ORIGINAL ARTICLE

Cell Proliferation WILEY

Transcriptional states and chromatin accessibility during bovine myoblasts proliferation and myogenic differentiation

Qian Li | Yahui Wang | Xin Hu | Yapeng Zhang | Hongwei Li | Qi Zhang | Wentao Cai | Zezhao Wang | Bo Zhu | Lingyang Xu | Xue Gao | Yan Chen | Huijiang Gao | Junya Li | Lupei Zhang

Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, Key Laboratory of Animal (Poultry) Genetics Breeding and Reproduction, Ministry of Agriculture and Rural Affairs, Beijing, China

Correspondence
Junya Li and Lupei Zhang, Institute of Animal Science, Chinese Academy of Agricultural Sciences, Key Laboratory of Animal (Poultry) Genetics Breeding and Reproduction, Ministry of Agriculture and Rural Affairs, Beijing 100193, China.
Email: lijunya@caas.cn (J. L) and zhanglupei@caas.cn (L. Z.)

Abstract

Objectives: Although major advances have been made in bovine epigenome study, the epigenetic basis for fetal skeletal muscle development still remains poorly understood. The aim is to recapitulated the time course of fetal skeletal muscle development *in vitro*, and explore the dynamic changes of chromatin accessibility and gene expression during bovine myoblasts proliferation and differentiation.

Methods: PDGFR- cells were isolated from bovine fetal skeletal muscle, then cultured and induced myogenic differentiation *in vitro* in a time-course study (P, D0, D2, and D4). The assay for transposase-accessible chromatin sequencing (ATAC-seq) and RNA sequencing (RNA-seq) were performed.

Results: Among the enriched transcriptional factors with high variability, we determined the effects of *MAFF*, *ZNF384*, and *KLF6* in myogenesis using RNA interference (RNAi). In addition, we identified both stage-specific genes and chromatin accessibility regions to reveal the sequential order of gene expression, transcriptional regulatory, and signal pathways involved in bovine skeletal muscle development. Further investigation integrating chromatin accessibility and transcriptome data was conducted to explore cis-regulatory regions in line with gene expression. Moreover, we combined bovine GWAS results of growth traits with regulatory regions defined by chromatin accessibility, providing a suggestive means for a more precise annotation of genetic variants of bovine growth traits.

Conclusion: Overall, these findings provide valuable information for understanding the stepwise regulatory mechanisms in skeletal muscle development and conducting beef cattle genetic improvement programs.

1 | INTRODUCTION

Skeletal muscle, which accounts for about 40% of total body weight,[1] is a highly dynamic tissue of the body. The growth and maintenance of skeletal muscle play an important role in crucial

---

Qian Li and Yahui Wang contributed equally to this work and shared the first authorship.

physical and metabolic function in human, while in animal production, affecting growth efficiency and meat quality.[2] The growth of skeletal muscle fibres is initiated during embryonic stage,[3] then is fixed during the foetal stage, and remains almost unchanged in number after birth.[4] Notably, the foetal stage is the critical period of myogenesis, when the majority of skeletal muscle fibres form.[5] It has become established that many specific signalling pathways and transcription factors (TFs) are involved in myogenesis.[5–8] Wingless and Int (Wnt), paired box 3 (*Pax3*), and *Pax7* are necessary myogenic regulatory factors (*MRFs*) that regulate the myogenic lineage differentiation of mesenchymal stem cells during embryo stage.[5,9–12] Although previous studies provide strong evidence of the contribution by *MRFs* to skeletal muscle development,[13,14] the precise mechanisms underlying skeletal muscle development remain unclear. Halstead et al. reported the dynamic changes of chromatin accessibility during bovine pre-implantation development, indicating their important roles in species-specific embryonic genome activation.[15] In addition, a comprehensive functional annotation of bovine rumen epithelial cells has been well established and provided important information to explore butyrate-induced biological effects in bovine.[16] As Functional Annotation of Animal Genomes (FAANG) Consortium proposed, additional efforts have been made to investigate chromatin accessibility related to skeletal muscle development.[17–19] However, particularly in bovine models, previous studies in the context of skeletal muscle chromatin accessibility have mainly focused on comparative analysis across distant species to identify tissue-specific regulatory elements.[20–22] Recently, Cao et al. described comparative analysis of enhancers between adult and embryo bovine muscle using chromatin accessibility.[23] Despite the foetal stage is critical for skeletal muscle development, only limited evidence of foetal muscle growth and development was available from prior studies, and these findings may ignore that skeletal muscle development is a dynamic process comprised of multiple developmental stages. In our study, we reconstructed the course of myogenic differentiation in vitro, providing an ideal model to define when and how these regulatory changes occur during skeletal muscle development.

2 | MATERIALS AND METHODS

2.1 | Primary cells isolation and culture

The myoblasts were enzymatically isolated from longissimus dorsi tissues of bovine foetuses at 90 days, and cultured in low-glucose DMEM with 10% foetal bovine serum (growth medium) as described previously.[24] At 90% confluence, the cells were trypsinized with 0.25% trypsin–EDTA (Gibco, Grand Island, NY) and passaged to culture plates. At 100% confluence, the growth medium was exchanged by low-glucose DMEM with 5% horse serum (differentiation medium). The differentiation medium was then changed every 2 days.

2.2 | RNA extraction, reverse transcription and quantitative real-time PCR

Total RNA was isolated using the TRIzol reagent (Invitrogen Life Technologies). RNA concentration and quality were determined by NanoPhotometer N50 (Implen, Munich, Germany). Next, the qRT-PCR was performed using ABI QuantStudio 7 Flex system (Life, Carlsbad, CA) in accordance with the instructions of KAPA SYBR® FAST qPCR Kit (KAPABiosystems, Wilmington, MA). The primer sequences involved in qRT-PCR were listed in Table S1.

2.3 | siRNA transfection

Small interfering RNA (siRNA) were supplied by RiboBio (Guangzhou, China), and the sequences are listed in Table S2. With the help of Lipofectamine™ RNAiMAX (Invitrogen Life Technologies, Carlsbad, CA) reagent, the siRNAs against *MAFF*, *ZNF384*, *KLF6* (*siMAFF*, *siZNF384*, *siKLF6*) or siRNA control (NC) were transfected into myoblasts at a final concentration of 50 nM.

2.4 | CCK8 assay

After interference of *MAFF* in myoblasts, CCK8 reagent was added at 12, 24, 36, 48 and 60 h. After continuously incubating for 2 h, the absorbance (OD values) at 450 nm was then measured.

