JABG
Journal of Animal Breeding & Genomics

JABG. 2023 June, 7(2)25-35
https://doi.org/10.12972/jabng.20230004

Print ISSN 1226-5543
Online ISSN 2586-4297

Research Article

Identification of SNPs and microRNAs in the massively whole-genome sequencing data of Korean native cattle

Hong-Sik Kong[1,2], Yoonseok Lee[1,2,*]

[1] School of Biotechnology, Hankyong National University, Gyeonggi-do, Anseong, 17579, Republic of Korea

[2] Center for Genetic Information, Hankyong National University, Gyeonggi-do, Anseong, 17579, Republic of Korea

*Corresponding author : Yoonseok Lee, Tel: 031-670-5332, Fax: 031-670-5339, E-mail: yoonseok95@hknu.ac.kr

Abstract

This study aims to enhance the accuracy of genome selection for economic traits in Korean native cattle through the identification of single nucleotide polymorphisms (SNPs) located in microRNA sequences associated with lipid metabolism. We assembled and mapped the Hanwoo genome using whole-genome sequencing data from the NCBI's Sequence Read Archive and the National Institute of Animal Science. This allowed us to discover a total of 25,982,458 variants (21,795,577 SNPs and 4,186,881 Indels). Annotation revealed that a significant number of these variants were located in intergenic and intronic regions, while others were found to alter the DNA, creating amino acid substitutions, premature stop codons, or silent mutations. A high proportion of variants were observed in the protein coding regions and miRNA, with 24 and 74,205 variants identified in Impact_LOW and Impact_MODIFIER, respectively. In addition, we identified 18 putative miRNAs that were differentially expressed in subcutaneous and intramuscular fat tissues. Especially, bta-mir-143 and bta-mir-145 were highly expressed in intramuscular fat tissue. These results could support the impact on the potential regulatory roles of miRNAs in bovine adipogenesis, and could prove beneficial for improving the accuracy of Hanwoo breeding programs.

Keywords: Korean native cattle, MicroRNA, Single nucleotide polymorphism, Whole-genome sequence, Lipid metabolism

Introduction

The selection of Hanwoo genome for economic traits has been evaluated through traditional breeding program using whole-genome SNP 50K panel (Lee et al., 2015). Genome selection usually predicts individual genetic merits to increase performance of economic traits, and improving accuracy of this selection model is an important factor (Lopez et al., 2019; Xiang et al., 2021). Recently, several studies have developed customized Hanwoo 50K SNP panels to increase the accuracy of Hanwoo genome selection. As a results, the accuracy of genome selection was found to increase for backfat thickness, Longissimus muscle area, carcass weight and marbling score in Hanwoo steer populations.

Furthermore, Lee et al (2019) reported an increase in the accuracy of genome selection for economic traits by using high-density SNP panels containing causal SNP located in linkage disequilibrium (Lopez et al., 2019; Lopez et al., 2021). Improvement in the accuracy of the breeding program has been reported, especially with the use of bovine 50K SNP panels containing causal SNPs of myostatin (Lee et al., 2019). Although

Received May 17, 2023 **Revised** May 17, 2023 **Accepted** June 26, 2023

causal genetic variation has a direct role in phenotype variation, it is necessary to consider the combined effects of multiple causal SNPs because each individual SNP typically has only a small effect. Additionally, there are limitations in understanding the mechanisms related to lipid synthesis in cattle due to the small effect of causal genetic variations. Furthermore, the expression genetic variation of the causal gene plays an important role in regulating lipid synthesis and can increase the accuracy of genome selection. Therefore, to improve the accuracy of prediction in livestock breeding, a large amount of genomic information that can directly affect protein variations is required, in addition to the bovine 50K SNP panel.

Noncoding RNA (ncRNA) plays an important role in regulating gene expression (Ibanez-Escriche and González-Recio, 2011; VanRaden, 2008). MicroRNA (miRNAs) are a class of endogenous small non-coding RNAs, approximately 22 nucleotides in length, that play a role in post-transcriptional regulation by targeting mRNA degradation or inhibiting their translation (Ha and Kim, 2014). Furthermore, miRNAs, which play a critical role in metabolism by controlling the expression of key genes, are small non-coding RNAs that bind to the 3' untranslated regions (UTRs) of target mRNAs, leading to mRNA degradation, deadenylation, or translational repression. However, interactions of miRNAs with other regions, including the 5' UTR, coding sequence, and gene promoters, have also been reported. Under certain conditions, miRNAs can activate translation or regulate transcription (Broughton et al., 2016). The interaction of miRNAs with their target genes is dynamic and dependent on many factors, such as the subcellular location of miRNAs, the abundance of miRNAs and target mRNAs, and the affinity of miRNA-mRNA interactions (Vasudevan, 2012; Makarova et al., 2016). miRNAs can be secreted into extracellular fluids and transported to target cells via vesicles, such as exosomes, or by binding to proteins, including Argonautes. Extracellular miRNAs function as chemical messengers to mediate cell-cell communication (Hayes et al., 2014; Wang et al., 2016; Huang, 2017).

