.1007/s11250-011-9989-8.

Chen S, Henderson A, Petriello MC, Romano KA, Gearing M, Miao J, Schell M, Sandoval-Espinola WJ, Tao J, Sha B, Graham M, Crooke R, Kleinridders A, Balskus EP, Rey FE, Morris AJ, Biddinger SB. 2019. Trimethylamine N-oxide binds and activates PERK to promote

metabolic dysfunction. *Cell Metabolism* **30(6)**:1141–1151 e1145 DOI 10.1016/j.cmet.2019.08.021.

Chiang YT, Jin T. 2014. p21-Activated protein kinases and their emerging roles in glucose homeostasis. *American Journal of Physiology-Endocrinology and Metabolism* **306(7)**:E707–E722 DOI 10.1152/ajpendo.00506.2013.

Cooke AR, Clark ED. 1976. Effect of first part of duodenum on gastric emptying in dogs: response to acid, fat, glucose, and neural blockade. *Gastroenterology* **70(4)**:550–555 DOI 10.1016/S0016-5085(76)80494-5.

DiGiacomo K, Norris E, Dunshea FR, Hayes BJ, Marett LC, Wales WJ, Leury BJ. 2018. Responses of dairy cows with divergent residual feed intake as calves to metabolic challenges during midlactation and the nonlactating period. *Journal of Dairy Science* **101(7)**:6474–6485 DOI 10.3168/jds.2017-12569.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29(1)**:15–21 DOI 10.1093/bioinformatics/bts635.

Elangeeb ME, Elfaki I, Eleragi AMS, Ahmed EM, Mir R, Alzahrani SM, Bedaiwi RI, Alharbi ZM, Mir MM, Ajmal MR, Tayeb FJ, Barnawi J. 2024. Molecular dynamics simulation of Kir6.2 variants reveals potential association with diabetes mellitus. *Molecules* **29(8)**:1904 DOI 10.3390/molecules29081904.

Ge M, Li C, Zhang Z. 2024. SNP-based and kmer-based eQTL analysis using transcriptome data. *Animals* **14(20)**:2941 DOI 10.3390/ani14202941.

Herd RM, Bishop SC. 2000. Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. *Livestock Production Science* **63(2)**:111–119 DOI 10.1016/S0301-6226(99)00122-0.

Hewitt EW. 2003. The MHC class I antigen presentation pathway: strategies for viral immune evasion. *Immunology* **110(2)**:163–169 DOI 10.1046/j.1365-2567.2003.01738.x.

Higgins MG, Fitzsimons C, McClure MC, McKenna C, Conroy S, Kenny DA, McGee M, Waters SM, Morris DW. 2018. GWAS and eQTL analysis identifies a SNP associated with both residual feed intake and GFRA2 expression in beef cattle. *Scientific Reports* **8**:14301 DOI 10.1038/s41598-018-32374-6.

Hogarth KA, Costford SR, Yoon G, Sondheimer N, Maynes JT. 2018. DNM1L variant alters baseline mitochondrial function and response to stress in a patient with severe neurological dysfunction. *Biochemical Genetics* **56(1–2)**:56–77 DOI 10.1007/s10528-017-9829-2.

Ibragimov E, Pedersen AO, Xiao L, Cirera S, Fredholm M, Karlskov-Mortensen P. 2022. Analysis of merged transcriptomic and genomic datasets to identify genes and pathways underlying residual feed intake in growing pigs. *Scientific Reports* **12**:21946 DOI 10.1038/s41598-022-26496-1.

Iqbal M, Pumford NR, Tang ZX, Lassiter K, Ojano-Dirain C, Wing T, Cooper M, Bottje W. 2005. Compromised liver mitochondrial function and complex activity in low feed efficient broilers are associated with higher oxidative stress and differential protein expression. *Poultry Science* **84(6)**:933–941 DOI 10.1093/ps/84.6.933.