2.5 | mRNA-seq library construction, sequencing and data processing

mRNAs were purified from total RNAs using oligo-dT beads and then fragmented with Mg^{2+}. RNA-seq libraries were constructed by following steps: reverse transcription, the end repair of cDNA, poly(A) tail and index addition. All libraries were sequenced on the HiSeq2500 platform following a PE150 strategy. RNA-seq datasets from each sequencing library were trimmed with Trim_Galore 0.6.3 to remove Illumina adapter sequences and low-quality sequences. Trimmed reads were then aligned to bovine genome (ARS-UCD1.2) using STAR 2.7.3a.[25] Gene expression counts were calculated by the –quantMode GeneCounts functionality in STAR. For downstream analysis, RNA-seq raw counts were normalized by computing counts per million and then formed to Z scores. Differential gene expression analysis was performed by DESeq2 (adjusted *p*-value <0.01).[26,27]

2.6 | ATAC-seq library construction, sequencing and data processing

The ATAC-seq libraries were performed as TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme, Nanjing, China). Briefly, 5×10^4 cells were

counted and resuspended in precooled lysis buffer. The unfixed nucleuses were incubated with Tn5 transposase at 37°C for 30 min. Amplified the fragments for 16 cycles and purified using the VAHTS RNA Clean Beads (Vazyme, Nanjing, China). Libraries were quantitated using Qubit 4 Fluorometer (Invitrogen, Singapore), and then sequenced on Illumina HiSeq2500 platform following a PE150 strategy. ATAC-seq datasets from each sequencing library were first trimmed with Trim Galore 0.6.3 to remove Illumina adapter sequences and low-quality sequences. Trimmed reads were then aligned to bovine genome (ARS-UCD1.2) using Bowtie2 2.3.5.[28] ATAC-seq peak calling for each individual replicate was performed by MACS2 2.1.2.[29] To establish a common peak set, peak summits were extended ±250 bps to obtain 500 bp fixed-width peaks.[30] Overlapping peaks in each sample were handled using an iterative removal procedure with the strongest signal ($-\log_{10}$ FDR) until all peaks were accounted. FeatureCounts[31] was used to quantify the raw read counts from the common peak set including all replicates. For downstream analysis, ATAC-seq raw counts were normalized by computing counts per million and then formed to Z scores. Differential chromatin accessibility analysis was performed by DESeq2 (adjusted p-value < 0.01 and |log2FC| > 2). Peak annotation was performed by R package ChIPseeker.[32]

2.7 | Clustering and visualization of ATAC-seq and RNA-seq signal

For clustering, the normalized counts of all genes and the top 50,000 open chromatin peaks (ranked by adjusted p-value) were used for principal component analysis (PCA) and Spearman correlation analysis using prcomp_irlba and Complexheatmap.[33] The coverage tracks were generated by deeptools 3.3.0 bamCoverage.[34] Visualization of ATAC-seq and RNA-seq coverage track was performed by IGV 2.6.2.[35] Each read was extended by 250 bp and the genome-wide signal was generated by the normalized read counts (CPM) per bin. TF deviation and variability analysis were performed via the motifmatchr package, and CisBP transcription factors from the 'human_pwms_v2' dataset.[36] Transcription factor activity scoring was performed using the 501-bp fixed-width peak set, which combined replicates for each differentiated time point.

2.8 | Identification of stage-specific genes and stage-specific peaks

We used a Shannon-entropy-based to identify differentiation stage-specific genes/peaks as previously described.[37,38] We first selected those with entropy scores less than a predefined threshold (stage-specific genes: 1.6; stage-specific peaks: 1.8) as candidates. Then the ATAC-seq peak was highest active in this stage (normalized CPM > 1), and its high activity (normalized CPM > 0) could not be observed at more than two additional stages. The enrichment of known motifs for stage-specific peaks was detected by findMotifsGenome.pl function of the HOMER package.[39] After filtering out motifs of TFs which were not expressed in all stages ($MAX_{(CPM)} < 1$), the top enriched known motifs were reported.

2.9 | Functional annotation enrichment analysis

Gene Ontology (GO) enrichment analysis for the tested genes was performed by DAVID V 6.8.[40,41] KEGG pathway enrichment was conducted by KOBAS 3.0.[42]

2.10 | Integration of ATAC-seq and RNA-seq

Based on genomic region, consensus peaks were classified into six groups: Promoter (TSS ±2.5 kb), 5′UTR, 3′UTR, downstream, exon, intergenic and intron. Spearman's correlation coefficients were calculated between gene expression levels (normalized CPM) and chromatin accessibility levels (normalized CPM) of each genomic region group. We further defined genes/peaks based on the previous definition[43]: Genes defined as HA/HE that maximum value of ATAC-seq/RNA-seq reads was higher than the 70th percentile; Genes/peaks were defined as MA/ME that maximum value of ATAC-seq/RNA-seq reads was below the 50th percentile. Each gene associated with promoter-region accessibility was assigned to four groups: HA–HE, high accessibility/high expression; MA–ME, medium-low accessibility/medium-low expression; HA–ME, high accessibility/medium-low expression; and MA–HE, medium-low accessibility/high expression.

2.11 | Enrichment analysis based on empirical sampling

Overall, 3804 human housekeeping genes were converted successfully to 3256 bovine orthologous housekeeping genes using the BioMart tool.[44,45] Enrichment analysis was conducted on housekeeping genes enriched in HA–MA group and gene set randomly sampled 10,000 times. The empirical p-value was calculated as:

$$p-\text{value}\frac{1+\sum_{i=1}^{10,000}\left(\text{number of }\left(\text{proportion}_{(\text{random sample})} > \text{proportion}_{(\text{HA−MA group})}\right)\right)}{10,001}.$$

QTL regions and associated SNP that affect cattle anatomy and growth were derived from Animal QTL Database (Cattle QTLdb: https://www.animalgenome.org/cgi-bin/QTLdb/BT/index).

2.12 | Genome-wide association analyses and GWAS enrichment analysis

A total of 1233 Chinese Simmental beef cattle were derived from Ulgai, Xilingol League, and Inner Mongolia, China from 2008 to 2015 as previously described.[46,47] Phenotypic data including carcass weight, average daily gain, liveweight, dressing percentage, meat percentage and pure meat weight were collected. DNA samples were genotyped on Illumina BovineHD 770 K SNP array, and then imputed to whole-genome sequence level.[47] Genome-Wide Association Analyses were performed by GCTA version 1.93.0.[48,49] The annotation of GWAS Summary Statistics was lifted over from UMD 3.1 to ARS-UCD1.2 via the UCSC liftOver tool.[50] To check whether the SNP effects were more enriched in chromatin peaks than background regions, GWAS signals enrichment analysis were
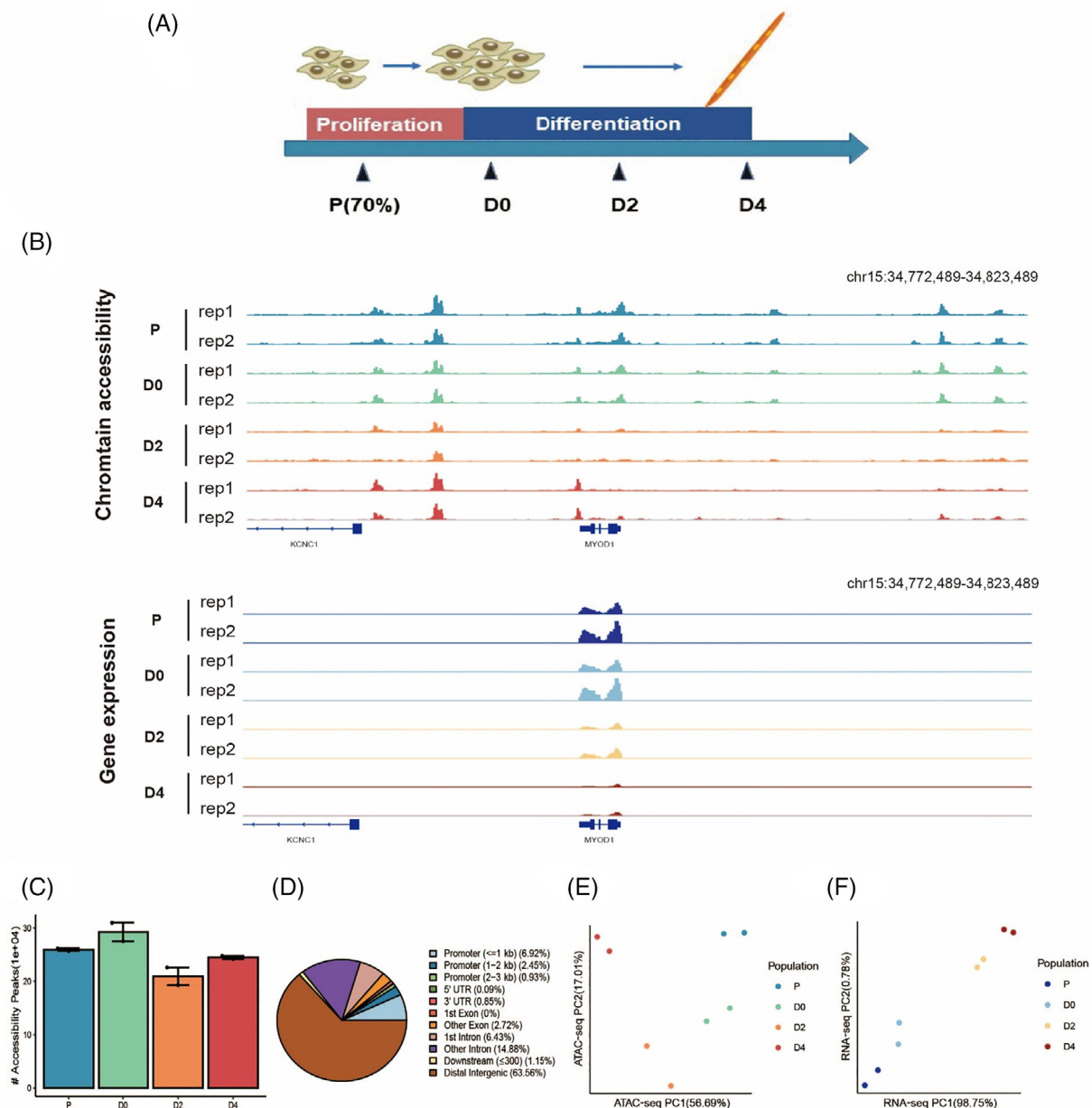

