Identification of SNPs potentially associated with residual feed intake in 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 750021, China;
[2] Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service, United States Department of Agriculture, Beltsville, MD 20705, USA

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

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. 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.

Keywords: RNA-Seq; single nucleotide polymorphisms; beef cattle; hypothalamus; duodenum

1. Introduction

Feed cost encompasses more than 70% of the total input cost in cattle production, making feed utilization a crucial metric for evaluating production expenditures (Patience et al. 2015). Efficient feed utilization can reduce herd maintenance costs by 9-10%, lower feed intake by 10-12%, and mitigate methane emissions by 15-20% (Moore et al. 2009). Consequently, optimizing feed utilization and minimizing production costs are essential for livestock development. Residual feed intake represents the disparity between the average daily feed intake (ADFI) and the average expected feed intake (AEFI) required to maintain production levels(Koch et al. 1963). RFI enables a precise assessment of feed utilization efficiency in livestock, distinctly isolating the impact of animal growth traits and rates(Richardson & Herd 2004).

Moreover, RFI has emerged as a potential candidate for genetic improvement owing to its moderate heritability (0.28-0.58) (Moore et al. 2009) and its significant genetic variability within and between species(Archer & Bergh 2000; Herd & Bishop 2000) . In our previous research, we found that RFI is related to multiple factors, including gut microbiota(Zhou et al. 2023), circRNA-miRNA interaction(Zhao et al. 2023), and expressed genes at the transcriptional level (Yang et al. 2023; Yang et al. 2021; Yang et al. 2022). These explorations provide significant support for a comprehensive analysis of the RFI phenotype and show that the factors influencing the RFI phenotype of cattle are numerous and complex.

The hypothalamus and duodenum are critical organs in animal feed intake, energy metabolism, and digestion. Previous studies have highlighted the role of the hypothalamic arcuate nucleus in appetite regulation, wherein neuropeptide Y (NPY) and agouti-related peptide (AGRP) promote animal feeding. In contrast, α-MSH (α-Melanocyte-stimulating hormone), an anorexigenic neuropeptide, induces satiety (Perkins et al. 2014). The duodenum, a significant organ involved in 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, and gut microbes play a crucial role in host carbohydrate, amino acid, lipid metabolism, and other nutrient metabolic processes (Genomes Project et al. 2015; Olivier 2003). Thus, the close association of the hypothalamus and duodenum with feeding efficiency underscores their significance as focal points in studies investigating RFI in beef cattle.

Single nucleotide polymorphism (SNP) is a form of genetic variation that occurs at the genomic level, resulting from a single nucleotide variant in the DNA sequence. These variants can potentially impact gene expression and function, consequently influencing individual phenotypes and disease susceptibilities (Kim & Misra 2007). Due to the low cost of RNA-seq data and high availability in various databases, explore coding region variants from RNA-Seq data 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). Therefore, multiple studies have shown that exploring SNPs in RNA-Seq data is a very cost-effective method. A wealth of research has been dedicated to extracting single nucleotide

Dikomentari [f1]: why efficient feed can reduce methane ?

Dikomentari [f2]: What other research might be added here?

polymorphisms 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 been pivotal in identifying SNPs potentially linked to the immune response and the growth performance of Penaeus vannamei (Santos et al. 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). Additionally, in animal husbandry, RNA-Seq SNP data has been instrumental in revealing potential causal mutations that are relevant to pig production traits and shedding light on the intricacies of RNA editing (Martinez-Montes 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.

2. Materials and Methods

1.1 Experimental animals and data collection

Based on our previous study (Yang et al. 2021), 30 Qinchuan bulls with similar age (15±1 mouths) 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) (Yang et al. 2021).

1.2 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-hour 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. Twenty tissue samples were washed with PBS, cut into blocks, and placed into sterile, enzyme-free freezing tubes for storage in liquid nitrogen. Total RNA was extracted from 500 mg of tissue samples using TRIzol method (TaKaRa Bio, China), following the manufacturer's instructions. 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 $\geqslant$ 500 ng/μL, 28S:18S > 1.0, and RIN $\geqslant$ 7. Library construction was performed, and the library's initial quantification was carried out using Qubit 2.0. The

Dikomentari [f3]: for cattle, is there any other research besides pigs?

Dikomentari [f4]: Why are 30 cattle samples used?