Iqbal M, Pumford NR, Tang ZX, Lassiter K, Wing T, Cooper M, Bottje W. 2004. Low feed efficient broilers within a single genetic line exhibit higher oxidative stress and protein expression in breast muscle with lower mitochondrial complex activity. *Poultry Science* **83(3)**:474–484 DOI 10.1093/ps/83.3.474.

Jehl F, Degalez F, Bernard M, Lecerf F, Lagoutte L, Desert C, Coulee M, Bouchez O, Leroux S, Abasht B, Tixier-Boichard M, Bed'hom B, Burlot T, Gourichon D, Bardou P, Acloque H,

Foissac S, Djebali S, Giuffra E, Zerjal T, Pitel F, Klopp C, Lagarrigue S. 2021. RNA-Seq data for reliable SNP detection and genotype calling: interest for coding variant characterization and cis-regulation analysis by allele-specific expression in livestock species. *Frontiers in Genetics* **12**:655707 DOI 10.3389/fgene.2021.655707.

Jiang B, Liu M, Li P, Zhu Y, Liu Y, Zhu K, Zuo Y, Li Y. 2022. RNA-seq reveals a novel porcine lncRNA MPHOSPH9-OT1 induces CXCL8/IL-8 expression in ETEC infected IPEC-J2 cells. *Frontiers in Cellular and Infection Microbiology* **12**:996841 DOI 10.3389/fcimb.2022.996841.

Kamal MM, Teeya ST, Rahman MM, Talukder MEK, Sarmin S, Wani TA, Hasan MM. 2024. Prediction and assessment of deleterious and disease causing nonsynonymous single nucleotide polymorphisms (nsSNPs) in human FOXP4 gene: an in-silico study. *Heliyon* **10(12)**:e32791 DOI 10.1016/j.heliyon.2024.e32791.

Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME, Genome Aggregation Database C, Neale BM, Daly MJ, MacArthur DG. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581(7809)**:434–443 DOI 10.1038/s41586-020-2308-7.

Kim S, Misra A. 2007. SNP genotyping: technologies and biomedical applications. *Annual Review of Biomedical Engineering* **9(1)**:289–320 DOI 10.1146/annurev.bioeng.9.060906.152037.

Koch RM, Swiger LA, Chambers D, Gregory KE. 1963. Efficiency of feed use in beef cattle. *Journal of Animal Science* **22(2)**:486–494 DOI 10.2527/jas1963.222486x.

Kolath WH, Kerley MS, Golden JW, Keisler DH. 2006. The relationship between mitochondrial function and residual feed intake in Angus steers. *Journal of Animal Science* **84(4)**:861–865 DOI 10.2527/2006.844861x.

Lagarrigue S, Martin L, Hormozdiari F, Roux PF, Pan C, van Nas A, Demeure O, Cantor R, Ghazalpour A, Eskin E, Lusis AJ. 2013. Analysis of allele-specific expression in mouse liver by RNA-Seq: a comparison with Cis-eQTL identified using genetic linkage. *Genetics* **195(3)**:1157–1166 DOI 10.1534/genetics.113.153882.

Lam S, Zeidan J, Miglior F, Suarez-Vega A, Gomez-Redondo I, Fonseca PAS, Guan LL, Waters S, Canovas A. 2020. Development and comparison of RNA-sequencing pipelines for more accurate SNP identification: practical example of functional SNP detection associated with feed efficiency in Nellore beef cattle. *BMC Genomics* **21**:703 DOI 10.1186/s12864-020-07107-7.

Laskar FS, Bappy MNI, Hossain MS, Alam Z, Afrin D, Saha S, Ali Zinnah KM. 2023. An in silico approach towards finding the cancer-causing mutations in human MET gene. *International Journal of Genomics* **2023**:9705159 DOI 10.1155/2023/9705159.