The international bovine genome sequencing and HapMap projects have enabled the discovery of substantial numbers of single nucleotide polymorphisms (SNPs) throughout the cattle genome. These markers have contributed to the development of SNP marker panels, which are widely used to increase the accuracy of selection and implement economic breeding programs in cattle. Recent advances in massively parallel sequencing technologies, also known as next-generation sequencing (NGS), have further enabled the discovery of large amounts of genetic variation and causal mutations. Several studies have successfully applied NGS technologies in cattle and have shown that many SNPs and insertions-deletions (InDels) remain to be detected, particularly in diverse cattle breeds or multiple individuals. The genome of one Proven Hanwoo bull was sequenced to a 45.6-fold coverage using the ABI SOLiD system. Therefore, 4.7 million single nucleotide polymorphisms and 0.4 million small indels were identified by comparison with the Btau4.0 reference assembly (Lee et al., 2013). Choi et al (2015) reported the discovery of 10.4 and 13.5 million SNPs in the whole genome of Hanwoo at an average coverage of approximately 10.71 and 26-fold, respectively, by mapping to the UMD 3.1 reference bovine genome (Choi et al., 2015; Choi et al., 2014). Therefore, the aim of this study is to identify single nucleotide polymorphisms (SNPs) located in microRNA sequences associated with lipid metabolism, using massively whole-genome sequence information obtained from open-source databases on NCBI for Korean native cattle.

Materials and methods

Collection of sequence reads and Database

All of the raw sequence data were obtained from two sources: the Sequence Read Archive (SRA) of NCBI (n=56), and reads that were supported by the National Institute of Animal Science (n=19). Additionally, the ARS-UCD1.2 (GCF_002263795.1) was utilized as the reference genome. SNP and Indel positions, comprising approximately 97 million SNPs and indels, were acquired from dbSNP 140 (https://ftp.ensembl.org/pub/release-109/variation/vcf/bos_taurus/).

Data pre-processing and mapping

Seventy-four sets of raw data were sequenced on an Illumina HiSeq system. Quality control was performed on the raw reads, including the removal of adapter sequences and bases with low quality, using the fastp software (version 0.23.1) with default parameters. The trimmed reads were mapped to the bovine reference genome, specifically the ARS-UCD1.2 version, using bwa-mem. The -M option was used to mark shorter split hits as secondary alignments. Default parameters were used for all other steps. After mapping, the resulting SAM files were converted to BAM files and sorted simultaneously using SortSam in Picard tools. PCR duplicates were removed from the mapped reads using MarkDuplicates of GATK (version 4.2.6.1). Finally, variant calling was performed using only the uniquely mapped reads.

Base quality optimization and Variant calling

We used GATK (version 4.2.6.1) BaseRecalibrator, ApplyBQSR, and AnalyzeCovariates modules to recalibrate base quality scores. This process entailed constructing an error model using known covariates from all base calls, and then adjusting the dataset according to this model. We followed the best practice guidelines recommended for variant discovery and genotyping using GATK (version 4.2.6.1) with default parameters for all commands. First, genotype likelihoods were calculated separately for each sequenced animal using GATK HaplotypeCaller, which resulted in gVCF (genomic Variant Call Format) files for each sample. The gVCF files from the 74 samples were consolidated using GATK GenomicsDBImport. Subsequently, GATK GenotypeGVCFs was applied to genotype polymorphic sequence variants for all samples simultaneously.

Hard filtering

In the first step, the raw VCF files were separated, and different thresholds were applied for SNPs and indels using GATK SelectVariants with the '-select-type SNP' and '-select-type INDEL' options. The GATK VariantFiltration module was then used to filter the raw VCF files. Quality control on the raw sequencing variants and genotypes was performed following the guidelines recommended for GATK HaplotypeCaller.

For SNPs, variants were retained if they satisfied the following criteria: a QualByDepth (QD) of less than 2.0, a variant quality score recalibration score (QUAL) of less than 30.0, a StrandOddsRatio (SOR) of more than 3.0, a FisherStrand (FS) of more than 60.0, an RMSMappingQuality (MQ) of less than 40.0, a MappingQualityRankSumTest (MQRankSum) of less than -12.5, and a ReadPosRankSumTest (ReadPosRankSum) of less than -8.0. For indels, variants were retained if they met the following criteria: QualByDepth (QD) < 2.0, variant quality score recalibration score (QUAL) < 30.0, FisherStrand (FS) > 200.0, ReadPosRankSumTest (ReadPosRankSum) < -20.0. After applying the filtering step to the SNP and indel calls, the GATK SortVcf tool was used to merge the variant calls into a single VCF file. Next, the filtered variants were removed from the merged VCF using the GATK SelectVariants tool, resulting in a final dataset containing only the high-quality variant calls.