Dikomentari [f5]: standard feed pattern? what is the nutrient content?

library with an insert size was verified using an Agilent 2100, and the effective concentration of the library (effective concentration > 2 nM) was accurately determined using qRT-PCR. Finally, pair-end sequencing data (raw data) with 150bp 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.

1.3 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 quality of raw data was assessed using the fastQC software (version 0.11.9). Subsequently, the Trimmomatic software (version 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 5bp to remove bases with an average Phred score below 20, and discarding reads shorter than 75bp. 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 software (version 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 software (version 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.

1.4 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 software (version 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 (Figure 1).

1.5 SNPs recognition, filtering and annotation

BCFtools (version 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 (version 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 (version 5.1d) software with the built-in ARS-UCD1.2.105 database. The thresholds of above software are referenced from previous study(Lam et al. 2020).

1.6 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 (version 5.1d) software 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.

1.7 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 pvalue < 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.

1.8 Statistical analysis

The data from the experiment were analyzed and visualized using R (version 4.3.0) and Prism (version 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.

3. Results

3.1 RNA-seq sequencing data quality and comparisons

From the RNA-seq data of the hypothalamus and duodenum, we obtained a total of 1019 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 (Figure 2). These results demonstrate the high quality of the obtained data, reducing the impact of sequencing errors on subsequent analysis.

3.2 SNPs screening and analysis

Based on the results of mapping 20 samples to the bovine reference genome, SNPs in the high and low RFI groups were identified using a method that combines RFI phenotype and 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 (Figure 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 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%.

3.3 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 (Figure 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 (Figure 4B). The chromosomal distribution of SNPs within the high and low RFI groups is delineated in figure 4C and figure 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, un-translation 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 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 (Figure 4E). SNPs located in intergenic regions may be found in new genes or gene portions that have not been annotated yet.

3.4 Influence prediction and amino acid change

To understand the potential effects of SNPs mutations on codons, the genetic effects of SNPs mutations were evaluated, the results indicated 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 (Figure 5A, 5B; Table 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. To analyze the potential impact of SNPs on genes and proteins, the effect of intergroup-specific SNPs (Table 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 (Figure 5C, 5D). By identifying SNP loci that significantly impacted both the high and low RFI groups and 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).

Dikomentari [f6]: add supporting references here?

3.5 Genes function annotation of high-impact SNP loci

GO functional annotation and enrichment analysis were conducted for the aforementioned genes (Figure 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; 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 et al. 2004; Patterson et al. 2005; Xiao et al. 2018; Ying 2006; Ying 2007; Ying 2008). Particularly, 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 (Figure 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).

3.6 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 (Figure 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 et al. 2017).

3.7 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 (Table 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.

4. Discussion

Improving animal feed utilization and reducing production costs have always been the primary focus of livestock production, as efficient feed utilization is crucial for the growth

and development of animals. In addition to the development of high-quality feeds through breeding selection, exploring the genetic factors that affect feed efficiency in beef cattle is an important research direction for breeders. The hypothalamus serves as a central control center for feeding regulation, where it interacts with groups of neurons to produce signals that stimulate or suppress appetite, ultimately influencing food intake (Perkins et al. 2014; Sartin et al. 2011). Meanwhile, the duodenum, as the initial segment of the small intestine, plays a vital role in the digestion and absorption of nutrients, specifically carbohydrates and micronutrients(Cooke & Clark 1976; Reeves & Chaney 2004). The brain-gut axis, composed of the central nervous system, enteric nervous system, and autonomic nervous system, facilitates the complex communication between the gut and the brain through neurohumoral pathways (Margolis et al. 2021). Axes such as the hypothalamic-pituitary-adrenal (HPA) axis, which is part of the brain-gut axis, have been identified as important factors contributing to the variability of residual feed intake (DiGiacomo et al. 2018).