FIGURE 1 Comprehensive profiling of the chromatin accessibility and transcriptional landscape during bovine myoblast proliferation and myogenic differentiation. (A) Schematic illustration of myogenic differentiation *in vitro*. P, proliferation; D0, D2, D4, myogenic differentiation after 0, 2 and 4 days. (B) Normalized coverage (CPM) of chromatin accessibility and mRNA levels around *MYOD* gene for each time point. (C) Number of chromatin accessibility regions detected at P, D0, D2 and D4 time points. (D) Genomic distribution of chromatin accessibility peaks from the common peak set, namely promoter (TSS ± 2.5 kb), 5′UTR, 3′UTR, downstream, exon, intergenic and intron. (E) and (F) Principal component analysis (PCA) plot of chromatin accessibility and gene expression during myogenesis in two dimensions (PCA1 and PCA2). The normalized counts of all genes and the top 50,000 open chromatin peaks (ranked by $-\log_{10}$ FDR) were used for plot

performed using Perl scripts (sumGSE, https://github.com/WentaoCai/GWAS_enrichment).[51]

3 | RESULTS

3.1 | Comprehensive profiling of the chromatin accessibility and transcriptional landscape

To investigate the molecular control during bovine myoblast proliferation and myogenic differentiation, we isolated PDGFR⁻ cells from bovine foetal skeletal muscle, then cultured and induced myogenic differentiation *in vitro* in a time-course study. Next, we profiled the transcriptome and chromatin accessibility during bovine myogenesis at four important time points, including proliferation (P) and myogenic differentiation after 0, 2, and 4 days (D0, D2, and D4) (Figure 1A,B). Overall, we generated an average of 61.23 million uniquely mapped reads for ATAC-seq library and an average of 21.55 million mapped reads for RNA-seq library (Tables S3 and S4). Fragment size distribution displayed a clear nucleosomal periodicity in ATAC-seq and the expression pattern of major myogenic

regulatory factors in RNA-seq demonstrated the robustness of our dataset (Figure S1A,B). To enable direct comparison of peaks across different time points during bovine myogenesis, peaks called from all replicates (range 192,928 to 309,729) were pooled together and identified regions of open chromatin, resulting in a 501-bp fixed-width peak set with 379,553 accessible chromatin peaks (Figure 1C). To predict genomic features of functional regulatory elements, annotation of accessible regions was performed (Figure 1D). As expected, the majority of consensus peaks identified in the peak set were assigned to distal intergenic (63.56%), intronic (21.21%) and promoter (10.30%) regions throughout the genome, consistent with previously reported profiles of chromatin accessibility in cattle.[20] We also performed the Spearman correlation and principal component analysis on both gene expression and chromatin accessibility profiles, and our data suggested a continuous trajectory of bovine myogenesis and a high correlation between biological replicates (Figures 1E,F and S2A,B). Thus, examples of the chromatin accessibility and transcriptomic dynamic changes during bovine myoblast proliferation and myogenic differentiation at myogenic marker genes were shown to demonstrate the robustness of our dataset (Figures 1B and S2C).
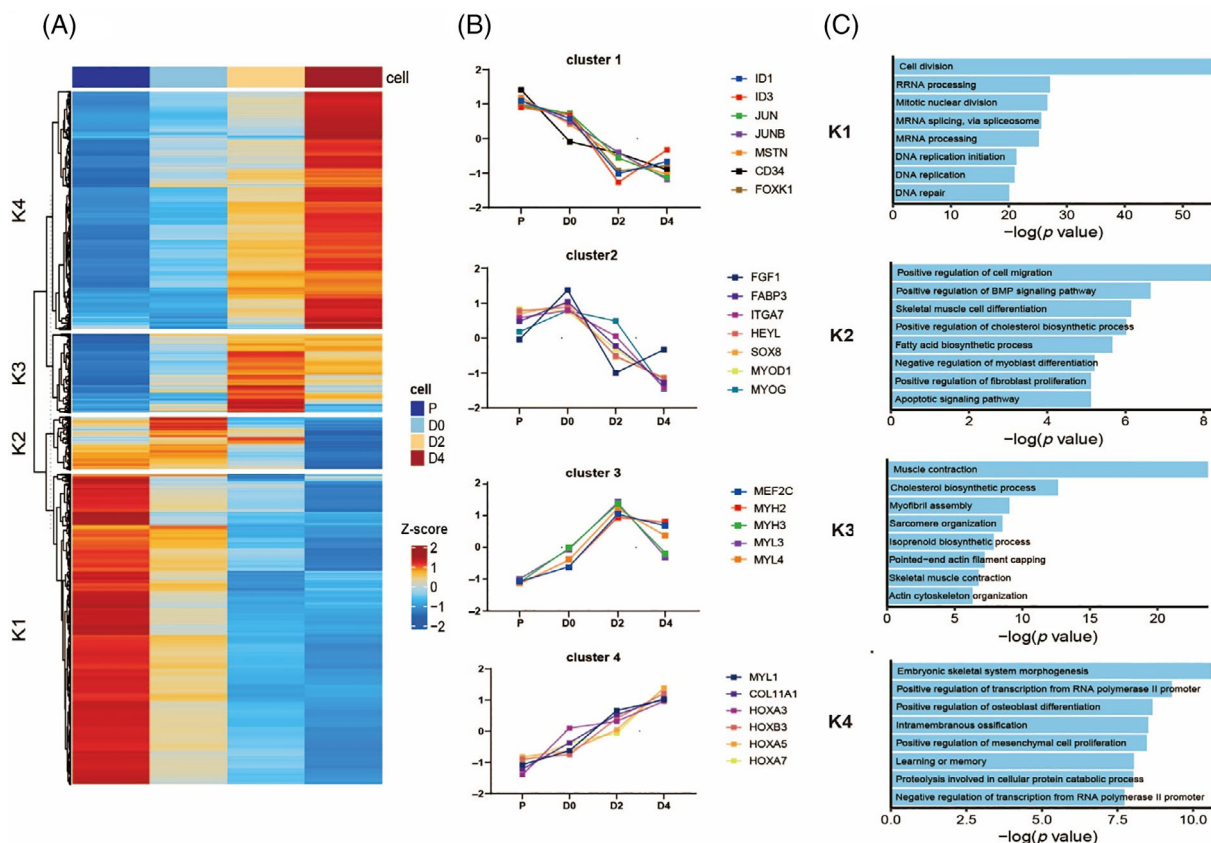