Variant functional annotation and Computing environment

After variant filtration, SNPs and InDels were functionally annotated using SnpEff (version 5.1) to assign each variant a functional class and provide various fields of information for coding variants, such as identifying the affected transcript and protein. The SnpEff database ARS-UCD1.2.105 was used for functional annotation. The functional classes assigned to both SNPs and InDels were exonic, intronic, intergenic, splice site acceptor, splice site donor, splice site region, downstream, upstream, UTR 3 prime, and UTR 5 prime. Functional classes exclusively assigned to InDels were conserved in-frame deletion and insertion, and disruptive in-frame deletion and insertion. Functional classes exclusively assigned to SNPs were stop lost, stop retained, initiator codon, and synonymous. All computations were executed on a desktop computer equipped with a 16-core AMD RyzenTM 9 5950X processor (with a base frequency rated at 3.4 GHz) and 125 GB of random-access memory.

Results and Discussion

Hanwoo sequence assembly and Mapping

We collected fifty-five and nineteen sequence reads from sequence read archive (SRA) of NCBI and the National Institute of Animal Science, respectively. As shown in Table 1, the SRR data obtained from BioProject Accession no. PRJNA318149, PRJNA288838, and PRJNA210523 are expressed as hanwoo_ncbi19, hanwoo_ncbi125, and hanwoo_ncbi34 for Korean native cattle, respectively. The data obtained from the National Institute of Animal Science are expressed as hanwoo_nias.

Table 1. Summary of sequencing and short read alignment results.

	hanwoo_ncbi19	hanwoo_ncbi25	hanwoo_ncbi34	hanwoo_nias[1)]
BioProject Accession no.	PRJNA318149	PRJNA288838	PRJNA210523	-
Platform model	Illumina HiSeq2000	Illumina Genome Analyzer Iix	Illumina HiSeq2000	-
Number of individuals	20	14	21	19
Paired-end length(bp)	200	152-168	202	180-300
Average reads per individual	73.47M	25.54M	393.18M	389.75M
Average sequence depth per individual	2.29X	0.43X	10.67X	12.55X
Average map reads per individual	72.17M	23.84M	388.91M	380.70M
Average read mapping rate	98.19%	93.09%	98.91%	97.67%
Average read properly paired	94.87%	86.71%	92.45%	95.05%
Average coverage rate	33.47%	33.22%	99.44%	99.24%
Duplication ratio	0.10	0.01	0.12	0.06

[1)]The sequenced reads were obtained from the National Institute of Animal Science

Paired-end length ranges from 152 bp and 300 bp, and the minimum average reads per individual is 25.54 million, while the maximum is 389.75 million. A total of 73.47 million (98.19%), 25.54 million (93.09%), 393.18 million (98.91%), and 380.70 million (97.67%) high-qulity-trimmed and filtered reads, respectively, were successfully mapped to the bovine reference genome (ARS-UCD1.2), with sequencing coverage rate of 33.47%, 33.22%, 99.44%, and 99.24%, respectively.

Variant identification and comparison

High-quality, trimmed, and filtered sequence reads were mapped to the bovine reference genome assembly, resulting in the detection of 25,982,458 variants, which include both SNPs and Indels. Of these, 21,795,577 were SNPs and 4,186,881 were Indels, with an average variant rate of one variant every 101 bases. Table 2 presents the number and rate of variants in the Hanwoo chromosome, where Chromosome 1 has the highest number of variants, while Chromosome 25 has the lowest. According to a study reported by Choi et al. (2015), they discovered a total of 13,544,560 SNPs, of which 22.3% were found to be novel, by mapping the bovine reference assembly (UMD 3.1). Furthermore, in a study by Choi et al. (2014), 10.4 million SNPs and 1,063,267 InDels were found to align with the bovine reference assembly (ARS-UCD1.2).

Compared to their results, although our coverage depth is similar or lower, the number of individuals used in our study is fairly high and more than sufficient to detect high-confidence variants. In this study, we performed a comprehensive genome analysis of seventy-four Hanwoo cattle using whole-genome sequencing reads from the Sequence Read Archive (SRA) data of NCBI. Our analysis revealed a total of 25,982,458 variants across the entire Hanwoo genome. This finding contrasts with previous studies by Choi et al. (2014) and Choi et al. (2015), who reported detecting a total of 11,534,445 SNP and InDel variants in the whole genome of two Hanwoo, along with 13,544,560 SNPs in ten Hanwoo.