RNA-Seq, as a second-generation transcriptome sequencing method, provides a diverse range of research 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 methods collectively unravel the intricate complexities of organisms. Residual feed intake, a quantitative trait, is influenced by multiple genes and shaped by various physiological metabolic processes (Arthur et al. 2001). Previous studies have successfully identified genes associated with feed utilization efficiency through SNP screening. For instance, Marc et al. demonstrated a strong association between the variant rs43555985 and RFI (P=8.28E-06) (Higgins et al. 2018). Bolormaa et al. identified 111 and 75 significantly associated SNPs with RFI ($p < 0.001$) using the 10K and 50K SNP microarray data, respectively(Bolormaa et al. 2011). Lima et al. employed comprehensive GWA, AWM, and RNA-Seq analyses to identify the *PRUNE2* gene as a potential candidate affecting feed efficiency (Lima et al. 2016). Several tools have also 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 et al. 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. By employing high-throughput transcriptome sequencing, we obtained data from 20 samples and merged the tissue data based on RFI groups for SNPs identification, aiming to enhance the accuracy of SNP functional annotation.

SNPs are DNA sequence polymorphisms caused by single nucleotide variants. In this study, the obtained SNPs data from 20 samples showed that there were 270,410 specific SNPs in the high RFI group and 255,120 specific SNPs in the low RFI group. Among them, the high RFI group had 11,991 pure homozygous SNPs, while the low RFI group had 14,007 SNPs. More than 70% of SNPs in both groups were located in the intron region, followed by the intergenic region, due to unspliced transcripts (premature transcripts) and unannotated regions (Jehl et al. 2021). Only 1.11% and 1.38% of SNPs in high and low RFI groups were located in the intron region. This distribution is expected since intron regions are generally subject to greater selection pressure than non-coding regions (Zhao et al. 2003). Additionally,

the analysis of transition and transversion types revealed a transition-to-transversion ratio (Ts/Tv) of 2.55 in the high RFI group and 2.50 in the low RFI group, consistent with previous studies reporting a higher frequency of transition mutations compared to transversions (Nandanpawar et al. 2023; Raizada & Souframanien 2019; van Deventer et al. 2020). The relatively consistent Ts/Tv values further confirm the accuracy of SNP identification in this study (Arabnejad et al. 2018). Based on SNPs annotations, functionally significant SNPs specific to the high and low RFI groups were selected using SnpEff software, and the respective genes harboring these SNPs were identified. A total of 83 genes were found in the high RFI group, while 97 genes were identified in the low RFI group. The corresponding code scripts can be found in Script S1.

GO and KEGG pathway enrichment analyses were performed separately for the two gene sets. At the molecular function level, the enriched GO terms in both groups were primarily associated with protein binding and enzyme binding. Additionally, in the low RFI group, specific enrichment was observed for terms related to NADH, such as 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, is the first enzyme involved in mitochondrial oxidative phosphorylation and is often referred to as the "entry enzyme". It catalyzes the transfer of electrons from NADH to coenzyme Q and plays a key role in energy metabolism within the mitochondrial inner membrane (Nakamaru-Ogiso et al. 2010). Genes *ND4*, *ND5*, and *ND6* in the related pathway are core subunits of mitochondrial respiratory chain NADH dehydrogenase (complex I) and are critical for its catalytic activity and assembly (UniProt 2023). At the same time, previous genomic analyses have suggested that the *ND (2,3,4,4L,5,6)* gene cluster may be the major effect genes causing changes in feed efficiency (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 et al. 2020). Animals with higher feed efficiency tend to exhibit lower oxidative stress phenomena (Bottje & Carstens 2009; Iqbal et al. 2005; Iqbal et al. 2004). KEGG pathway enrichment analysis also identified several pathways associated with energy metabolism. In the high RFI group, pathways such as thyroid hormone synthesis, pancreatic secretion, gastric acid secretion, and cAMP signaling pathway were enriched. The enriched pathways in the low RFI group included thermogenesis, parathyroid hormone synthesis, secretion and action,

Dikomentari [f7]: Explain in detail how an efficient feed indicates lower oxidative stress?

pancreatic secretion, triglyceride metabolism, and alanine metabolism. One of the enriched genes, *ATP1A2*, plays a role in ATP hydrolysis and facilitates the exchange of sodium and potassium ions across the plasma membrane, creating a sodium-potassium ion electrochemical gradient for the active transport of various nutrients (Lingrel 1992). Additionally, *SLC9A4* has been found to 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, which are involved in the regulation of 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* have been linked to 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* has been implicated in the regulation of glucose uptake (Chiang & Jin 2014). *SMAD4* has been associated with aerobic glycolysis and obesity (Li et al. 2020). *MAP3K7* can induce adipocyte differentiation through *PPARγ* signaling (Zhang et al. 2017). *PGM2L1* has been 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* acts in synergy with coenzyme Q, participating in the cross-membrane redox system to protect cells against oxidative stress (Lopez-Bellon et al. 2022). These genes will serve as target genes in subsequent cellular and molecular experiments to validate their associations with the RFI trait.

