

Transcription of Endogenous Retrovirus Group K Members and Their Neighboring Genes in Chicken Skeletal Muscle Myoblasts

Tomohide Takaya^{1,2,3}, Yuma Nihashi¹, Tamao Ono² and Hiroshi Kagami²

¹ Department of Science and Technology, Graduate School of Medicine, Science and Technology, Shinshu University, Nagano 399-4598, Japan

² Department of Agricultural and Life Science, Faculty of Agriculture, Shinshu University, Nagano 399-4598, Japan

³ Department of Biomolecular Innovation, Institute for Biomedical Sciences, Shinshu University, Nagano 399-4598, Japan

Skeletal muscle myoblasts are myogenic precursor cells that generate myofibers during muscle development and growth. We recently reported that broiler myoblasts, compared to layer myoblasts, proliferate and differentiate more actively and promptly into myocytes, which corresponds well with the muscle phenotype of broilers. Furthermore, RNA sequencing (RNA-seq) revealed that numerous genes are differentially expressed between layer and broiler myoblasts during myogenic differentiation. Based on the RNA-seq data, we herein report that chicken myoblasts transcribe endogenous retrovirus group K member (ERVK) genes. In total, 16 ERVKs were highly expressed in layer myoblasts and two (termed BrK1 and BrK2) were significantly induced in broiler myoblasts. These transcribed ERVKs had a total of 182 neighboring genes within ± 100 kb on the chromosomes, of which 40% were concentrated within ± 10 kb of the ERVKs. We further investigated whether the transcription of ERVKs affects the expression of their neighboring genes. BrK1 had two neighboring genes; *LOC107052719* was overlapping with BrK1 and down-regulated in the broiler myoblasts, and *FAM19A2* was upregulated in the broiler myoblasts as well as BrK1. BrK2 had 14 neighboring genes, and only one gene, *LOC772243*, was differentially expressed between layer and broiler myoblasts. *LOC772243* was overlapping with BrK2 and suppressed in the broiler myoblasts. These data indicate that the transcription of ERVKs may impact the expression of their neighboring genes in chicken myoblasts.

Key words: endogenous retrovirus group K, RNA sequencing, skeletal muscle myoblast

J. Poult. Sci., 58: 79–87, 2021

Introduction

Broiler chickens are the breeds genetically selected for the rapid development of skeletal muscle tissue with a high feed efficiency in order to improve meat production. The skeletal muscle of the broilers is composed of a higher number of myofibers presenting larger diameters and faster growth compared to other breeds such as layers (Scheuermann *et al.*, 2004). Myofibers are multinuclear contractile myocytes and have dozens of stem cells (named satellite cells) between the basal lamina and plasma membrane. In the process of mus-

cle development and growth, satellite cells are transformed into the myogenic precursor cells termed myoblasts. Myoblasts proliferate through several rounds of cell divisions and then differentiate into mononuclear myocytes. Finally, myocytes fuse to form nascent myotubes, or fuse to preexisting myofibers (Dumont *et al.*, 2015). Thus, the characteristics of myoblasts directly reflect the muscle phenotype of the animals.

We recently reported that myoblasts isolated from UK Chunky (UKC) broilers actively proliferate and promptly differentiate compared to those of White Leghorn (WL) layers. Comprehensive gene expression analyses by RNA sequencing (RNA-seq) revealed that numerous genes are differentially expressed between WL and UKC myoblasts during myogenic differentiation. UKC myoblasts highly express cell cycle genes in growing condition and muscular component genes in differentiation stages, which corresponds well with the cellular properties of UKC myoblasts and eventually with the muscle phenotype of broilers (Nihashi *et al.*, 2019b). However, major myogenic regulatory factors (MRFs) such as MyoD and myogenin were not

Received: March 18, 2020, Accepted: May 29, 2020

Released Online Advance Publication: July 25, 2020

Correspondence: Dr. Tomohide Takaya, Department of Agricultural and Life Science, Faculty of Agriculture, Shinshu University, 8304 Minami-minowa, Kami-ina, Nagano 399-4598, Japan.

(E-mail: ttakaya@shinshu-u.ac.jp)

The Journal of Poultry Science is an Open Access journal distributed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. To view the details of this license, please visit (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).

upregulated in UKC myoblasts as opposed to WL myoblasts. It is not fully understood why UKC myoblasts eventually achieve higher myogenic ability. Therefore, the precise understanding of the regulatory mechanisms of gene expression in myoblasts will contribute to improvements in chicken breeding and meat production.

Through the analyses of RNA-seq data, we found that chicken myoblasts transcribe endogenous retrovirus group K members (ERVks). The RNA genome of the retrovirus, which commonly consists of a 5' long terminal repeat (LTR), *gag* (group-specific antigen), *pro* (protease), *pol* (polymerase), *env* (envelope), and 3' LTR, is reverse-transcribed and integrated into the genome of a host cell as a provirus. The provirus inserted into a germ cell is vertically inherited in that line. Such heritable proviruses are endogenous retroviruses (ERVs), which are able to be transcribed to make copies unless silenced (Gifford *et al.*, 2018). Typical vertebrate genomes have thousands of ERVs. In chickens, eleven ERV families occupy > 21 Mb, which is ~2% of the genome (Huda *et al.*, 2008). Chicken ERVs mainly consist of ERVK and ERVL (Barr *et al.*, 2005). Although ERVKs use a lysine (K) tRNA to bind to a viral primer binding site for reverse transcription, this does not imply that each ERVK member is phylogenetically related.