Table 2. Number of variants in each bovine chromosome and its putative impact by sequence ontology term

Chromosome	Length	Variants	Variants rate	Sequence ontology term	No. of variant	Percent	Putative impact[1)]
1	158,534,110	1,575,413	100	3_prime_UTR_variant	142,721	0.35%	MODIFIER
2	136,231,102	1,267,981	107	5_prime_UTR_premature_start_codon_gain_variant	6,599	0.02%	LOW
3	121,005,158	1,234,674	98	5_prime_UTR_truncation	5	0%	MODERATE
4	120,000,601	1,304,963	91	5_prime_UTR_variant	50,207	0.12%	MODIFIER
5	120,089,316	1,169,173	102	bidirectional_gene_fusion	19	0%	HIGH
6	117,806,340	1,203,217	97	conservative_inframe_deletion	933	0.00%	MODERATE
7	110,682,743	1,042,062	106	conservative_inframe_insertion	866	0.00%	MODERATE
8	113,319,770	1,068,891	106	disruptive_inframe_deletion	1,687	0.00%	MODERATE
9	105,454,467	986,444	106	disruptive_inframe_insertion	922	0.00%	MODERATE
10	103,308,737	1,026,131	100	downstream_gene_variant	2,008,226	4.86%	MODIFIER
11	106,982,474	997,382	107	exon_loss_variant	15	0%	HIGH
12	87,216,183	1,096,475	79	frameshift_variant	8,218	0.02%	HIGH
13	83,472,345	823,401	101	gene_fusion	27	0%	HIGH
14	82,403,003	813,667	101	initiator_codon_variant	36	0%	LOW
15	85,007,780	961,573	88	intergenic_region	17,100,123	41.42%	MODIFIER
16	81,013,979	787,362	102	intragenic_variant	42	0%	MODIFIER
17	73,167,244	738,685	99	intron_variant	19,542,318	47.33%	MODIFIER
18	65,820,629	686,664	95	missense_variant	158,605	0.38%	MODERATE
19	63,449,741	639,081	99	non_coding_transcript_exon_variant	28,212	0.07%	MODIFIER
20	71,974,595	709,114	101	non_coding_transcript_variant	116	0%	MODIFIER
21	69,862,954	688,501	101	splice_acceptor_variant	1,947	0.01%	HIGH
22	60,773,035	569,336	106	splice_donor_variant	2,030	0.01%	HIGH
23	52,498,615	804,268	65	splice_region_variant	48,353	0.12%	MODERATE
24	62,317,253	657,676	94	start_lost	324	0.00%	HIGH
25	42,350,435	449,236	94	start_retained_variant	6	0%	LOW
26	51,992,305	500,744	103	stop_gained	2,228	0.01%	HIGH
27	45,612,108	495,004	92	stop_lost	331	0.00%	HIGH
28	45,940,150	493,246	93	stop_retained_variant	161	0%	LOW
29	51,098,607	614,032	83	synonymous_variant	217,282	0.53%	LOW
X	139,009,144	578,062	240	transcript_ablation	10	0%	HIGH
Total	2,628,394,923	25,982,458	101	upstream_gene_variant	1,964,455	4.76%	MODIFIER

1) HIGH, a high impact on protein alteration; MODERATE, missense variant; LOW, synonymous variant; MODIFIER, non-coding variants.

Functional annotation and miRNA

The discovered variants in Hanwoo whole-genome were functionally annotated for every single SNP and Indel using SnpEff (version 5.1) as described by Cingolani et al (2012). Out of the total SNPs and Indels discovered in this study, 17,100,123 variants (41.47%) and 19,500,395 variants (47.29%) were located in intergenic and intronic regions, respectively. Most of the variants are identified in the intronic region.

Out of the three functional classes, 159,155 missense were found to alter the DNA and result in amino acid substitutions. Additionally, 2,069 nonsense were identified that generate a point mutation resulting in a premature stop codon or a nonsense codon in the transcribed mRNA. and 217,438 silent were found.

In previous studies, Choi et al (2015) identified 36,511 missense, 399 nonsense, and 49,723 silent mutations in the whole-genome of Hanwoo (n=10), while Iqbal et al (2019) reported 155,251 missense, 1,132 nonsense, and 255,296 silent mutations in the whole-genome of indicine cattle from Pakistan (n=20). Our study found a different number of missense, nonsense, and silent mutations compared to these previous studies. However, our results were similar to those reported by Iqbal et al (2019).

Recently, several researchers reported that variants located in non-coding regions play an important role in regulating gene expression. Table 2 presents the number of variants and their putative impact based on sequence ontology terms. In our study, we identified 142,721 and 56,797 variants in the 3' and 5' untranslated regions, respectively. Our results showed a higher number of variants compared to a previous study on the whole-genome sequence of Hanwoo. However, the variant numbers reported by Iqbal et al (2019) were similar to ours.

The number of variants for miRNA and protein coding are presented in Table 3. Most of the variants were observed in protein coding regions rather than in miRNA. For variant impact in miRNA, 24 variants were identified in Impact_LOW, and 74,205 variants were identified in Impact_MODIFIER. For variant effect in miRNA, 37,404 variants were identified in down_gene_variant, and 36,428 variants were identified in upstream_gene_variant. In addition, we identified 397 non_coding_transcript_exon_variant and 24 splice_region_variant.

miRNA differentially expressed in subcutaneous and intramuscular fat tissues

In a previous study, the conformation of the 22 most abundant miRNA expressions by qRT-PCR indicated that they were highly accumulated not only in subcutaneous fat tissue but also in intramuscular fat tissue in bovines (Wang et al., 2015). As a result, four of the miRNAs (bta-mir-143, bta-mir-206, bta-mir-378, and bta-mir-133a) were found to be differentially regulated in fat deposition between subcutaneous and intramuscular fat tissue.