RFI, a key economic trait in feed efficiency research, requires a deep understanding of the genetic mechanisms associated with SNP loci and their impact on RFI regulation in beef cattle. This knowledge 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.

Although we have obtained a large number of 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.

Dikomentari [f8]: Can be improved by other references or research.

5. 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.

Acknowledgments

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.

References

Anand P, Kumar SV, Ravi K, Simmi TJTAH, and Production. 2021. Differential gene expression in duodenum of colored broiler chicken divergently selected for residual feed intake. 53:1-10.

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

Archer JA, and 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* 65:47-55. 10.1016/s0301-6226(99)00181-5

Arthur PF, Archer JA, Johnston DJ, Herd RM, Richardson EC, and Parnell PF. 2001. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. *J Anim Sci* 79:2805-2811. 10.2527/2001.79112805x

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

Berger F, Ramirez-Hernandez MH, and Ziegler M. 2004. The new life of a centenarian: signalling functions of NAD(P). *Trends Biochem Sci* 29:111-118. 10.1016/j.tibs.2004.01.007

Bolormaa S, Hayes BJ, Savin K, Hawken R, Barendse W, Arthur PF, Herd RM, and Goddard ME. 2011. Genome-wide association studies for feedlot and growth traits in cattle. *J Anim Sci* 89:1684-1697. 10.2527/jas.2010-3079

Bottje WG, and Carstens GE. 2009. Association of mitochondrial function and feed efficiency in poultry and livestock species. *J Anim Sci* 87:E48-63. 10.2527/jas.2008-1379

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

Charoensook R, Gatphayak K, Sharifi AR, Chaisongkram C, Brenig B, and Knorr C. 2012. Polymorphisms in the bovine HSP90AB1 gene are associated with heat tolerance in Thai indigenous cattle. *Trop Anim Health Prod* 44:921-928. 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, and Biddinger SB. 2019. Trimethylamine N-Oxide Binds and Activates PERK to Promote Metabolic Dysfunction. *Cell Metab* 30:1141-1151 e1145. 10.1016/j.cmet.2019.08.021

Chiang YT, and Jin T. 2014. p21-Activated protein kinases and their emerging roles in glucose homeostasis. *Am J Physiol Endocrinol Metab* 306:E707-722. 10.1152/ajpendo.00506.2013

Cooke AR, and Clark ED. 1976. Effect of First Part of Duodenum on Gastric Emptying in Dogs: Response to Acid, Fat, Glucose, and Neural Blockade. *Gastroenterology* 70:550-555. 10.1016/s0016-5085(76)80494-5

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

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

Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, and Abecasis GR. 2015. A global reference for human genetic variation. *Nature* 526:68-74. 10.1038/nature15393

Herd RM, and Bishop SC. 2000. Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. *Livestock Production Science* 63:111-119. 10.1016/s0301-6226(99)00122-0

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

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

Hogarth KA, Costford SR, Yoon G, Sondheimer N, and Maynes JT. 2018. DNM1L Variant Alters Baseline Mitochondrial Function and Response to Stress in a Patient with Severe Neurological Dysfunction. *Biochem Genet* 56:56-77. 10.1007/s10528-017-9829-2

Iqbal M, Pumford NR, Tang ZX, Lassiter K, Ojano-Dirain C, Wing T, Cooper M, and 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. *Poult Sci* 84:933-941. 10.1093/ps/84.6.933

Iqbal M, Pumford NR, Tang ZX, Lassiter K, Wing T, Cooper M, and 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. *Poult Sci* 83:474-484. 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, and 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. *Front Genet* 12:655707. 10.3389/fgene.2021.655707

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

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, and MacArthur DG. 2020. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581:434-443. 10.1038/s41586-020-2308-7

Kim S, and Misra A. 2007. SNP genotyping: technologies and biomedical applications. *Annu Rev Biomed Eng* 9:289-320. 10.1146/annurev.bioeng.9.060906.152037