Li J, Sun YBY, Chen W, Fan J, Li S, Qu X, Chen Q, Chen R, Zhu D, Zhang J, Wu Z, Chi H, Crawford S, Oorschot V, Puelles VG, Kerr PG, Ren Y, Nilsson SK, Christian M, Tang H, Chen W, Bertram JF, Nikolic-Paterson DJ, Yu X. 2020. Smad4 promotes diabetic nephropathy by modulating glycolysis and OXPHOS. *EMBO Reports* **21(2)**:e48781 DOI 10.15252/embr.201948781.

Li Z, Zhang Z, Yan P, Huang S, Fei Z, Lin K. 2011. RNA-Seq improves annotation of protein-coding genes in the cucumber genome. *BMC Genomics* **12**:540 DOI 10.1186/1471-2164-12-540.

Lima AOD, Oliveira PSN, Tizioto PC, Somavilla AL, Diniz WJS, Silva JVD, Andrade SCS, Boschiero C, Cesar ASM, Souza MM, Rocha MIP, Afonso J, Buss CE, Mudadu MA, Mourao GB, Coutinho LL, Regitano LCA. 2016. 0318 PRUNE2 gene has a potential effect on residual feed intake in Nellore cattle. *Journal of Animal Science* **94**:152–153 DOI 10.2527/jam2016-0318.

Lingrel JB. 1992. Na,K-ATPase: isoform structure, function, and expression. *Journal of Bioenergetics and Biomembranes* **24(3)**:263–270 DOI 10.1007/BF00768847.

Liu J, Shen Q, Bao H. 2022. Comparison of seven SNP calling pipelines for the next-generation sequencing data of chickens. *PLOS ONE* **17(1)**:e0262574 DOI 10.1371/journal.pone.0262574.

Liu Z, Tan X, Jin Q, Zhan W, Liu G, Cui X, Wang J, Meng X, Zhu R, Wang K. 2024. Multiomics analyses of Jining Grey goat and Boer goat reveal genomic regions associated with fatty acid and amino acid metabolism and muscle development. *Animal Bioscience* **37(6)**:982–992 DOI 10.5713/ab.23.0316.

Lopez-Bellon S, Rodriguez-Lopez S, Gonzalez-Reyes JA, Buron MI, de Cabo R, Villalba JM. 2022. CYB5R3 overexpression preserves skeletal muscle mitochondria and autophagic signaling in aged transgenic mice. *Geroscience* **44(4)**:2223–2241 DOI 10.1007/s11357-022-00574-8.

Lopez-Maestre H, Brinza L, Marchet C, Kielbassa J, Bastien S, Boutigny M, Monnin D, Filali AE, Carareto CM, Vieira C, Picard F, Kremer N, Vavre F, Sagot MF, Lacroix V. 2016. SNP calling from RNA-seq data without a reference genome: identification, quantification, differential analysis and impact on the protein sequence. *Nucleic Acids Research* **44(Suppl 6)**:e148 DOI 10.1093/nar/gkw655.

Luo F, Adetunji MO, Lamont SJ, Abasht B, Schmidt CJ. 2019. Variant analysis pipeline for accurate detection of genomic variants from transcriptome sequencing data. *PLOS ONE* **14(9)**:e0216838 DOI 10.1371/journal.pone.0216838.

Margolis KG, Cryan JF, Mayer EA. 2021. The microbiota-gut-brain axis: from motility to mood. *Gastroenterology* **160(5)**:1486–1501 DOI 10.1053/j.gastro.2020.10.066.

Martinez-Montes AM, Fernandez A, Perez-Montarelo D, Alves E, Benitez RM, Nunez Y, Ovilo C, Ibanez-Escriche N, Folch JM, Fernandez AI. 2017. Using RNA-Seq SNP data to reveal potential causal mutations related to pig production traits and RNA editing. *Animal Genetics* **48(2)**:151–165 DOI 10.1111/age.12507.

Moore SS, Mujibi FD, Sherman EL. 2009. Molecular basis for residual feed intake in beef cattle. *Journal of Animal Science* **87**:E41–E47 DOI 10.2527/jas.2008-1418.