Table 4 shows 18 putative miRNAs expressed in bovine adipocyte and their 1,179 variants. We identified 941 miRNAs through functional annotation. Then, by comparing them to the 22 miRNAs that were differentially expressed in intramuscular and subcutaneous fat tissue, we identified 18 putative miRNAs. As shown in Table 4, MODIFIER identified 1,179 variants in 18 miRNA sequences as non-coding variants. Of these 1,179 variants, 650 variants were identified as downstream_gene_variant, 3 variants as non-coding_transcript_exon_variant, and 526 variants as upstream_gene_variants. Among the putative 18 miRNAs, two miRNAs (bta-mir-143 and bta-mir-145) were found to be highly expressed in intramuscular fat tissue.

MicroRNAs (miRNAs) are non-coding RNAs that regulate gene expression in many biological processes. Wang et al (2013) identified 88 differentially expressed miRNAs, and 30 of them were significantly changed between intramuscular and subcutaneous fat. Romao et al (2012) found that miRNA expression was not equal between fat depots and diets, detecting a total of 207 miRNAs in both bovine subcutaneous and visceral fat tissues from different diets. They also found that some miRNAs were tissue-specific, while others were diet-specific. Moreover, high-fat diets increased the number of miRNAs detected per animal.

In this study, we identified 18 putative miRNAs that were differentially expressed in intramuscular and subcutaneous fat tissue of bovine. Some of these miRNAs, including miR-143, miR-145, let-7, and miR-92a, are known to play important regulatory roles in adipose tissue

Table 3. The number of variants for each impact and effect of miRNA and protein coding

Variant impact and effect[1)]	miRNA	Protein_coding
Impact_HIGH	-	14,242
Impact_LOW	24	262,346
Impact_MODERATE	-	162,798
Impact_MODIFIER	74,205	22,670,210
3_prime_UTR_variant	-	142,721
5_prime_UTR_premature_start_codon_gain_variant	-	6,599
5_prime_UTR_truncation	-	5
5_prime_UTR_variant	-	50,207
conservative_inframe_deletion	-	933
conservative_inframe_insertion	-	866
disruptive_inframe_deletion	-	1,687
disruptive_inframe_insertion	-	922
downstream_gene_variant	37,404	1,719,046
exon_loss_variant	-	15
frameshift_variant	-	8,217
initiator_codon_variant	-	36
intron_variant	-	19,123,043
missense_variant	-	158,540
non_coding_transcript_exon_variant	397	-
non_coding_transcript_variant	-	116
splice_acceptor_variant	-	1,849
splice_donor_variant	-	1,929
splice_region_variant	24	47,124
start_lost	-	324
start_retained_variant	-	6
stop_gained	-	2,227
stop_lost	-	330
stop_retained_variant	-	160
synonymous_variant	-	217,252
transcript_ablation	-	-
upstream_gene_variant	36,428	1,676,119

1) HIGH, a high impact on protein alteration; MODERATE, missense variant; LOW, synonymous variant; MODIFIER, non-coding variants.

and adipogenesis. High expression signal values of let-7 family members, let-7c, let-7d, and let-7e, were observed in subcutaneous fat, which is consistent with previous studies demonstrating that let-7 miRNA is upregulated during mouse adipocyte differentiation and regulates adipogenesis by targeting HMGA2 transcription factor (Sun et al., 2009). We also observed significant up-regulation of bta-miR-143 in bovine intramuscular fat tissues (Table 4 and Fig. 2). The mammalian orthologs of bta-miR-143, which potentially target extracellular signal-regulated protein kinase 5 (ERK5), have been implicated in obesity as they are up-regulated during adipogenic differentiation (He et al., 2013; Li et al., 2011; Takanabe et al., 2008). Additionally, miR-145 was found to be more than 5.46-fold up-regulated in intramuscular fat tissues, suggesting its role in adipose biology. A recent report revealed that up-regulated miR-145 expression inhibits porcine preadipocyte differentiation by targeting IRS1 (Guo et al., 2012). The present study provides important insights into the regulation of miRNAs in bovine adipose tissue, which could be useful for understanding adipogenesis and the pathogenesis of metabolic diseases.

Table 4. highly abundant miRNAs differentially expressed in bovine subcutaneous and intramuscular fat tissue and related genetic variants.