In humans, the mRNAs or the proteins derived from ERVKs play etiological roles in neurologic diseases (Manghera *et al.*, 2014), cancer (Downey *et al.*, 2015), and amyotrophic lateral sclerosis (Manghera *et al.*, 2016). Accumulating evidence suggests that ERVK transcription potentially influences the physiological homeostasis of the host animal. As one of the mechanisms, ERVK is thought to regulate the expression of its neighboring genes. The LTR of ERVK serves as the promoter harboring multiple binding sites for transcription factors such as YY1 (Knössl *et al.*, 1999), Sp1/3 (Fuchs *et al.*, 2011), and nuclear factor- κ B (NF- κ B) (Manghera and Douville, 2013). The transcription factors recruited on the LTR of ERVK can modulate expression of not only ERVK itself but also its neighboring genes.

The human expressed sequence tag database revealed that skeletal muscles express ERVK but not ERVE, ERVH, or ERVW (Stauffer *et al.*, 2004), which suggests that ERVK promoters are responsive in myogenic cells. In this study, we surveyed loci and neighboring genes of the ERVKs transcribed in chicken myoblasts. RNA-seq data demonstrated that the transcription of ERVKs may affect the expression of their neighboring genes in chicken myoblasts.

Materials and Methods

Chicken Myoblasts

All experimental procedures were conducted in accordance with the Regulations for Animal Experimentation of Shinshu University. The animal experimentation protocol was approved by the Committee for Animal Experiments of Shinshu University.

Myoblasts from hindlimb muscle of WL or UKC chickens at embryonic day 10 were isolated, primary-cultured, and induced to differentiate as previously described (Takaya *et*

al., 2017; Nihashi *et al.*, 2019a, b). Myoblasts were cultured on dishes coated with collagen type I-C (Cellmatrix; Nitta Gelatin, Osaka, Japan) at 37°C with 5% CO₂ throughout the experiments. Undifferentiated myoblasts were maintained in growth medium (GM) consisting of RPMI1640 (Nacalai, Osaka, Japan), 20% fetal bovine serum (FBS) (HyClone; GE Healthcare, UT, USA), 1% non-essential amino acids (Wako, Osaka, Japan), 1% chicken embryo extract (US Biological, MA, USA), 2 ng/mL basic fibroblast growth factor (Wako), and a mixed solution of 100 units/mL penicillin and 100 μ g/mL streptomycin (PS) (Nacalai). To initiate myogenic differentiation, 2.0 \times 10⁵ myoblasts in GM were seeded on 60 mm dishes. Next day (defined as day 0), the myoblasts were induced to differentiate by replacing GM with differentiation medium consisting of DMEM (Nacalai), 2% FBS, and PS. Myoblasts from each breed were subjected to experiments three times independently. Total RNA was isolated from the myoblasts using NucleoSpin RNA Plus (Macherey-Nagel, Düren, Germany) on days 0, 1, and 2.

RNA-seq Data

The RNA-seq data used in this study have been reported (Nihashi *et al.*, 2019b) and deposited into the DDBJ Sequence Read Archive (DRA; Research Organization of Information and Systems, National Institute of Genetics, Mishima, Japan) with the accession number DRA007964. Briefly, myoblasts from each breed were subjected to RNA-seq in triplicate on differentiation days 0, 1, and 2. Expression levels of 26,640 transcripts were quantified as reads per kilobase per million reads (RPKM) or as transcripts per kilobase per million (TPM). The false discovery rate (FDR) was employed to correct the *p* values.

Survey for ERVK-neighboring Genes

Loci of the ERVKs transcribed in chicken myoblasts were confirmed using Gene Tools (<https://www.ncbi.nlm.nih.gov/gene/>) (NCBI; National Center for Biotechnology Information, MD, USA). Gene annotations were performed using NCBI Gallus gallus Annotation Release 104 as a reference genome (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Gallus_gallus/104/). The ERVK-neighboring gene was defined as the gene that has a start or end position located within \pm 100 kb of the ERVKs.

Statistical Analysis

The data was presented as mean \pm standard error. Statistical differences between two groups were evaluated by unpaired two-tailed Student's *t* test. Statistical differences between more than three groups were evaluated by FDR *p* value. Statistical significance was set at *p* value < 0.05.

Results

The Transcribed ERVKs and Their Neighboring Genes in Chicken Myoblasts

RNA-seq detected the transcription of 44 ERVs in WL or UKC myoblasts on differentiation days 0, 1, or 2. One of them, *LOC100857682*, was an ERVFC1 encoding envelope polyprotein. The other 43 ERVs were ERVKs. We termed these ERVKs transcribed in chicken myoblasts "chMB-ERVKs". As shown in Table 1, four chMB-ERVKs were