FIGURE 2 Genome-wide transcription profiles of differentially expressed genes (DEGs) from time-point pairwise comparisons. (A) Heatmap of differentially expressed genes from time-point pairwise comparisons, yielding four clusters of co-regulated gene sets. (B) Examples of co-regulated gene sets showed similar gene expression pattern in each cluster. (C) Enrichment gene ontology (GO) terms of co-regulated gene sets in each cluster. Only GO terms for GOTERM_BP_DIRECT and *p*-value <0.05 were considered

3.2 | Time-dependent modules of transcriptional regulation during bovine skeletal muscle development

To characterize transcriptional modules during bovine skeletal muscle development, we defined gene expression profiles at four differentiated time points and identified a total of 4655 differentially expressed genes (DEGs) among each pairwise comparison. We also employed k-means clustering to separate differentially expressed genes, yielding four clusters of co-regulated gene sets that showed a wide range of expression patterns (Figure 2A,B). For example, many up-regulated genes in the early stages of myoblasts proliferation were related to cell proliferation and cell cycle progressions such as *ID1, ID3, JUN*,

JUNB and *MSTN*.[52,53] *HOX* genes (*HOXA3, HOXB3, HOXA5, HOXA7*) involved in embryonic skeletal system morphogenesis were strongly up-regulated in the later stages of myogenic differentiation.[54] Likewise, the marker genes of myogenic myogenesis including *MYOD*, *MYOG* showed maximum expression activity in the early stages, while *MYHs* (*MYH2, MYH3*), *MYLs* (*MYL3, MYL4*) in the later stages. We further performed Gene Ontology (GO) analysis on the differentially expressed genes within each cluster and identified associated GO terms to gain insights into variable changes of gene expression during the bovine skeletal muscle development (Figure 2C). The genes in cluster K1 demonstrated a highly significant enrichment for cell division-related processes, suggesting their potential roles in
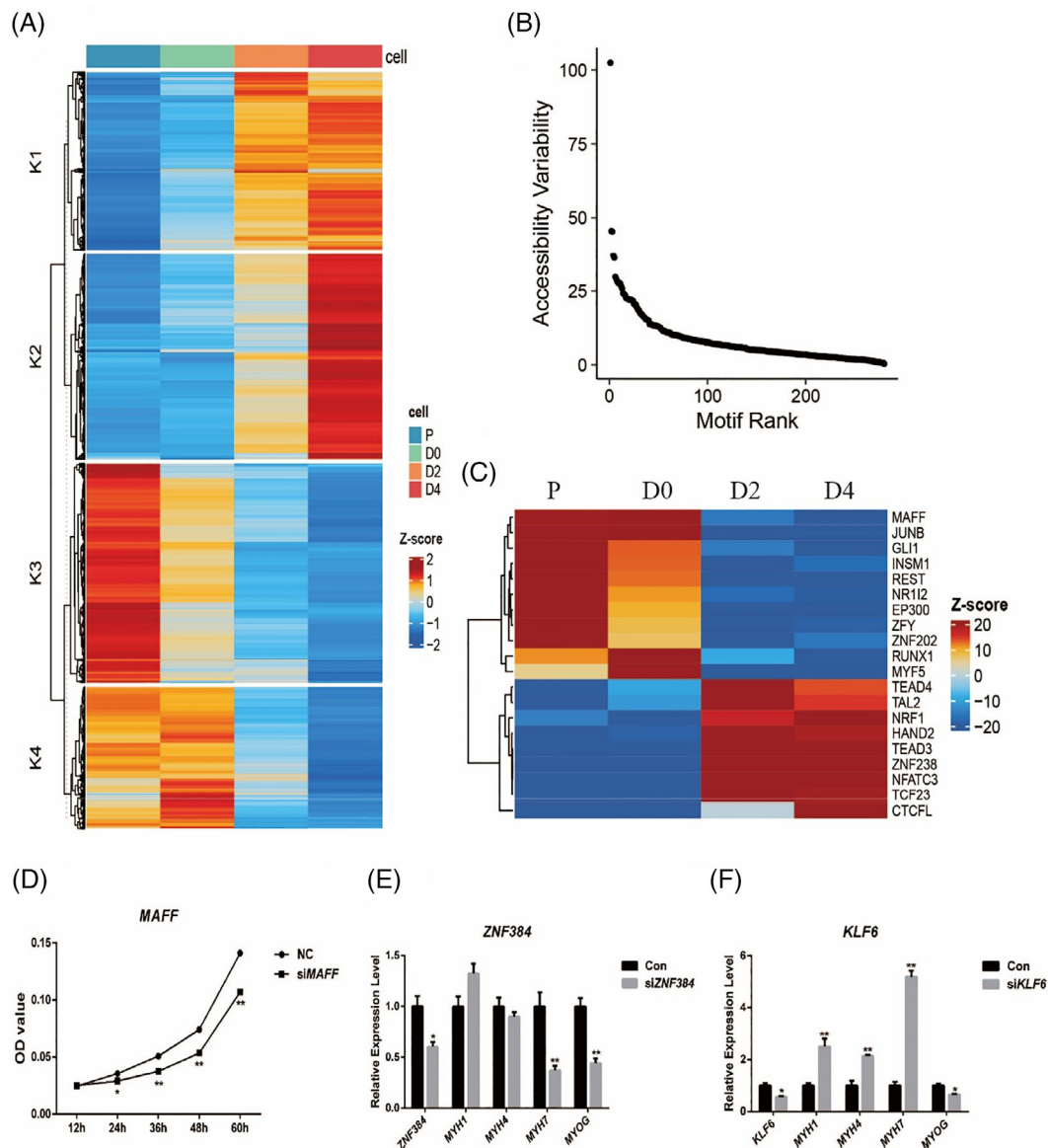

FIGURE 3 Dynamic of chromatin landscape and TF regulatory during myogenesis. (A) Heatmap of differentially accessible chromatin regions, yielding four clusters of co-regulated peak sets. (B) Rank order plot of the transcription factor variability in the OCRs during myogenesis. (C) Heatmap displaying the top 20 most variably motifs. (D) Cell proliferation was measured by CCK8 assay of expressing siMAFF. (E, F) Relative gene expression of muscle-specific genes (*MYH1, MYH4, MYH7* and *MYOG*) following siRNA knockdown of *ZNF384* and *KLF6*.
*p < 0.05, **p < 0.01

myoblasts proliferation. The genes in cluster K4 were significantly enriched embryonic skeletal muscle morphogenesis, consistent with the process of myotubes development.[55,56] Notably, skeletal myogenesis showed high dynamic in cluster K2 and cluster K3, reflecting the peak of myogenesis progress. Taken together clustering of gene expression profiles demonstrated that each cluster comprised unique gene expression pattern may reveal time-dependent modules of transcriptional regulation during skeletal muscle development.