Nakamaru-Ogiso E, Han H, Matsuno-Yagi A, Keinan E, Sinha SC, Yagi T, Ohnishi T. 2010. The ND2 subunit is labeled by a photoaffinity analogue of asimicin, a potent complex I inhibitor. *FEBS Letters* **584(5)**:883–888 DOI 10.1016/j.febslet.2010.01.004.

Nandanpawar P, Sahoo L, Sahoo B, Murmu K, Chaudhari A, Pavan Kumar A, Das P. 2023. Identification of differentially expressed genes and SNPs linked to harvest body weight of genetically improved rohu carp, *Labeo rohita*. *Frontiers in Genetics* **14**:1153911 DOI 10.3389/fgene.2023.1153911.

Ng PC, Henikoff S. 2003. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research* **31(13)**:3812–3814 DOI 10.1093/nar/gkg509.

Nkrumah JD, Okine EK, Mathison GW, Schmid K, Li C, Basarab JA, Price MA, Wang Z, Moore SS. 2006. Relationships of feedlot feed efficiency, performance, and feeding behavior

with metabolic rate, methane production, and energy partitioning in beef cattle. *Journal of Animal Science* **84(1)**:145–153 DOI 10.2527/2006.841145x.

Nolfi-Donegan D, Braganza A, Shiva S. 2020. Mitochondrial electron transport chain: oxidative phosphorylation, oxidant production, and methods of measurement. *Redox Biology* **37(1)**:101674 DOI 10.1016/j.redox.2020.101674.

Olivier M. 2003. A haplotype map of the human genome. *Physiological Genomics* **13(1)**:3–9 DOI 10.1152/physiolgenomics.00178.2002.

Parthiban V, Gromiha MM, Schomburg D. 2006. CUPSAT: prediction of protein stability upon point mutations. *Nucleic Acids Research* **34**:W239–W242 DOI 10.1093/nar/gkl190.

Patience JF, Rossoni-Serao MC, Gutierrez NA. 2015. A review of feed efficiency in swine: biology and application. *Journal of Animal Science and Biotechnology* **6(1)**:33 DOI 10.1186/s40104-015-0031-2.

Patterson RL, van Rossum DB, Kaplin AI, Barrow RK, Snyder SH. 2005. Inositol 1,4,5-trisphosphate receptor/GAPDH complex augments Ca2+ release via locally derived NADH. *Proceedings of the National Academy of Sciences of the United States of America* **102(5)**:1357–1359 DOI 10.1073/pnas.0409657102.

Peng Z, Cheng Y, Tan BC, Kang L, Tian Z, Zhu Y, Zhang W, Liang Y, Hu X, Tan X, Guo J, Dong Z, Liang Y, Bao L, Wang J. 2012. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nature Biotechnology* **30(3)**:253–260 DOI 10.1038/nbt.2122.

Perkins SD, Key CN, Garrett CF, Foradori CD, Bratcher CL, Kriese-Anderson LA, Brandebourg TD. 2014. Residual feed intake studies in Angus-sired cattle reveal a potential role for hypothalamic gene expression in regulating feed efficiency. *Journal of Animal Science* **92(2)**:549–560 DOI 10.2527/jas.2013-7019.

Quinn EM, Cormican P, Kenny EM, Hill M, Anney R, Gill M, Corvin AP, Morris DW. 2013. Development of strategies for SNP detection in RNA-seq data: application to lymphoblastoid cell lines and evaluation using 1000 Genomes data. *PLOS ONE* **8(3)**:e58815 DOI 10.1371/journal.pone.0058815.

Raizada A, Souframanien J. 2019. Transcriptome sequencing, de novo assembly, characterisation of wild accession of blackgram (*Vigna mungo* var. *silvestris*) as a rich resource for development of molecular markers and validation of SNPs by high resolution melting (HRM) analysis. *BMC Plant Biology* **19(1)**:358 DOI 10.1186/s12870-019-1954-0.