GeneName	GeneId	TranscriptId	MODIFIER	downstream_ gene_variant	non coding_transcript _exon_variant	upstream_gene _variant
mir- let-7c	ENSBTAG0 0000029912	ENSBTAT00 000042291	49	30	0	19
mir- let-7e	ENSBTAG0 0000029880	ENSBTAT00 000042259	64	27	0	37
bta- let-7d	ENSBTAG0 0000029853	ENSBTAT00 000042232	111	72	0	39
bt a-mir- 125b-1	ENSBTAG0 0000029925	ENSBTAT00 000042304	87	38	0	49
bt a-mir- 125b-2	ENSBTAG0 0000029827	ENSBTAT00 000042206	23	15	0	8
bta-m ir-126	ENSBTAG0 0000029979	ENSBTAT00 000042358	102	45	0	57
bta-m ir-143	ENSBTAG0 0000030114	ENSBTAT00 000042493	41	27	0	14
bta-m ir-145	ENSBTAG0 0000029857	ENSBTAT00 000042236	43	27	0	16
bta-m ir-16a	ENSBTAG0 0000036389	ENSBTAT00 000050889	53	24	0	29
bta-m ir-195	ENSBTAG0 0000030113	ENSBTAT00 000042492	44	32	0	12
bta-m ir-23a	ENSBTAG0 0000029901	ENSBTAT00 000042280	25	16	0	9
b ta-mir -26a-1	ENSBTAG0 0000029937	ENSBTAT00 000042316	42	25	0	17
b ta-mir -26a-2	ENSBTAG0 0000029869	ENSBTAT00 000042248	62	39	0	23
bt a-mir- 320a-1	ENSBTAG0 0000029761	ENSBTAT00 000042140	34	23	0	11
bt a-mir- 320a-2	ENSBTAG0 0000029760	ENSBTAT00 000042139	101	52	2	47
bta-m ir-342	ENSBTAG0 0000030073	ENSBTAT00 000042452	123	64	1	58
b ta-mir -92a-2	ENSBTAG0 0000030059	ENSBTAT00 000042438	42	32	0	10
b ta-mir -2325b	ENSBTAG0 0000048892	ENSBTAT00 000069205	133	62	0	71

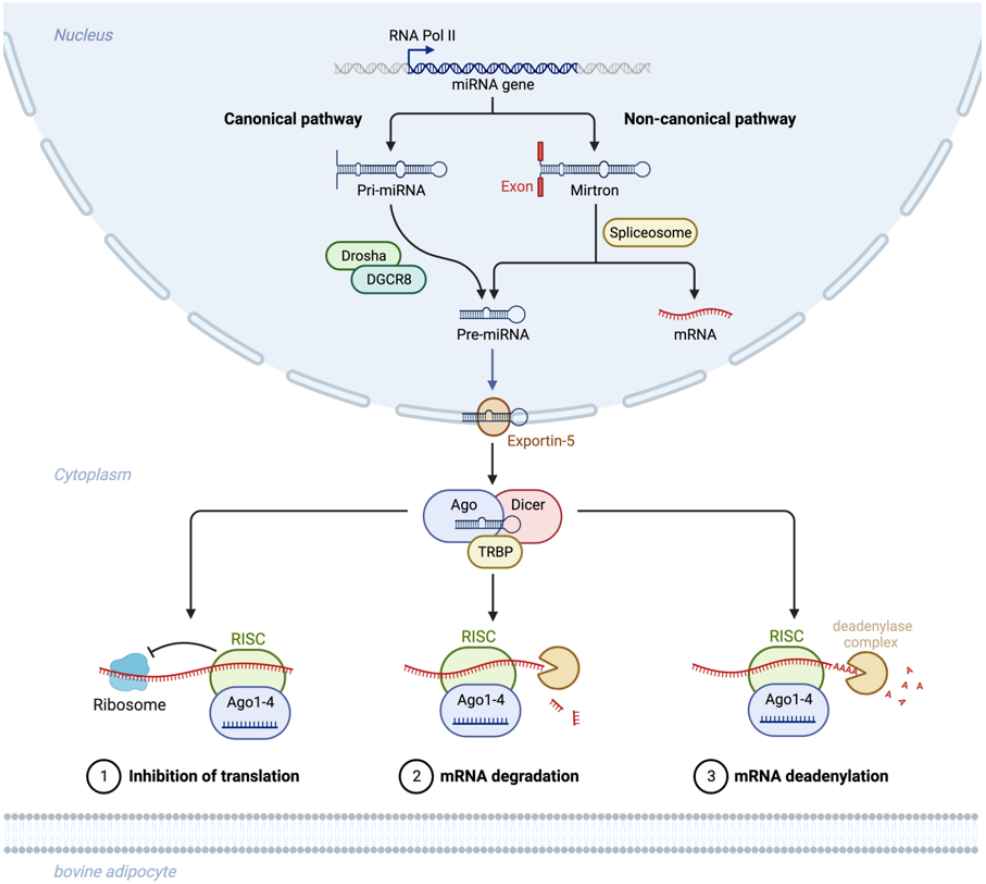

Figure 2. An overview of microRNA biogenesis and its role in bovine adipocytes.

Conclusion

In this study, we carried out a massive genomic analysis of Korean native cattle using whole-genome sequencing data. We discovered a significant number of variants, including both SNPs and Indels, with a high rate of variants located within intergenic and intronic regions. Notably, we identified more variants than previous studies, despite a similar or lower coverage depth. This discrepancy may be attributed to our larger sample size, which enhances the detection of high-confidence variants.