3.3 | Dynamics of the chromatin accessibility and TF regulatory network for bovine skeletal muscle development

We next sought to investigate dynamic changes in chromatin accessibility during bovine myoblasts proliferation and myogenic differentiation. We detected a total of 42,211 differentially accessible chromatin regions (DARs) among each pairwise comparison, which clustered into four groups via k-means clustering method. Dramatic and widespread changes in chromatin accessibility were evident at four differentiated time points, which exhibited a similar profile with gene expression profile (Figure 3A). Remarkably, we found that massive differential accessibility regions occurred at P and D4 time points, and less appeared at D0 and D2 time points, supporting changing and stable chromatin accessibility states during bovine skeletal muscle differentiation. Cluster K3 was comprised of 12,421 open chromatin regions (OCRs) that activated at P time point, and became less accessible at the later stages. In contrast, cluster K2 was composed of 11,652 OCRs that displayed relatively closed chromatin accessible at the early stages, and activated at D4 time point. Cluster K4 consisted of 8033 OCRs gained accessible during the whole period of myoblasts proliferation, while cluster K1 contained 10,105 OCRs remaining accessible during the whole period of myoblast proliferation.

Transcription factors and their regulatory programs mediate the key developmental events in myoblasts proliferation and differentiation. Here, we applied chromVAR to dissect transcription factor variability in the OCRs at four time points (Figures 3B,C and Data S1). Among the most variably accessible sequence motifs determined, we identified important TF motifs such as *MAFF*, *JUNB*, *Runx1*, *MYF5*, *TEAD3* and *TEAD4*. Many TFs from the group were directly related to skeletal muscle cell proliferation and differentiation, and their transcription regulation activity was consistent with their relevant regulatory roles.[57–60] Combined with differential chromatin accessibility, we observed profound loss of chromatin accessibility at promotors region of *NR1I2* (Figure S3A). Although we found *MAFF*, *ZNF384* and *KLF6* exhibited dynamic activity in four time points (Figure S3B), their roles in bovine myoblasts proliferation or differentiation have not been reported. We investigated their effects on myogenesis in cell models using RNA interference. Myoblasts proliferation was inhibited after *MAFF* silencing (Figure 3D). Silencing *ZNF384* impaired *MYOG* and *MYH7* expression, while silencing *KLF6* improved the *MYH1*, *MYH4* and *MYH7* expression but inhibited *MYOG* expression (Figure 3E,F). These findings suggest an integrative role for these factors in the regulatory mechanisms of bovine myoblasts proliferation and myogenic differentiation.
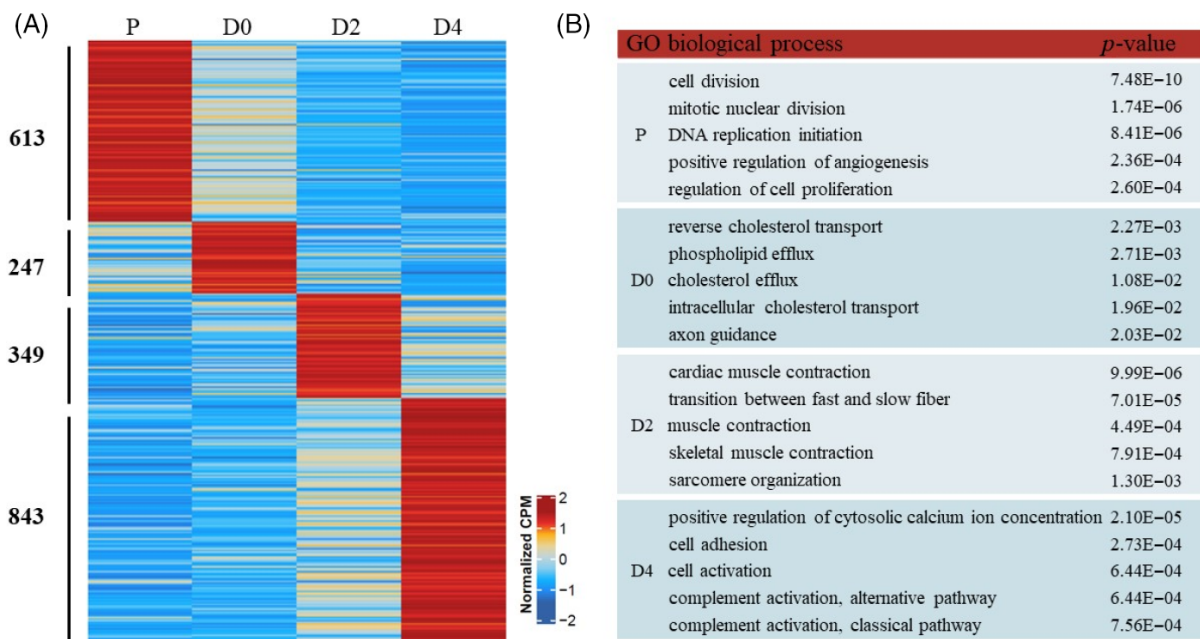

GO biological process		*p*-value
P	cell division	7.48E−10
	mitotic nuclear division	1.74E−06
	DNA replication initiation	8.41E−06
	positive regulation of angiogenesis	2.36E−04
	regulation of cell proliferation	2.60E−04
D0	reverse cholesterol transport	2.27E−03
	phospholipid efflux	2.71E−03
	cholesterol efflux	1.08E−02
	intracellular cholesterol transport	1.96E−02
	axon guidance	2.03E−02
D2	cardiac muscle contraction	9.99E−06
	transition between fast and slow fiber	7.01E−05
	muscle contraction	4.49E−04
	skeletal muscle contraction	7.91E−04
	sarcomere organization	1.30E−03
D4	positive regulation of cytosolic calcium ion concentration	2.10E−05
	cell adhesion	2.73E−04
	cell activation	6.44E−04
	complement activation, alternative pathway	6.44E−04
	complement activation, classical pathway	7.56E−04

FIGURE 4 Characterization of stage-specific genes. (A) Heatmap of stage-specific genes identified at each time point. (B) Table showing the top five GO terms enriched at P, DO, D2 and D4, respectively. Only GO terms for GOTERM_BP_DIRECT and *p*-value < 0.05 were considered

3.4 | Stage-specific gene expression during bovine myoblasts proliferation and myogenic differentiation

To characterize the persistent and stage-specific genetic mechanisms underlying skeletal muscle myogenesis, we next identified gene sets that temporally regulated at different time points. Using Shannon-entropy-based method, we identified 613, 247, 349 and 843 stage-specific genes at P, D0, D2 and D4 time point, respectively (Figure 4A). To further investigate whether specific genes at each time point have a specific gene regulatory pattern, we applied GO enrichment analysis (Figure 4B). We observed that the GO terms of specific genes at each time point were different from those of differentially expressed genes to some extent. Stage-specific genes at P and D2 time points were enriched for GO terms related to cell division and skeletal muscles development, respectively. Interestingly, we found distinct GO terms enriched in D0 and D4 time points. Reverse cholesterol transport ($p < 2.27 \times 10^{-3}$), cholesterol efflux ($p < 1.08 \times 10^{-2}$) and intracellular cholesterol transport ($p < 1.96 \times 10^{-2}$) were significantly enriched in stage-specific genes at D0 time point, while positive regulation of cytosolic calcium ion concentration ($p < 2.10 \times 10^{-5}$), cell adhesion ($p < 2.73 \times 10^{-4}$) and cell activation ($p < 6.44 \times 10^{-4}$) were significantly enriched in stage-specific genes at D4 time point. Altogether, stage-specific genes expression at each time point appears to establish a key restriction point for bovine myoblasts proliferation and myogenic differentiation.