Table 1. **The 43 ERVKs transcribed in chicken myoblasts (chMB-ERVKs)**

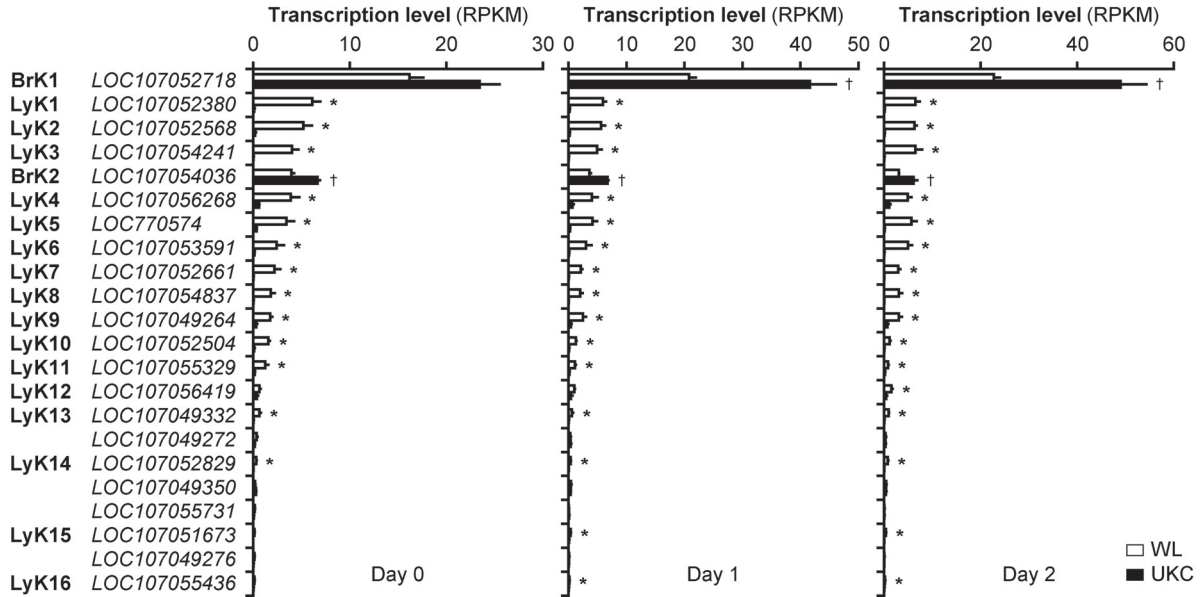
Chr ¹	Gene	Product
1	<i>LOC107050425</i>	ERVK-25 Pol protein-like
1	<i>LOC107050867</i>	ERVK-18 Pol protein-like
1	<i>LOC107051640</i>	ERVK-113 Pol protein-like
1	<i>LOC107051645</i>	ERVK-18 Pol protein-like
1	<i>LOC107052718</i>	(BrK1) ERVK-11 Pol protein
1	<i>LOC107054241</i>	(LyK3) ERVK-25 Pol protein-like
1	<i>LOC107054837</i>	(LyK8) ERVK-18 Pol protein-like
1	<i>LOC107055731</i>	ERVK-25 Pol protein-like
2	<i>LOC770574</i>	(LyK5) ERVK-25 Pol protein
2	<i>LOC107051673</i>	(LyK15) ERVK-18 Pol protein-like
2	<i>LOC107052504</i>	(LyK10) ERVK-18 Pol protein-like
2	<i>LOC107052568</i>	(LyK2) ERVK-113 Gag polyprotein-like
2	<i>LOC107052661</i>	(LyK7) ERVK-18 Pol protein-like
2	<i>LOC107052685</i>	ERVK-11 Pol protein-like
2	<i>LOC107052829</i>	(LyK14) ERVK-113 Pol protein-like
2	<i>LOC107052830</i>	ERVK-113 Pol protein-like
3	<i>LOC107053122</i>	ERVK-18 Pol protein-like
4	<i>LOC107053216</i>	ERVK-113 Pol protein-like
4	<i>LOC107056268</i>	(LyK4) ERVK-18 Pol protein-like
5	<i>LOC107053591</i>	(LyK6) ERVK-18 Pol protein-like
5	<i>LOC107056419</i>	(LyK12) ERVK-18 Pol protein-like
8	<i>LOC107054036</i>	(BrK2) ERVK-9 Pol protein-like
18	<i>LOC107057172</i>	ERVK-18 Pol protein-like
18	<i>LOC107057173</i>	ERVK-8 Gag polyprotein-like
28	<i>LOC107055329</i>	(LyK11) ERVK-18 Pol protein-like
33	<i>LOC107049264</i>	(LyK9) ERVK-8 Gag polyprotein-like
W	<i>LOC107049272</i>	ERVK-8 Gag polyprotein-like
W	<i>LOC107049273</i>	ERVK-9 Pol protein-like
W	<i>LOC107049274</i>	ERVK-18 Pol protein-like
W	<i>LOC107049276</i>	ERVK-18 Pol protein-like
W	<i>LOC107049277</i>	ERVK-9 Pol protein-like
W	<i>LOC107055436</i>	(LyK16) ERVK-8 Pol protein-like
W	<i>LOC107055437</i>	ERVK-18 Pol protein-like
W	<i>LOC107055442</i>	ERVK-18 Pol protein-like
W	<i>LOC107055443</i>	ERVK-25 Pol protein-like
W	<i>LOC107055445</i>	ERVK-25 Pol protein-like
Z	<i>LOC107049326</i>	ERVK-18 Pol protein-like
Z	<i>LOC107049332</i>	(LyK13) ERVK-113 Pol protein-like
Z	<i>LOC107049350</i>	ERVK-25 Pol protein-like
Z	<i>LOC107051888</i>	ERVK-25 Pol protein-like
Z	<i>LOC107051891</i>	ERVK-8 Pol protein-like
Z	<i>LOC107052380</i>	(LyK1) ERVK-25 Pol protein-like
ND ²	<i>LOC107049474</i>	ERVK-9 Pol protein-like

¹Chr; chromosome.²ND; chromosome was not defined.

Gag polyprotein genes and the remaining were Pol protein genes. Twenty-two chMB-ERVKs showed significant transcription levels (RPKM > 0.1). Sixteen of these were transcribed to a strikingly higher degree in layer WL myoblasts (named LyK1–LyK16) and two were significantly upregulated in broiler UKC myoblasts (named BrK1 and BrK2) on at least one of the differentiation days (FDR *p* value < 0.05) (Fig. 1A). This clearly indicates that some chMB-ERVKs were differentially transcribed between WL and UKC myoblasts throughout myogenic differentiation.