Koch RM, Swiger LA, Chambers D, and Gregory KE. 1963. Efficiency of Feed Use in Beef Cattle. *Journal of Animal Science* 22:486-494. 10.2527/jas1963.222486x

Kolath WH, Kerley MS, Golden JW, and Keisler DH. 2006. The relationship between mitochondrial function and residual feed intake in Angus steers. *J Anim Sci* 84:861-865. 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, and Lusis AJ. 2013. Analysis of allele-specific expression in mouse liver by RNA-Seq: a comparison with Cis-eQTL identified using genetic linkage. *Genetics* 195:1157-1166. 10.1534/genetics.113.153882

Lam S, Zeidan J, Miglior F, Suarez-Vega A, Gomez-Redondo I, Fonseca PAS, Guan LL, Waters S, and 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. 10.1186/s12864-020-07107-7

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, and Yu X. 2020. Smad4 promotes diabetic nephropathy by modulating glycolysis and OXPHOS. *EMBO Rep* 21:e48781. 10.15252/embr.201948781

Li Z, Zhang Z, Yan P, Huang S, Fei Z, and Lin K. 2011. RNA-Seq improves annotation of protein-coding genes in the cucumber genome. *BMC Genomics* 12:540. 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, and Regitano LCA. 2016. 0318 PRUNE2 gene has a potential effect on residual feed intake in Nellore cattle. *Journal of Animal Science* 94:152-153. 10.2527/jam2016-0318

Lingrel JB. 1992. Na,K-ATPase: isoform structure, function, and expression. *J Bioenerg Biomembr* 24:263-270. 10.1007/BF00768847

Liu J, Shen Q, and Bao H. 2022. Comparison of seven SNP calling pipelines for the next-generation sequencing data of chickens. *PLoS One* 17:e0262574. 10.1371/journal.pone.0262574

Liu Z, Tan X, Jin Q, Zhan W, Liu G, Cui X, Wang J, Meng X, Zhu R, and 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. *Anim Biosci* 37:982-992. 10.5713/ab.23.0316

Lopez-Bellon S, Rodriguez-Lopez S, Gonzalez-Reyes JA, Buron MI, de Cabo R, and Villalba JM. 2022. CYB5R3 overexpression preserves skeletal muscle mitochondria and autophagic signaling in aged transgenic mice. *Geroscience* 44:2223-2241. 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, and 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 Res* 44:e148. 10.1093/nar/gkw655

Luo F, Adetunji MO, Lamont SJ, Abasht B, and Schmidt CJ. 2019. Variant analysis pipeline for accurate detection of genomic variants from transcriptome sequencing data. *Plos One* 14. 10.1371/journal.pone.0216838

Margolis KG, Cryan JF, and Mayer EA. 2021. The Microbiota-Gut-Brain Axis: From Motility to Mood. *Gastroenterology* 160:1486-1501. 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, and Fernandez AI. 2017. Using RNA-Seq SNP data to reveal potential causal mutations related to pig production traits and RNA editing. *Anim Genet* 48:151-165. 10.1111/age.12507

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

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

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

Nkrumah JD, Okine EK, Mathison GW, Schmid K, Li C, Basarab JA, Price MA, Wang Z, and Moore SS. 2006. Relationships of feedlot feed efficiency, performance, and feeding behavior with metabolic rate, methane production, and energy partitioning in beef cattle. *J Anim Sci* 84:145-153. 10.2527/2006.841145x

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

Olivier M. 2003. A haplotype map of the human genome. *Physiol Genomics* 13:3-9. 10.1152/physiolgenomics.00178.2002

Patience JF, Rossoni-Serao MC, and Gutierrez NA. 2015. A review of feed efficiency in swine: biology and application. *J Anim Sci Biotechnol* 6:33. 10.1186/s40104-015-0031-2

Patterson RL, van Rossum DB, Kaplin AI, Barrow RK, and Snyder SH. 2005. Inositol 1,4,5-trisphosphate receptor/GAPDH complex augments Ca2+ release via locally derived NADH. *Proc Natl Acad Sci U S A* 102:1357-1359. 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, and Wang J. 2012. Comprehensive analysis of RNA-Seq data reveals extensive RNA editing in a human transcriptome. *Nat Biotechnol* 30:253-260. 10.1038/nbt.2122