R Core Team. 2023. *R: a language and environment for statistical computing*. Version 4.3.0. Vienna: R Foundation for Statistical Computing. *Available at https://www.r-project.org*.

Reeves PG, Chaney RL. 2004. Marginal nutritional status of zinc, iron, and calcium increases cadmium retention in the duodenum and other organs of rats fed rice-based diets. *Environmental Research* **96(3)**:311–322 DOI 10.1016/j.envres.2004.02.013.

Richardson EC, Herd RM. 2004. Biological basis for variation in residual feed intake in beef cattle. 2. Synthesis of results following divergent selection. *Australian Journal of Experimental Agriculture* **44(5)**:431 DOI 10.1071/ea02221.

Sakuta H, Lin CH, Hiyama TY, Matsuda T, Yamaguchi K, Shigenobu S, Kobayashi K, Noda M. 2020. SLC9A4 in the organum vasculosum of the lamina terminalis is a [Na(+)] sensor for the control of water intake. *Pflugers Archiv* **472(5)**:609–624 DOI 10.1007/s00424-020-02389-y.

Santos CA, Andrade SCS, Freitas PD. 2018. Identification of SNPs potentially related to immune responses and growth performance in *Litopenaeus vannamei* by RNA-seq analyses. *PeerJ* **6(3–4)**:e5154 DOI 10.7717/peerj.5154.

Sartin JL, Whitlock BK, Daniel JA. 2011. Triennial Growth Symposium: neural regulation of feed intake: modification by hormones, fasting, and disease. *Journal of Animal Science* **89(7)**:1991–2003 DOI 10.2527/jas.2010-3399.

Sims D, Sudbery I, Ilott NE, Heger A, Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nature Reviews Genetics* **15(2)**:121–132 DOI 10.1038/nrg3642.

Song C, Huang Y, Yang Z, Ma Y, Chaogetu B, Zhuoma Z, Chen H. 2019. RNA-seq analysis identifies differentially expressed genes insubcutaneous adipose tissuein Qaidamford cattle, cattle-yak, and angus cattle. *Animals* **9**:1077 DOI 10.3390/ani9121077.

Sun YX, Tsuboi K, Okamoto Y, Tonai T, Murakami M, Kudo I, Ueda N. 2004. Biosynthesis of anandamide and N-palmitoylethanolamine by sequential actions of phospholipase A2 and lysophospholipase D. *Biochemical Journal* **380(3)**:749–756 DOI 10.1042/BJ20040031.

Tanaka T, Ishikawa G, Ogiso-Tanaka E, Yanagisawa T, Sato K. 2019. Development of genome-wide SNP markers for barley via reference-based RNA-seq analysis. *Frontiers in Plant Science* **10**:577 DOI 10.3389/fpls.2019.00577.

Tang X, Baheti S, Shameer K, Thompson KJ, Wills Q, Niu N, Holcomb IN, Boutet SC, Ramakrishnan R, Kachergus JM, Kocher J-PA, Weinshilboum RM, Wang L, Thompson EA, Kalari KR. 2014. The eSNV-detect: a computational system to identify expressed single nucleotide variants from transcriptome sequencing data. *Nucleic Acids Research* **42(22)**:e172 DOI 10.1093/nar/gku1005.

Tang H, Thomas PD. 2016. PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics* **32(14)**:2230–2232 DOI 10.1093/bioinformatics/btw222.

The 1000 Genomes Project Consortium. 2015. A global reference for human genetic variation. *Nature* **526(7571)**:68–74 DOI 10.1038/nature15393.

UniProt C. 2023. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Research* **51(D1)**:D523–D531 DOI 10.1093/nar/gkac1052.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, Banks E, Garimella KV, Altshuler D, Gabriel S, DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Current Protocols in Bioinformatics* **43(1)**:11.10.11–11.10.33 DOI 10.1002/0471250953.bi1110s43.

Van Deventer R, Rhode C, Marx M, Roodt-Wilding R. 2020. The development of genome-wide single nucleotide polymorphisms in blue wildebeest using the DArTseq platform. *Genomics* **112(5)**:3455–3464 DOI 10.1016/j.ygeno.2020.04.032.