It has been reported that an intensive gene expression spreads into its physically neighboring genes to activate their expression. The radius of such an intergenic transcriptional ripple effect is estimated to be around 100 kb on the chromosome (Ebisuya *et al.*, 2008). To assess the influence of chMB-ERVK on its neighbor transcription, the loci of the 43 chMB-ERVKs and their neighboring genes on the *Gallus gallus* genome were carefully surveyed. One chMB-ERVK, *LOC107055731*, had no neighboring genes. However, the remaining 42 chMB-ERVKs had a total of 182 genes within

A



B

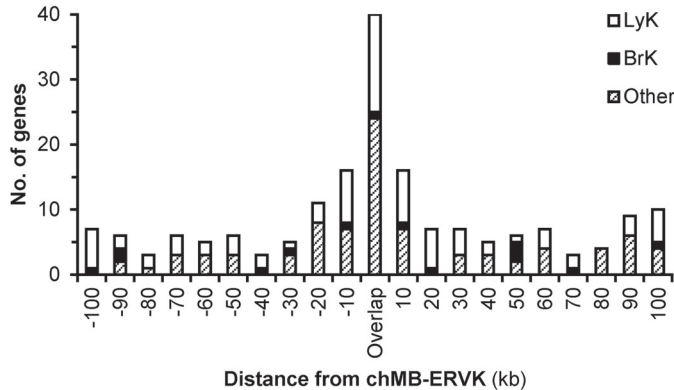


Fig. 1. **Transcription of ERVVs and neighboring genes in chicken myoblasts.** (A) Transcription levels (RPKM) of the 22 chMB-ERVVs with RPKM > 0.1 in WL and UKC myoblasts on differentiation days 0, 1, and 2. * FDR $p < 0.05$ vs UKC. † FDR $p < 0.05$ vs WL. (B) Total number of neighboring genes and distances from the 43 chMB-ERVVs (16 LyKs, two BrKs, and others) on the chicken chromosomes.

± 100 kb. Remarkably, 72 of the 182 genes (39.6%) were intensively located within ± 10 kb of chMB-ERVVs (Fig. 1B). In other words, 37/43 (86.0%) of chMB-ERVVs had been integrated within ± 10 kb of the genes. In humans, ERVVs are considered to be randomly and independently inserted into the genome (Kurdyukov *et al.*, 2001). In chickens, ERVVs are enriched in gene-sparse chromosomes and intergenic regions. Indeed, only 11.2% of chicken ERVVs are present in active transcription units (Barr *et al.*, 2005). Contrary to such propensities of ERVVs, most of the chMB-ERVVs had neighboring genes. It is thought that the ERVVs

integrated near the genes have remained to be transcribed in chicken myoblasts.

A total of 182 chMB-ERVV-neighboring genes comprised 158 unique genes (Supplemental Table), of which 96 (60.8%) encoded proteins, and 44 (27.8%) were non-coding (Table 2). Intriguingly, the 96 coding genes included 15 ERVVs and four envelope genes, indicating that chMB-ERVVs form clusters with ERVVs on the chromosomes. For instance, LyK2 and LyK10 were overlapping each other. Another major class of chMB-ERVV-neighboring genes was olfactory receptors but almost all of them (20/22) were

LyK4-neighboring genes. The chicken genome has at least 554 olfactory receptor genes (Niimura and Nei, 2005), and they form genomic clusters on chromosomes as a result of duplication (International Chicken Genome Sequencing Consortium, 2004). LyK4 was perhaps accidentally integrated into the loci associated with an olfactory receptor gene cluster. This would explain why the olfactory receptor became one of the main components of chMB-ERVK-neighboring genes.

Transcription of chMB-ERVK-neighboring Genes in Chicken Myoblasts

It was subsequently investigated whether chMB-ERVKs affect the transcription of their neighboring genes. RNA-seq

data showed that the expression levels of all 79 neighboring genes of 16 LyKs were not significantly different between WL and UKC myoblasts except for a combination of LyK2 and LyK10 as the neighboring genes of each other. LyKs that highly expressed in WL myoblasts are considered not to influence their neighboring gene transcription.

Next, the expression levels of a total of 16 neighboring genes of BrK1 and BrK2 were dissected (Table 3). BrK1 had two neighboring genes, *LOC107052719* and *FAM19A2* (Fig. 2A). *LOC107052719* expression was significantly suppressed in UKC myoblasts compared to those in WL myoblasts at early stages of myogenic differentiation both in RPKM and TPM values (Fig. 2C). It is of interest that BrK1 and *LOC107052719* were overlapping on the identical strand and showed opposite transcription patterns. Conversely, *FAM19A2* was strikingly upregulated in UKC myoblasts throughout differentiation (Fig. 2D), which is in accordance with the BrK1 transcription.

BrK2 had 14 neighboring genes, of which seven were located upstream, five were located downstream, and two were overlapping with BrK2 (Fig. 2B). RNA-seq data showed that expression levels of BrK2-neighboring genes except for *LOC772243* were not different between WL and UKC myoblasts. Only the expression of *LOC772243* was significantly downregulated in UKC myoblasts as opposed to BrK2 (Fig. 2E). *LOC772243* encodes for a murine leukemia virus (MLV)-related proviral envelope polyprotein-like protein. MLV originally exists as an exogenous virus and serves as an ERV like ERVK (Stocking and Kozak, 2008). The BrK2 and *LOC772243*, which overlap on the same strand and negatively regulate the transcription of one another, might be an important instance of retrovirus-mediated gene expression in host cells.

Table 2. Classification of chMB-ERVK-neighboring genes

Class	No. of genes	(%)
Coding gene		
Characterized endogenous gene (except for olfactory receptor)	28	17.7%
Olfactory receptor	22	13.9%
ERVK	15	9.5%
Envelope glycoprotein gp95-like	4	2.5%
Predicted ORF ¹	27	17.1%
Non-coding gene		
lincRNA ²	38	24.1%
miRNA ³	5	3.2%
snRNA ⁴	1	0.6%
Uncharacterized gene	18	11.4%

¹ ORF; open reading frame.

² lincRNA; long intergenic non-coding RNA.

³ miRNA; microRNA.

⁴ snRNA; small nuclear RNA.