3.5 | Stage-specific open chromatin reveals distinct TF binding events during bovine myoblasts proliferation and myogenic differentiation

As lineage commitment and myogenic differentiation relies on the activity of specific transcriptional programs, we asked whether the stage-specific ATAC-seq peaks harbour specific TFs regulating skeletal muscle development. A total of 3424, 731, 343 and 3022 stage-specific peaks were identified at P, D0, D2 and D4 time points, respectively (Figure 5A). Interestingly, we observed a progressive increase in the number of stage-specific peaks at P and D4 time points and the majority of stage-specific peaks fell into enhancer regions (distal intergenic and intronic regions; Figure S4). The enriched GO terms were involved in a variety of biological functions, while only a small part of GO terms was directly related with skeletal muscle growth and development (Figure S5). The enriched pathways were Rap1, cAMP, Wnt, Hippo and MAPK signalling pathways, which were associated with muscle growth and development (Figure 5B). Next, we investigated the characters of TF binding in stage-specific peaks. Notably, we revealed a series of TFs that showed specific enrichment at different time points (Figure 5C). We observed specific enrichment motif of the *ATF3*, *AP-1* and its family members (*FRA-1*, *FRA-2*, *JUNB*) at P time point, which is in accordance with the observation that *AP-1* complex is correlated with cell proliferation and transformation.[61,62] We also found that high motif enrichment of *MyoG*, *Myf5*, *HLH-1*
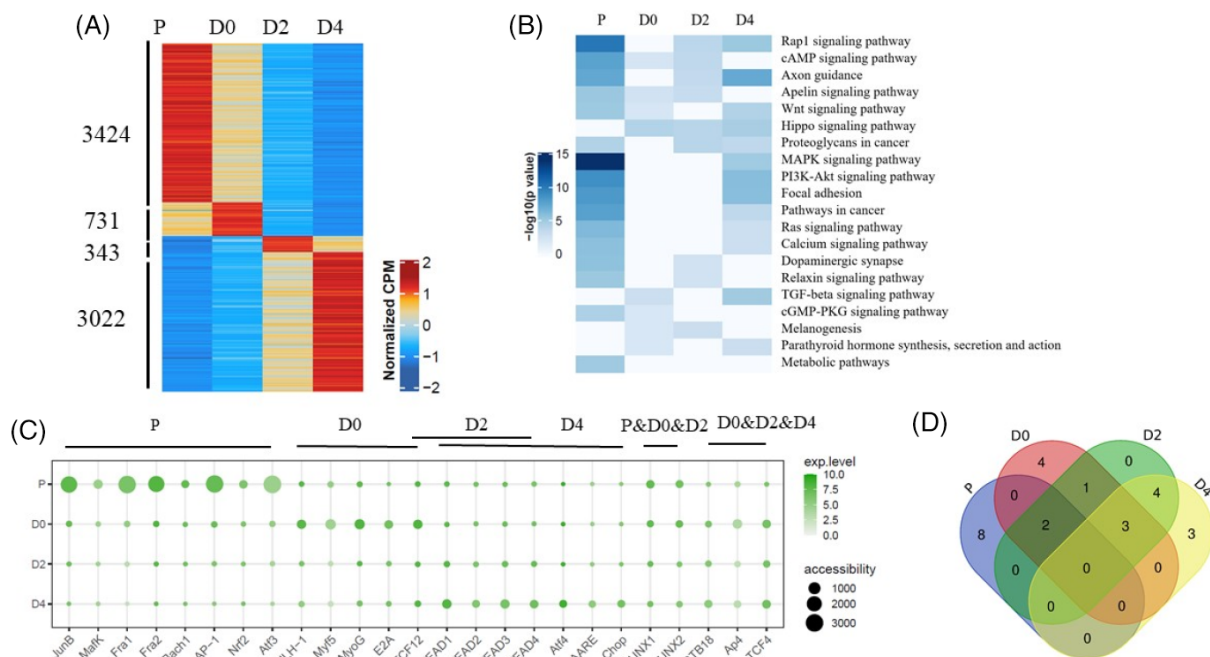

FIGURE 5 Characterization of stage-specific peaks. (A) Heatmap of stage-specific peaks identified at each time point. (B) Heatmap displaying the top 20 significantly enriched KEGG pathways across four time points. *p*-value < 0.05 is statistically significant. (C) The top 10 known TF binding motifs enrichment within stage-specific accessible regions. The dot colour represents gene expression level of TFs and the dot size represents the motif-enrichment value identified by HOMER (−log *p* value). (D) Veen diagram showing the overlap of top 10 known TF binding motifs enrichment at each time point
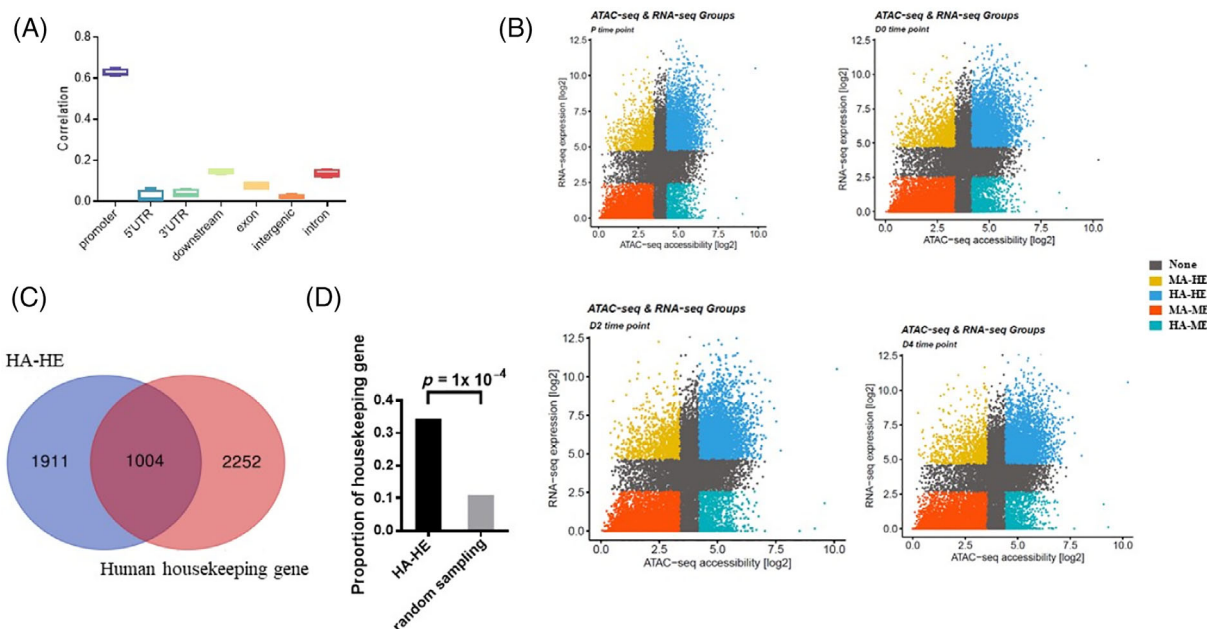
FIGURE 6 Chromatin accessibility at promoter positively correlates with gene expression. (A) Boxplot of the Spearman correlation between normalized counts of RNA-seq (log2CPM) and normalized counts of ATAC-seq (log2CPM) at promoter (TSS ± 2.5 kb), 5'UTR, 3'UTR, downstream, exon, intergenic and intron. (B) Scatter plot showing positive relationships between gene expression level and chromatin accessibility level at promoter for P, D0, D2 and D4 time points. Based on the level of gene expression and chromatin accessibility, genes are divided into four groups: HA–HE: high accessibility/high expression; MA–ME: medium-low accessibility/medium-low expression; HA–ME: high accessibility/medium-low expression; MA–HE: medium-low accessibility/high expression. (C) Veen diagram showing the overlap between HA–HE genes and bovine orthologous housekeeping genes. (D) Enrichment analysis of HA–HE genes in bovine orthologous housekeeping genes compared to random genes. Random selections of 2915 genes from all protein-coding genes
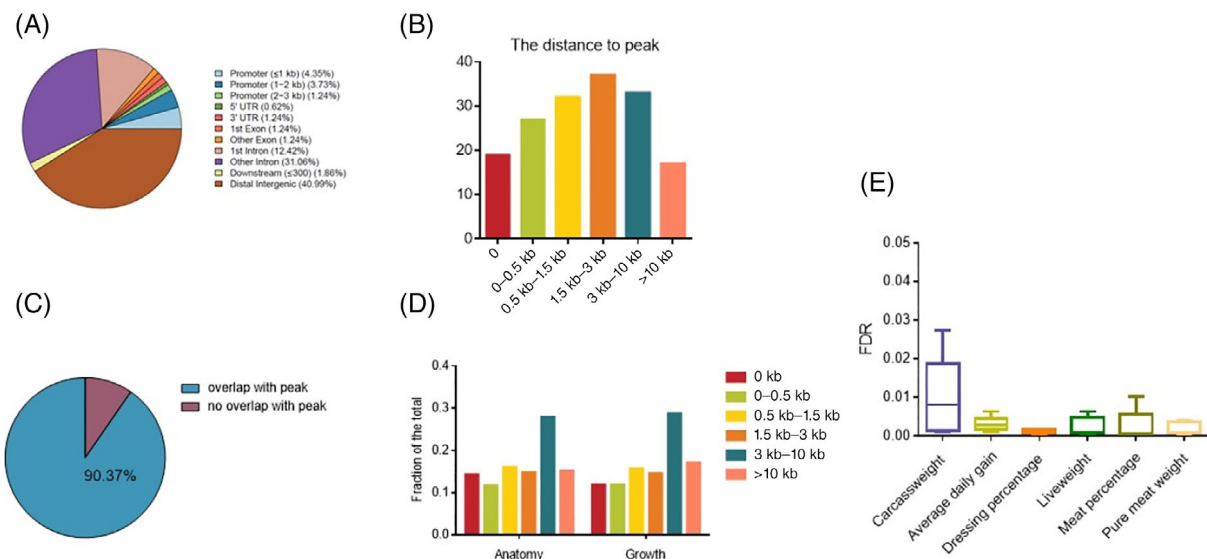