We further examined the functional implications of these variants, revealing a substantial number of missense, nonsense, and silent mutations. Our results showed a variance when compared to the findings of previous studies, but were comparable to those reported by Iqbal et al (2019). Importantly, we also observed a large number of variants in non-coding regions, which are known to play crucial roles in gene expression regulation. A particularly novel aspect of our research focused on the examination of miRNA expression and its variants. We discovered a substantial number of variants in the 3' and 5' untranslated regions, and identified 18 putative miRNAs differentially expressed between subcutaneous and intramuscular fat tissues. Two of these, bta-mir-143 and bta-mir-145, were found to be highly expressed in intramuscular fat tissue. This finding contributes to our understanding of the role miRNAs play in adipogenesis and metabolic diseases, and the potential implications of these results for future studies and applications are substantial.

In conclusion, our study provides new insights into the genomic landscape of Hanwoo cattle, significantly expanding our understanding of the genetic variability and functional impact of these variations. The data generated in this study will be valuable for future research in livestock genetics, including efforts to improve the productivity and health of Hanwoo cattle, and potentially other breeds as well. Furthermore, our findings on differentially expressed miRNAs in fat tissues pave the way for more nuanced studies on fat deposition in bovines, with implications for meat quality.

Acknowledgement

We are also grateful to the members of our research team who actively participated in data collection, analysis, and discussion. Their diligent efforts and expertise significantly contributed to the success of this project. Furthermore, we extend our thanks to the National Institute of Animal Science for providing the Hanwoo sequencing data that enabled us to carry out this research.

References

Broughton JP, Lovci MT, Huang JL, Yeo GW, Pasquinelli AE. 2016. Pairing beyond the Seed Supports MicroRNA Targeting Specificity. Mol Cell. 64: 320–33. doi: 10.1016/j.molcel.2016.09.004

Choi JW, Liao X, Stothard P, Chung WH, Jeon HJ, Miller SP, Choi SY, Lee JK, Yang B, Lee KT, Han KJ, Kim HC, Jeong D, Oh JD, Kim N, Kim TH, Lee HK, Lee SJ. 2014. Whole-Genome Analyses of Korean Native and Holstein Cattle Breeds by Massively Parallel Sequencing. PloS one. 9(7): e101127. doi: 10.1371/journal.pone.0101127

Choi JW, Choi BH, Lee SH, Lee SS, Kim HC, Yu DY, Chung WH, Lee KT, Chai HH, Cho YM, Lim DJ. 2015. Whole-Genome Resequencing Analysis of Hanwoo and Yanbian cattle to identify genome-wide SNPs and signatures of selection. Mol. Cells. 38(5): 466-473. doi: 10.14348/molcells.2015.0019

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w^{1118}; iso-2; iso-3. Fly (Austin). 6(2): 80–92. doi: 10.4161/fly.19695

Das A, Gupta MD, Shaha M, Lima A, Miazi OF, Das GB. 2023. Whole genome sequencing of Red Chittagong Cattle (RCC) cattle and insight into genetic variants in candidate genes for disease resistance. BioRxiv. In press. doi: 10.1101/2023.01.09.523278

Guo Y, Chen Y, Zhang Y, Zhang Y, Chen L, Mo D. 2012. Up-regulated miR-145 expression inhibits porcine preadipocytes differentiation by targeting IRS1. Int J Biol Sci.. 8(10):1408-17. doi: 10.7150/ijbs.4597.

Ha M, Kim VN. 2014. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 15: 509–24. doi: 10.1038/nrm3838

Hayes J, Peruzzi PP, Lawler S. 2014. MicroRNAs in cancer: biomarkers, functions and therapy. Trends Mol Med. 20:460–9. doi: 10.1016/j.molmed.2014.06.005

He Z, Yu J, Zhou C, Ren G, Cong P, Mo D, Chen Y, Liu X. 2013. MiR-143 is not essential for adipose development as revealed by in vivo antisense targeting. Biotechnol. Lett. 35(4): 499–507. doi: 10.1007/s10529-012-1112-3

Huang W. 2017. MicroRNAs: biomarkers, diagnostics, and therapeutics. Methods Mol Biol. 1617:57–67. doi: 10.1007/978-1-4939-7046-9_4

Ibañez-Escriche N, Gonzalez-Recio O. 2011. Review. Promises, pitfalls and challenges of genomic selection in breeding programs. Spanish Journal of Agricultural Research. 9(2): 404-413. doi: 10.5424/sjar/20110902-447-10

Iqbal N, Liu X, Yang T, Huang Z, Hanif Q, Asif M, Khan QM, Mansoor S. 2019. Genomic variants identified from whole-genome resequencing of indicine cattle breeds from Pakistan. PLoS One. 14(4): e0215065. doi: 10.1371/journal.pone.0215065.