Table 3. BrK1- and BrK2-neighboring genes

Gene	Product	Distance (bp)
BrK1: <i>LOC107052718</i>		
<i>LOC107052719</i>	Uncharacterized gene	Overlap
<i>FAM19A2</i>	Family with sequence similarity 19 member A2, C-C motif chemokine-like	93,542
BrK2: <i>LOC107054036</i>		
<i>SLC1A7</i>	Solute carrier family 1 member 7	-92,653
<i>PODN</i>	Podcan	-61,675
<i>SCP2</i>	Sterol carrier protein 2	-49,303
<i>LOC107054037</i>	Uncharacterized gene	-46,114
<i>ECHDC2</i>	Enoyl-CoA hydratase domain containing 2	-43,020
<i>COA7</i>	Cytochrome c oxidase assembly factor 7	-18,667
<i>MIR6549</i>	miRNA	-1,396
<i>LOC772243</i>	Envelope glycoprotein gp95-like	Overlap
<i>ZYG11B</i>	Zyg-11 family member B, cell cycle regulator	Overlap
<i>FAM159A</i>	Family with sequence similarity 159 member A	2,405
<i>GPX7</i>	Glutathione peroxidase 7	27,744
<i>ZCCHC11</i>	Zinc finger CCHC-type containing 11	39,929
<i>PRPF38A</i>	Pre-mRNA processing factor 38A	84,605
<i>ORC1</i>	Origin recognition complex subunit 1	89,080

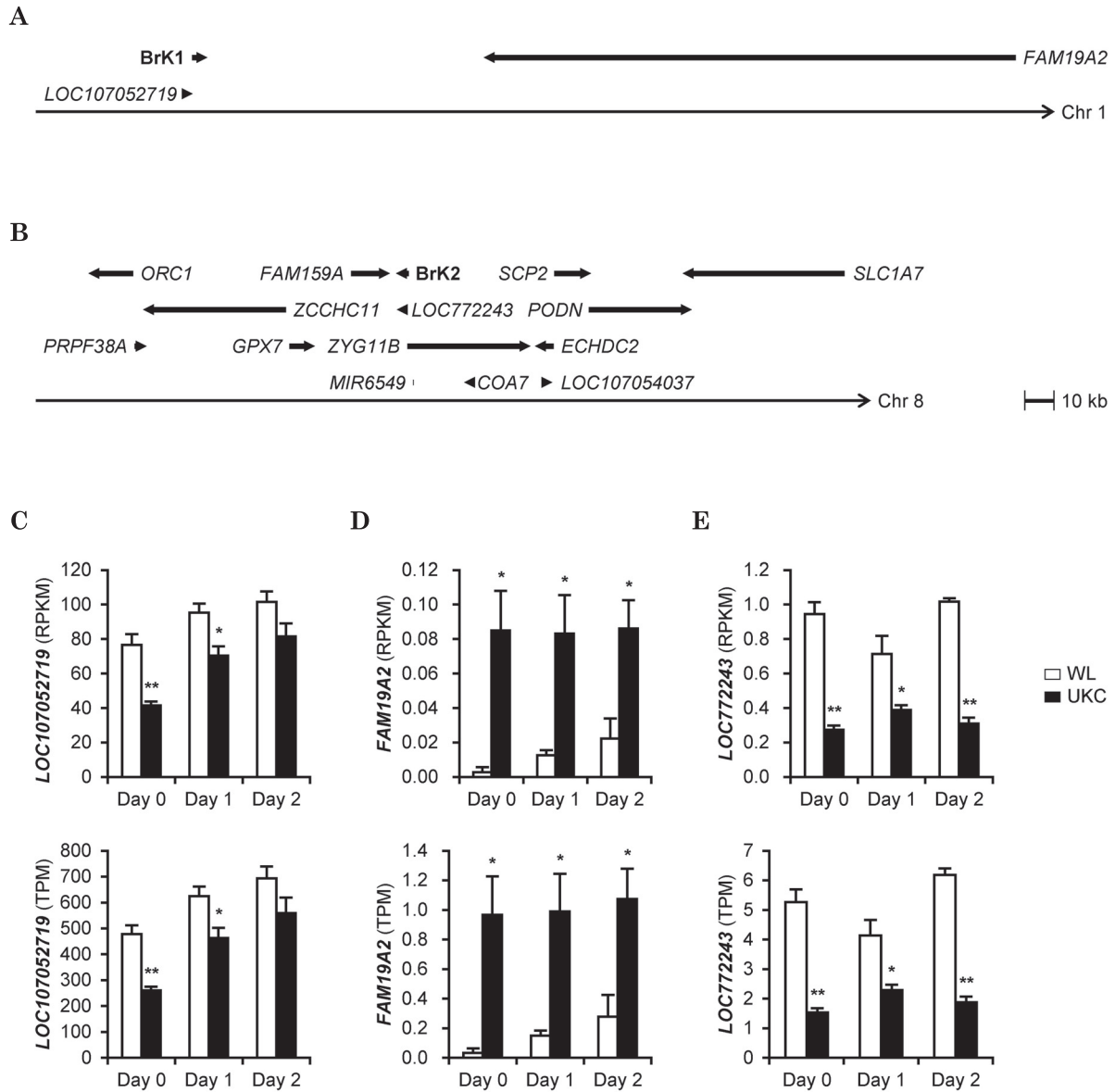


Fig. 2. Loci and transcription levels of BrK1- and BrK2-neighboring genes. (A and B) Physical maps of BrK1, BrK2, and their neighboring genes. Arrows indicate sequence direction (5'-3'). (A) BrK1 and its neighboring genes on chromosome 1. (B) BrK2 and its neighboring genes on chromosome 8. (C-E) The transcription levels defined by RPKM and TPM of *LOC107052719* (C), *FAM19A2* (D), and *LOC772243* (E) in WL and UKC myoblasts on differentiation days 0, 1, and 2. * $p < 0.05$, ** $p < 0.01$ vs WL (Student's *t* test at each day).