Xiao W, Wang RS, Handy DE, Loscalzo J. 2018. NAD(H) and NADP(H) redox couples and cellular energy metabolism. *Antioxidants & Redox Signaling* **28(3)**:251–272 DOI 10.1089/ars.2017.7216.

Yang C, Ding Y, Dan X, Shi Y, Kang X. 2023. Multi-transcriptomics reveals RLMF axis-mediated signaling molecules associated with bovine feed efficiency. *Frontiers in Veterinary Science* **10**:1090517 DOI 10.3389/fvets.2023.1090517.

Yang C, Han L, Li P, Ding Y, Zhu Y, Huang Z, Dan X, Shi Y, Kang X. 2021. Characterization and duodenal transcriptome analysis of Chinese beef cattle with divergent feed efficiency using RNA-Seq. *Frontiers in Genetics* **12**:741878 DOI 10.3389/fgene.2021.741878.

Yang C, Zhu Y, Ding Y, Huang Z, Dan X, Shi Y, Kang X. 2022. Identifying the key genes and functional enrichment pathways associated with feed efficiency in cattle. *Gene* **807(3)**:145934 DOI 10.1016/j.gene.2021.145934.

Yao Y, Liu S, Xia C, Gao Y, Pan Z, Canela-Xandri O, Khamseh A, Rawlik K, Wang S, Li B, Zhang Y, Pairo-Castineira E, D'Mellow K, Li X, Yan Z, Li CJ, Yu Y, Zhang S, Ma L, Cole JB, Ross PJ, Zhou H, Haley C, Liu GE, Fang L, Tenesa A. 2022. Comparative transcriptome in large-scale human and cattle populations. *Genome Biology* **23(1)**:176 DOI 10.1186/s13059-022-02745-4.

Ying W. 2006. NAD+ and NADH in cellular functions and cell death. *Frontiers in Bioscience* **11(1)**:3129–3148 DOI 10.2741/2038.

Ying W. 2007. NAD+ and NADH in brain functions, brain diseases and brain aging. *Frontiers in Bioscience* **12(1)**:1863–1888 DOI 10.2741/2194.

Ying W. 2008. NAD+/NADH and NADP+/NADPH in cellular functions and cell death: regulation and biological consequences. *Antioxidants & Redox Signaling* **10(2)**:179–206 DOI 10.1089/ars.2007.1672.

Zhang Y, O'Keefe RJ, Jonason JH. 2017. BMP-TAK1 (MAP3K7) induces adipocyte differentiation through PPARgamma signaling. *Journal of Cellular Biochemistry* **118(1)**:204–210 DOI 10.1002/jcb.25626.

Zhao K, Chen YH, Penner GB, Oba M, Guan LL. 2017. Transcriptome analysis of ruminal epithelia revealed potential regulatory mechanisms involved in host adaptation to gradual high fermentable dietary transition in beef cattle. *BMC Genomics* **18**:976 DOI 10.1186/s12864-017-4317-y.

Zhao L, Ding Y, Yang C, Wang P, Zhao Z, Ma Y, Shi Y, Kang X. 2023. Identification and characterization of hypothalamic circular RNAs associated with bovine residual feed intake. *Gene* **851(5)**:147017 DOI 10.1016/j.gene.2022.147017.

Zhao Z, Fu YX, Hewett-Emmett D, Boerwinkle E. 2003. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* **312**:207–213 DOI 10.1016/s0378-1119(03)00670-x.

Zhou X, Ma Y, Yang C, Zhao Z, Ding Y, Zhang Y, Wang P, Zhao L, Li C, Su Z, Wang X, Ming W, Zeng L, Kang X. 2023. Rumen and fecal microbiota characteristics of Qinchuan cattle with divergent residual feed intake. *Microorganisms* **11**:358 DOI 10.3390/microorganisms11020358.