Lee J, Kim JM, Garrick DJ. 2019. Increasing the accuracy of genomic prediction in pure-bred Limousin beef cattle by including cross-bred Limousin data and accounting for an F94L variant in MSTN. Animal genetics, 50(6): 621–633. doi: 10.1111/age.12846

Lee KT, Chung WH, Lee SY, Choi JW, Kim JW, Lim DJ, Lee SH, Jang GW, Kim BS, Choy YH, Liao X, Paul Stothard, Stephen S Moore, Lee SH, Ahn SM, Kim NS, Kim TH. 2013. Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity. BMC Genomics. 14: 519. doi: 10.1186/1471-2164-14-519

Lee SH, Cho YM, Lee JH, Oh SJ. 2015. Implementation of genomic selection in Hanwoo breeding program. Institute of Agricultural Science. 42(4): 397–406. doi: 10.7744/CNUJAS.2015.42.3.397

Li H, Zhang Z, Zhou X, Wang Z, Wang G, Han Z. 2011. Effects of microRNA-143 in the differentiation and proliferation of bovine intramuscular preadipocytes. Mol Biol Rep. 38: 4273–4280. doi: 10.1007/s11033-010-0550-z

Lopez BI, Lee SH, Park JE, Shin, DH, Oh JD, de Las Heras-Saldana S, van der Werf J, Chai HH, Park W, Lim D. 2019. Weighted Genomic Best Linear Unbiased Prediction for Carcass Traits in Hanwoo Cattle. Genes(Basel). 10(12): 1019. doi: 10.3390/genes10121019

Lopez BIM, An N, Srikanth K, Lee S, Oh JD, Shin DH, Park W, Chai HH, Park JE, Lim D. 2021. Genomic Prediction Based on SNP Functional Annotation Using Imputed Whole-Genome Sequence Data in Korean Hanwoo Cattle. Frontiers in genetics. 11: 603822. doi: 10.3389/fgene.2020.603822

Makarova JA, Shkurnikov MU, Wicklein D, Lange T, Samatov TR, Turchinovich AA,Tonevitsky AG. 2016. Intracellular and extracellular microRNA: An update on localization and biological role. Progress in histochemistry and cytochemistry. 51(3-4): 33–49. doi: 10.1016/j.proghi.2016.06.001

Mei C, Gui L, Hong J, Raza SHA, Aorigele C, Tian W, Garcia M, Xin Y, Yang W, Zhang S, Zan L. 2021. Insights into adaption and growth evolution: a comparative genomics study on two distinct cattle breeds from Northern and Southern China. Molecular Therapy. Nucleic Acids. 23: 959-967. doi: 10.1016/j.omtn.2020.12.028

Romao JM, Jin W, He M, McAllister T, Guan LL. 2012. Altered microRNA expression in bovine subcutaneous and visceral adipose tissues from cattle under different diet. PLoS ONE. 7(7): e40605. doi: 0.1371/journal.pone.0040605

Sun T, Fu M, Bookout AL, Kliewer SA, Mangelsdorf DJ. 2009. MicroRNA let-7 regulates 3T3-L1 adipogenesis. Mol Endocrinol. 23(6): 925–931. doi: 10.1210/me.2008-0298

Takanabe R, Ono K, Abe Y, Takaya T, Horie T, Wada H, Kita T, Satoh N, Shimatsu A, Hasegawa K. 2008. Up-regulated expression of microRNA-143 in association with obesity in adipose tissue of mice fed high-fat diet. Biochem Biophys Res Commun. 376(4): 728–732. doi: 10.1016/j.bbrc.2008.09.050

VanRaden PM. 2008. Efficient methods to compute genomic predictions. Journal of Dairy Science. 91(11): 4414–23. doi: 10.3168/jds.2007-0980

Vasudevan S. 2012. Posttranscriptional upregulation by microRNAs. Wiley Interdiscip Rev RNA. 3:311–30. doi: 10.1002/wrna.121

Wang J, Chen J, Sen S. 2016. MicroRNA as Biomarkers and Diagnostics. J Cell Physiol. 231:25–30. doi: 10.1002/jcp.25056

Wang HY, Zheng Y, Wang GL, Li HX. 2013. Identification of microRNA and bioinformatics target gene analysis in beef cattle intramuscular fat and subcutaneous fat. Mol. Biosyst. 9: 2154-2162. doi: 10.1039/C3MB70084D

Wang HY, Xiao SH, Wang M, Kim NH, Li HX, Wang GL. 2015. In silico identification of conserved microRNAs and their targets in bovine fat tissue. Gene. 559(2):119-128. doi: 10.1016/j.gene.2015.01.021

Xiang R, MacLeod IM, Daetwyler HD, de Jong G, O'Connor E, Schrooten C, Chamberlain A, Goddard M. 2021. Genome-wide fine-mapping identifies pleiotropic and functional variants that predict many traits across global cattle populations. Nat Commun. 12: 860. doi: 10.1038/s41467-021-21001-0