Discussion

This is the first report of the ERVKs that are transcribed in chicken skeletal muscle myoblasts. Generally in host cells, most ERVs are silenced by chromatin modifications such as hypermethylation (Schulz *et al.*, 2006). It is predicted that the 43 chMB-ERVks identified in this study are still epigenetically active in chicken myoblasts, which is supported by the fact that almost all chMB-ERVks are locating near

genes. Eighteen chMB-ERVks were found to be differentially transcribed between WL and UKC myoblasts during differentiation. Previous studies have reported that the myoblast genome is epigenetically modified during myogenic differentiation (Blum, 2014). We recently showed that proliferating and differentiating abilities are different between WL and UKC myoblasts (Nihashi *et al.*, 2019b), suggesting that the epigenome linking myoblast properties may also differ among the chickens. Genome-wide epigenetic com-

parison among the breeds will give an insight into the relationship between chromatin state and chMB-ERVK transcription.

chMB-ERVVs were physically associated with endogenous genes on chicken chromosomes despite the fact that ERVs tend to be integrated into intergenic regions (Barr *et al.*, 2005). This suggests that the loci having chMB-ERVVs are active in myoblasts. However, there is no information about the roles of chMB-ERVK-neighboring genes in the MRF-driven myogenic program. At present, the transcription of chMB-ERVVs and their neighboring genes seems not to be directly involved in the principal myogenic process regulated by MRFs. The functions of chMB-ERVK-neighboring genes during myoblast differentiation should be clarified in the future.

RNA-seq data showed that the transcription levels of LyK-neighboring genes were not different between WL and UKC myoblast while both BrK1-neighboring genes (*LOC107052719* and *FAM19A2*) and one BrK2-neighboring gene (*LOC772243*) were differentially expressed among the breeds. BrK1 and BrK2 displayed the first and second highest expression levels among chMB-ERVVs. Transcriptional ripple effect occurs at the locus intensively transcribed (Ebisuya *et al.*, 2008). This could be one of the reasons why BrK1 and BrK2 seemed to affect the expression of their neighboring genes.

FAM19A2 was strongly upregulated in UKC myoblasts in parallel with BrK1. *FAM19A2* (*TAF2* in human) encodes for the chemoattractant neurokinine potently expressed in the brain (Tom Tang *et al.*, 2004). A recent study reported that *FAM19A2* protein enhances the migration, recruitment, and proliferation of human mesenchymal stem cells without affecting differentiation into osteocytes and adipocytes (Jafari *et al.*, 2019). Mesenchymal stem cells can also differentiate into muscle cells. Therefore, it is possible that the myoblast-secreting *FAM19A2* protein recruits mesenchymal stem cells to muscle tissue during muscle development and growth in broilers. RNA-seq data obtained in this study is still insufficient to conclude that BrK1 transcription actually enhances *FAM19A2* expression. Although the transcriptional regulatory role of *FAM19A2* has not been studied, *FAM19A2* locus was reported to be clinically associated with various pathological conditions such as intellectual disability (Borg *et al.*, 2005), systemic lupus erythematosus (Lessard *et al.*, 2012), vital capacity (Parker *et al.*, 2014), and insulin sensitivity (Walford *et al.*, 2016) in humans. These studies suggest that *FAM19A2* expression can be altered by a variety of biological factors and processes. To clarify the relationship between BrK1 and *FAM19A2* transcription, the role of *FAM19A2* in myoblasts need to be further investigated.

The transcription levels of *LOC107052719* and *LOC772243* were significantly suppressed in UKC myoblasts in contrast to those of BrK1 and BrK2. *LOC107052719* and *LOC772243* were overlapping with BrK1 and BrK2 on the identical strands, respectively. Almost of such paired overlapping genes are expressed in parallel. However, a part of overlapping gene pairs exhibit the expression pattern that one is induced and another is suppressed in specific tissues (Chen *et al.*,

2019). BrK1/*LOC107052719* and BrK2/*LOC772243* may also be the gene pairs that are differentially expressed in UKC myoblasts. The transcription of another BrK2-overlapping gene, *ZYG11B*, was not altered between WL and UKC myoblasts. *ZYG11B* was encoded on the antisense strand of BrK2, and its overlapping region occupied only a small portion of the *ZYG11B* gene structure. The transcription level of BrK2, which was similar to those of LyKs might be inadequate to modulate the expression of its neighboring genes including *ZYG11B*.

In any case, the causality of transcription levels of chMB-ERVVs and their neighboring genes so far is ambiguous. To address this issue, the transcription of chMB-ERVVs in myoblasts need to be carefully investigated in various situations because transcriptional regulation in myoblasts can be different depending on developmental origin or phase. For example in mice, regulatory systems of MRFs are distinct between facial and trunk muscles (Lu *et al.*, 2002; Moncaut *et al.*, 2012). It is possible that the transcription of ERVVs in myoblasts may also differ between breeds, growth stages, or muscle tissues. The myoblasts used in this study were initially isolated from hindlimb muscles of chicken embryos. It has not been strictly compared yet whether cellular properties of chicken myoblasts are distinctive for muscle tissues. The broiler breast muscle, which is mostly composed of fast-type myofibers frequently shows acute and chronic inflammatory and myopathic changes (Kuttappan *et al.*, 2013; Hosotani *et al.*, 2020). Intriguingly, ERVK transcription is known to be activated by pathogenic transcription factors such as NF- κ B and interferon regulatory factors (Manghera and Douville, 2013). These findings suggest that ERVVs in myoblasts would be induced by extracellular factors or the environment. The transcription of chMB-ERVVs and their neighboring genes *in vivo* should be elucidated in further studies.