Perkins SD, Key CN, Garrett CF, Foradori CD, Bratcher CL, Kriese-Anderson LA, and Brandebourg TD. 2014. Residual feed intake studies in Angus-sired cattle reveal a potential role for hypothalamic gene expression in regulating feed efficiency. *J Anim Sci* 92:549-560. 10.2527/jas.2013-7019

Quinn EM, Cormican P, Kenny EM, Hill M, Anney R, Gill M, Corvin AP, and 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:e58815. 10.1371/journal.pone.0058815

Raizada A, and 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 Biol* 19:358. 10.1186/s12870-019-1954-0

Reeves PG, and 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. *Environ Res* 96:311-322. 10.1016/j.envres.2004.02.013

Richardson EC, and 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. 10.1071/ea02221

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

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

Sartin JL, Whitlock BK, and Daniel JA. 2011. Triennial Growth Symposium: neural regulation of feed intake: modification by hormones, fasting, and disease. *J Anim Sci* 89:1991-2003. 10.2527/jas.2010-3399

Sims D, Sudbery I, Ilott NE, Heger A, and Ponting CP. 2014. Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* 15:121-132. 10.1038/nrg3642

Song C, Huang Y, Yang Z, Ma Y, Chaogetu B, Zhuoma Z, and Chen H. 2019. RNA-Seq Analysis Identifies Differentially Expressed Genes Insubcutaneous Adipose Tissuein Qaidamford Cattle, Cattle-Yak, and Angus Cattle. *Animals (Basel)* 9. 10.3390/ani9121077

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

Tanaka T, Ishikawa G, Ogiso-Tanaka E, Yanagisawa T, and Sato K. 2019. Development of Genome-Wide SNP Markers for Barley via Reference- Based RNA-Seq Analysis. *Front Plant Sci* 10:577. 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, and Kalari KR. 2014. The eSNV-detect: a computational system to identify expressed single nucleotide variants from transcriptome sequencing data. *Nucleic Acids Research* 42:e172-e172. 10.1093/nar/gku1005

UniProt C. 2023. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res* 51:D523-D531. 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, and DePristo MA. 2013. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43:11 10 11-11 10 33. 10.1002/0471250953.bi1110s43

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

Xiao W, Wang RS, Handy DE, and Loscalzo J. 2018. NAD(H) and NADP(H) Redox Couples and Cellular Energy Metabolism. *Antioxid Redox Signal* 28:251-272. 10.1089/ars.2017.7216

Yang C, Ding Y, Dan X, Shi Y, and Kang X. 2023. Multi-transcriptomics reveals RLMF axis-mediated signaling molecules associated with bovine feed efficiency. *Front Vet Sci* 10:1090517. 10.3389/fvets.2023.1090517

Yang C, Han L, Li P, Ding Y, Zhu Y, Huang Z, Dan X, Shi Y, and Kang X. 2021. Characterization and Duodenal Transcriptome Analysis of Chinese Beef Cattle With Divergent Feed Efficiency Using RNA-Seq. *Front Genet* 12:741878. 10.3389/fgene.2021.741878

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

Ying W. 2006. NAD+ and NADH in cellular functions and cell death. *Front Biosci* 11:3129-3148. 10.2741/2038

Ying W. 2007. NAD+ and NADH in brain functions, brain diseases and brain aging. *Front Biosci* 12:1863-1888. 10.2741/2194

Ying W. 2008. NAD+/NADH and NADP+/NADPH in cellular functions and cell death: regulation and biological consequences. *Antioxid Redox Signal* 10:179-206. 10.1089/ars.2007.1672

Zhang Y, O'Keefe RJ, and Jonason JH. 2017. BMP-TAK1 (MAP3K7) Induces Adipocyte Differentiation Through PPARgamma Signaling. *J Cell Biochem* 118:204-210. 10.1002/jcb.25626

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

Zhao Z, Fu YX, Hewett-Emmett D, and Boerwinkle E. 2003. Investigating single nucleotide polymorphism (SNP) density in the human genome and its implications for molecular evolution. *Gene* 312:207-213. 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, and Kang X. 2023. Rumen and Fecal Microbiota Characteristics of Qinchuan Cattle with Divergent Residual Feed Intake. *Microorganisms* 11. 10.3390/microorganisms11020358