FIGURE 7 Cattle growth-related genetic variants associated with chromatin accessibility regions. (A) Genomic distribution of 163 stature-associated lead SNPs identified by Aniek C. Bouwman et al. (B) The distribution of the distance between 163 stature-associated lead SNPs and their nearest OCRs. Peak distance is grouped into 6 categories: 0 kb (located within the peak), 0–0.5, 0.5–1.5, 1.5–3, 3–10 and >10 kb. (C) The overlap between 163 stature-associated QTL confidence regions and OCRs. (D) Distribution of the distance between known QTL regions and their nearest OCRs. The known QTL regions and associated SNP for anatomy and growth traits are obtained from cattle QTLdb. (E) Boxplot showing the enrichment of GWAS signals for chromatin accessibility obtained from myoblasts proliferation peaks (cluster K3 and K4), myogenic differentiation peaks (cluster K1 and K2), stage-specific peaks, DARs peaks and OCRs peaks. The x-axis represents growth traits, including carcass weight, average daily gain, dressing percentage, liveweight, meat percentage and pure meat weight. The y-axis represents FDR (adjusted p values) obtained from enrichment analysis

appeared at D0 time and these TFs have previously been shown to be markers of muscle lineage specification,[63–65] suggesting that the differentiation program may be initiated at this stage. It is important to note that the TFs, *ZBTB18, TCF4, TEAD1, TEAD2, TEAD3* and *TEAD4*, enriched between D2 and D4 time point were largely consistent (Figure 5D), suggesting a continuous regulatory mechanism that modulates myogenic differentiation.

3.6 | Integrative analysis of chromatin accessibility and gene expression

Enhancer and promoter regions are key players in gene expression regulation through transcription initiation (promoters) and by amplifying such transcription initiation (enhancers).[66–69] Overall, we observed only chromatin accessibility at promoter was positively correlated with its gene expression [Spearman's R (promoter) $= 0.62$], whereas the correlation at other genomic regions was relatively weak (Spearman's $R < 0.14$; Figure 6A). To enhance our understanding of the positive relationship, we further divided genes into groups according to their promoter accessibility and expression levels (Figure 6B). As in previous reports, genes with both high expression level and accessibility at promoter region were enriched for housekeeping-like biological functions.[43,70] Therefore, we verified whether this regulatory pattern was also present during bovine skeletal muscle development. In general, we found a significantly higher proportion (proportion $= 34.42\%$, p-value $<10^{-4}$) of HA-HE genes were bovine orthologous housekeeping genes (Figure 6C,D). In summary, our results indicated a potential relationship between chromatin accessibility and gene expression, which is paramount to understanding the complexities of gene regulatory during bovine skeletal muscle development.

3.7 | Cattle growth-related genetic variants in the context of chromatin dynamics

To dissect genetic aspects of skeletal muscle development, we further investigated whether the open chromatin regions dynamics were associated with muscle-related traits. In this study, we first investigated the association between our data and 163 stature-associated lead SNPs identified by Bouwman et al.[71] (Figure 7A). Although very few of them were located in accessible chromatin peaks, the majority of these lead variants were located within 3 kb of OCRs (Figure 7B). There were 11.5% of lead variants located in OCRs, while 70.55% of lead variants were located within 3 kb. Since there is a reasonable chance that the lead SNP is not the causal variant, we also examined the fractions of OCRs that overlapped with QTL confidence regions and found that a great number (90.37%) of QTL confidence regions were overlapped with at least one OCRs (Figure 7C). A similar pattern was observed in context from cattle QTLdb. About 56.92% and 54.09% of known QTL regions and associated SNP located within 3 kb of OCRs for anatomy and growth traits, respectively (Figure 7D,

Table S5). Thus, we expect that the joint effect of genetic variations in OCRs might impact cattle growth-related traits. We categorized five groups: myoblasts proliferation peaks (cluster K3 and K4), myogenic differentiation peaks (cluster K1 and K2), stage-specific peaks, DARs peaks, OCRs peaks. Significant enrichments (FDR <0.05) were observed for carcass weight, average daily gain, dressing percentage, liveweight, meat percentage and pure meat weight in all five chromatin peak groups. (Figure 7E, Table S6). Notably, we found the joint SNP effects in OCRs were lower than those in other groups (Table S6). Given that the majority of genetic variants fell into non-coding regions, our data therefore may be suitable for the functional interpretation of GWAS results.