This study demonstrated that chicken myoblasts transcribe ERVVs, that transcription levels of these chMB-ERVVs are different among the breeds, and that chMB-ERVK transcription may be linked to the expression of their neighboring genes. These data suggest that chMB-ERVK-mediated gene expression may be involved in the molecular characteristics of myoblasts.

Acknowledgments

This study was supported in part by Grants-in-Aid from The Ito Foundation and from The Shinshu Foundation for Promotion of Agricultural and Forest Science to T. T., a Research Fellowship for Young Scientists from The Japan Society for the Promotion of Science (19J20888) and a Grant-in-Aid from The Fund of Nagano Prefecture to Promote Scientific Activity (H-29-3-12) to Y. N., and Grants-in-Aid from The Japan Society for the Promotion of Science to T. T. (19K05948), T. O. (16K07986), and H. K. (18K05939). RNA-seq data analysis was supported by the Cooperative Research Grant of the Genome Research for BioResource, NODAI Genome Research Center, Tokyo University of Agriculture.

Disclosure Statement

The authors declare no conflict of interest.

References

- Barr SD, Leipzig J, Shinn P, Ecker JR and Bushman FD. Integration targeting by avian sarcoma-leukosis virus and human immunodeficiency virus in the chicken genome. *Journal of Virology*, 79: 12035–12044. 2005.
- Blum R. Activation of muscle enhancers by MyoD and epigenetic modifiers. *Journal of Cellular Biochemistry*, 115: 1855–1867. 2014.
- Borg K, Stankiewicz P, Bocian E, Kruczek A, Obersztyn E, Lupski JR and Mazurczak T. Molecular analysis of a constitutional complex genome rearrangement with 11 breakpoints involving chromosomes 3, 11, 12, and 21 and a approximately 0.5-Mb submicroscopic deletion in a patient with mild mental retardation. *Human Genetics*, 118: 267–275. 2005.
- Chen CH, Pan CY and Lin WC. Overlapping protein-coding genes in human genome and their coincidental expression in tissues. *Scientific Reports*, 9: 13377. 2019.
- Downey RF, Sullivan FJ, Wang-Johanning F, Ambs S, Giles FJ and Glynn SA. Human endogenous retrovirus K and cancer: Innocent bystander or tumorigenic accomplice? *International Journal of Cancer*, 137: 1249–1257. 2015.
- Dumont NA, Bentzinger CF, Sincennes MC and Rudnicki MA. Satellite cells and skeletal muscle regeneration. *Comprehensive Physiology*, 5: 1027–1059. 2015.
- Ebisuya M, Yamamoto T, Nakajima M and Nishida E. Ripples from neighbouring transcription. *Nature Cell Biology*, 10: 1106–1113. 2008.
- Fuchs NV, Kraft M, Tondera C, Hanschmann KM, Löwer J and Löwer R. Expression of the human endogenous retrovirus (HERV) group HML-2/HERV-K does not depend on canonical promoter elements but is regulated by transcription factors Sp1 and Sp3. *Journal of Virology*, 85: 3436–3448. 2011.
- Gifford RJ, Blomberg J, Coffin JM, Fan H, Heidmann T, Mayer J, Stoye J, Tristem M and Johnson WE. Nomenclature for endogenous retrovirus (ERV) loci. *Retrovirology*, 15: 59. 2018.
- Hosotani M, Kawasaki T, Hasegawa Y, Wakasa Y, Hoshino M, Takahashi N, Ueda H, Takaya T, Iwasaki T and Watanabe T. Physiological and pathological mitochondrial clearance is related to pectoralis major muscle pathogenesis in broilers with wooden breast syndrome. *Frontiers in Physiology*, 11: 579. 2020.
- Huda A, Polavarapu N, Jordan IK and McDonald JF. Endogenous retroviruses of the chicken genome. *Biology Direct*, 3: 9. 2008.
- International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432: 695–716. 2004.
- Jafari A, Isa A, Chen L, Ditzel N, Zaher W, Harkness L, Johnsen HE, Abdallah BM, Clausen C and Kassem M. TAF2A2 induces skeletal (stromal) stem cell migration through activation of Rac1-p38 signaling. *Stem Cells*, 37: 407–416. 2019.
- Knössl M, Löwer R and Löwer J. Expression of the human endogenous retrovirus HTDV/HERV-K is enhanced by cellular transcription factor YY1. *Journal of Virology*, 73: 1254–1261. 1999.
- Kurdyukov SG, Lebedev YB, Artamonova II, Gorodentseva TN, Batrak AV, Mamedov IZ, Azhikina TL, Legchilina SP, Efimenko IG, Gardiner K and Sverdlov ED. Full-sized HERV-K (HML-2) human endogenous retroviral LTR sequences on human chromosome 21: map locations and evolutionary history. *Gene*, 273: 51–61. 2001.
- Kuttappan VA, Shivaprasad HL, Shaw DP, Valentine BA, Hargis BM, Clark FD, McKee SR, and Owens CM. Pathological changes associated with white striping in broiler breast muscles. *Poultry Science*, 92: 331–338. 2013.
- Lessard CJ, Adrianto I, Ice JA, Wiley GB, Kelly JA, Glenn SB, Adler AJ, Li H, Rasmussen A, Williams AH, Ziegler J, Comeau ME, Marion M, Wakeland BE, Liang C, Ramos PS, Grundahl KM, Gallant CJ, Alarcón-Riquelme ME, Alarcón GS, Anaya JM, Bae SC, Boackle SA, Brown EE, Chang DM, Cho SK, Criswell LA, Edberg JC, Freedman BI, Gilkeson GS, Jacob CO, James JA, Kamen DL, Kimberly RP, Kim JH, Martin J, Merrill JT, Niewold TB, Park SY, Petri MA, Pons-Estel BA, Ramsey-Goldman R, Reveille JD, Scofield RH, Song YW, Stevens AM, Tsao BP, Vila LM, Vyse TJ, Yu CY, Guthridge JM, Kaufman KM, Harley JB, Wakeland EK, Langefeld CD, Gaffney PM, Montgomery CG, Moser KL; BIOLUPUS Network; GENLES Network. Identification of IRF8, TMEM39A, and IKZF3-ZBP2 as susceptibility loci for systemic lupus erythematosus in a large-scale multiracial replication study. *American Journal of Human Genetics*, 90: 648–660. 2012.
- Lu JR, Bassel-Duby R, Hawkins A, Chang P, Valdez R, Wu H, Gan L, Shelton JM, Richardson JA and Olson EN. Control of facial muscle development by MyoR and capsulin. *Science*, 298: 2378–2381. 2002.
- Manghera M and Douville RN. Endogenous retrovirus-K promoter: a landing strip for inflammatory transcription factors? *Retrovirology*, 10: 16. 2013.
- Manghera M, Ferguson J and Douville R. Endogenous retrovirus-K and nervous system diseases. *Current Neurology and Neuroscience Reports*, 14: 488. 2014.
- Manghera M, Ferguson-Parry J, Lin R and Douville RN. NF- κ B and IRF1 induce endogenous retrovirus K expression via interferon-stimulated response elements in its 5' long terminal repeat. *Journal of Virology*, 90: 9338–9349. 2016.
- Moncaut N, Cross JW, Siligan C, Keith A, Taylor K, Rigby PW and Carvajal JJ. Myosin and TCF21 coordinate the maintenance of myogenic regulatory factor expression levels during mouse craniofacial development. *Development*, 139: 958–967. 2012.
- Nihashi Y, Ono T, Kagami H and Takaya T. Toll-like receptor ligand-dependent inflammatory responses in chick skeletal muscle myoblasts. *Developmental and Comparative Immunology*, 91: 115–122. 2019a.
- Nihashi Y, Umezawa K, Shinji S, Hamaguchi Y, Kobayashi H, Kono T, Ono T, Kagami H and Takaya T. Distinct cell proliferation, myogenic differentiation, and gene expression in skeletal muscle myoblasts of layer and broiler chickens. *Scientific Reports*, 9: 16527. 2019b.
- Niimura Y and Nei M. Evolutionary dynamics of olfactory receptor genes in fishes and tetrapods. *Proceedings of the National Academy of Sciences of the United States of America*, 102: 6039–6044. 2005.
- Parker MM, Foreman MG, Abel HJ, Mathias RA, Hetmanski JB, Crapo JD, Silverman EK, Beaty TH; COPDGene Investigators. Admixture mapping identifies a quantitative trait locus associated with FEV₁/FVC in the COPDGene Study. *Genetic Epidemiology*, 38: 652–659. 2014.

- Scheuermann GN, Bilgili SF, Tuzun S and Mulvaney DR. Comparison of chicken genotypes: myofiber number in pectoralis muscle and myostatin ontogeny. *Poultry Science*, 83: 1404–1412. 2004.
- Schulz WA, Steinhoff C and Florl AR. Methylation of endogenous human retroelements in health and disease. *Current Topics in Microbiology and Immunology*, 310: 211–250. 2006.
- Stauffer Y, Theiler G, Sperisen P, Lebedev Y and Jongeneel CV. Digital expression profiles of human endogenous retroviral families in normal and cancerous tissues. *Cancer Immunity*, 4: 2. 2004.
- Stocking C and Kozak CA. Murine endogenous retroviruses. *Cellular and Molecular Life Sciences*, 65: 3383–3398. 2008.
- Takaya T, Nihashi Y, Kojima S, Ono T and Kagami H. Autonomous xenogenic cell fusion of murine and chick skeletal muscle myoblasts. *Animal Science Journal*, 88: 1880–1885. 2017.
- Tom Tang Y, Emtage P, Funk WD, Hu T, Arterburn M, Park EE and Rupp F. TAFE: a novel secreted family with conserved cysteine residues and restricted expression in the brain. *Genomics*, 83: 727–734. 2004.
- Walford GA, Gustafsson S, Rybin D, Stančáková A, Chen H, Liu CT, Hong J, Jensen RA, Rice K, Morris AP, Mägi R, Tönjes A, Prokopenko I, Kleber ME, Delgado G, Silbernagel G, Jackson AU, Appel EV, Grarup N, Lewis JP, Montasser ME, Landenvall C, Staiger H, Luan J, Frayling TM, Weedon MN, Xie W, Morcillo S, Martínez-Larrad MT, Biggs ML, Chen YD, Corbaton-Anchuelo A, Færch K, Gómez-Zumaquero JM, Goodarzi MO, Kizer JR, Koistinen HA, Leong A, Lind L, Lindgren C, Machicao F, Manning AK, Martín-Núñez GM, Rojo-Martínez G, Rotter JI, Siscovick DS, Zmuda JM, Zhang Z, Serrano-Rios M, Smith U, Soriguer F, Hansen T, Jørgensen TJ, Linnenberg A, Pedersen O, Walker M, Langenberg C, Scott RA, Wareham NJ, Fritsche A, Häring HU, Stefan N, Groop L, O’Connell JR, Boehnke M, Bergman RN, Collins FS, Mohlke KL, Tuomilehto J, März W, Kovacs P, Stumvoll M, Psaty BM, Kuusisto J, Laakso M, Meigs JB, Dupuis J, Ingelsson E and Florez JC. Genome-wide association study of the modified Stumvoll Insulin Sensitivity Index identifies BCL2 and FAM19A2 as novel insulin sensitivity loci. *Diabetes*, 65: 3200–3211. 2016.