4 | DISCUSSION

Previous studies have provided powerful evidence for the crucial roles of cis-regulatory elements (CREs) in identifying causative variants of complex traits.[16,72,73] With driven efforts by FAANG, a growing body of studies engaged in the functional annotation of farm animal genome have emerged, further leading to characterize the genetic architecture of important economic traits and enhance the animal performance.[20,22,74,75] Although major advances have been made in bovine epigenome study, the epigenetic basis for foetal skeletal muscle development remains poorly understood. An important limitation for this study is the lack of ideal samples across multiple developmental time points of bovine skeletal muscle development. In our study, we recapitulated the time course of foetal bovine myoblasts proliferation and myogenic differentiation, which is an ideal model to study bovine early skeletal muscle development *in vitro*. Furthermore, we inferred the gene expression and transcription regulator dynamic change, investigated the major genes that exhibit maximum activity, and identified potential TFs motif that acts actively in accessibility peaks. As a result, the chromatin accessibility and gene expression profiles between time points exhibited widespread changes and period specificity. We comprehensively characterized co-expressed genes clusters and highly variable TFs during bovine myoblasts proliferation and myogenic differentiation, providing a novel insight to capture key transitions and turning points during skeletal muscle development. CREs play essential roles in development, whereas promoter and enhancer serve as the best understood types of CREs.[76–78] Notably, chromatin accessibility across the genome reflects specific binding events at CREs. Despite chromatin accessibility is a powerful marker of active regulatory elements, alone it provides only a partial view of the regulatory mechanisms underlying skeletal muscle development. Further investigations with additional epigenetic data, such as histone modification, are still necessary for accurate identification of CREs in the future.

In order to characterize the sequence of developmental steps leading to bovine myoblasts proliferation and myogenic differentiation, we applied the Shannon-entropy-based method to capture specification gene expression pattern and TF binding events. In terms of specific gene pattern, enrichment for GO terms at the early stages of myoblasts proliferation and myogenic differentiation was expected,

while GO terms enriched at the later stages exhibited distinct biological activities, denoting their specific and temporal roles. The later stages of myoblasts proliferation enriched cholesterol-related GO terms such as reverse cholesterol transport, cholesterol efflux and intracellular cholesterol transport, and the later stages of myogenic differentiation enriched GO terms associated with positive regulation of cytosolic calcium ion concentration, cell adhesion and cell activation. Notably, these biological processes were not observed in differentially expressed genes from each pairwise comparison, elucidating that the stage-specific expression may be responsible for the specific molecular mechanisms. During skeletal muscle development, myoblasts undergo a number of biochemical and morphological changes, including membrane cholesterol decreased during the early stages of myoblasts fusion.[79–83] For instance, previous studies had demonstrated that *ABCG1*- and *ABCA1*-mediated cholesterol efflux played an important role in reducing cell plasma membrane cholesterol content and apoA-IV encoded by *APOA4* gene is known to participate in promoting cholesterol efflux.[84–88] In addition, elevation in cytosolic calcium, membrane remodelling and adhesive interactions are considered to participate in the fusion of myoblasts into multinucleated myotubes.[89–94] Alternatively, these stage-specific genes may contribute to bovine myoblasts proliferation and myogenic differentiation by mediating these biological processes. Overall, these results further supported our prediction on the functions of stage-specific genes, suggesting the reliability of our data analysis. As for nearby genes of stage-specific peaks, we found significant decreases in the number of GO terms related to myogenesis, although muscle growth and development such as Rap1, cAMP, Wnt, Hippo and MAPK signalling pathways were identified. Indeed, the investigation on gene regulatory mechanisms regulated by stage-specific peaks in our study may be limited on the assumption that genes are regulated by proximal regulatory elements. It must be noted that this assumption is not always correct. Additionally, a series of TF binding events occurred as expected during bovine myoblasts proliferation and myogenic differentiation, including *ATF3*, *AP-1* family members, *MyoG*, *Myf5*, *HLH-1*, *ZBTB18*, *TCF4* and *TEAD* family. We propose that these TFs may facilitate our understanding of the sequential order of gene regulatory during skeletal muscle growth and development.

Previous studies have pointed to a strong relationship between gene expression and chromatin accessibility at promoter region.[43,95] However, given the genetic basis of complex cis-regulatory mechanisms, the relationship between chromatin accessibility and gene expression remains functionally uncharacterized at present. In this work, chromatin landscape was classified into six groups based on their genomic regions, and relationships were calculated between chromatin accessibility and gene expression in each group. We only observed a strong positive relationship between chromatin accessibility and their nearby gene expression in promoter group, which were consistent with the previous study.[43]Transcriptional regulation is governed not only by proximal cis-regulatory elements (promoter), but also by distal cis-regulatory elements (enhancer) that often located far away from their target gene.[96] While such a strong positive relationship was absent

in intergenic group, this is partly due to a remarkably large amount of intergenic regions widely distributed across the genome and the inaccurate method of nearby genes annotation. Furthermore, we provided evidence showing that at least one-third of the genes (high expression and high accessibility level at promoter region) were correlated with housekeeping genes. Considering the potential significance of the residual HA/HE genes, we suppose that this regulatory model is an essential step towards completing our understanding of the regulatory mechanism of skeletal muscle development.

Genome-wide association studies (GWAS) have proved to be an effective and promising approach for the identification of genetic variants associated with complex traits.[97] However, most of these genetic variants are located in non-coding regions and a push for annotation of cis-regulatory elements by epigenomic information have contributed to elucidate the functional relevance of non-coding genetic variants to some extent.[98] In this study, integration of chromatin landscape profiles with GWAS signals demonstrated that the GWAS signals of bovine growth traits were significantly enriched in chromatin accessibility of skeletal muscle development, which elucidated the activity of genetic variants in skeletal muscle development-related cis-regulatory elements. Further studies are warranted to prioritize variant SNPs and explain how these variant SNPs change TFs regulatory and genes expression.

5 | CONCLUSION

Overall, our study describes the dynamic changes of chromatin accessibility and gene expression during bovine myoblasts proliferation and differentiation in vitro, which is a good model to reconstruct the process of skeletal muscle development. In addition, we identified a series of stage-specific genes and TFs to enhance our understanding of the sequential regulation of skeletal muscle development. Integrations of chromatin accessibility with transcriptional expression profiles and GWAS signals provide an opportunity to explore the regulatory role of these cis-regulatory elements defined by chromatin accessibility. Taken together, our study demonstrates a step-wise dissection of the transcriptional regulation network for skeletal muscle development and provide a systematic understanding of the molecular circuits governing skeletal muscle development.

CONFLICT OF INTEREST

AUTHOR CONTRIBUTIONS

Qian Li and Yahui Wang performed the experiments, analysed the data and drafted the manuscript. Junya Li and Lupei Zhang conceived and supervised the experiments. Xin Hu, Yapeng Zhang, Hongwei Li, and Qi Zhang participated in data collection and analysis. Wentao Cai, Zezhao Wang, Bo Zhu, Lingyang Xu, Xue Gao, Yan Chen, and Huijiang Gao contributed to data analysis and manuscript preparation.

ETHICS STATEMENT

DATA AVAILABILITY STATEMENT

ORCID

Lupei Zhang https://orcid.org/0000-0001-7701-2331

REFERENCES

1. Janssen I, Heymsfield SB, Wang ZM, Ross R. Skeletal muscle mass and distribution in 468 men and women aged 18-88 yr. *J Appl Physiol*. 2000;1985(89):81-88. doi:10.1152/jappl.2000.89.1.81
2. Zhao L, Huang Y, Du M. Farm animals for studying muscle development and metabolism: dual purposes for animal production and human health. *Anim Front*. 2019;9:21-27. doi:10.1093/af/vfz015
3. Cossu G, Borello U. Wnt signaling and the activation of myogenesis in mammals. *EMBO J*. 1999;18:6867-6872. doi:10.1093/emboj/18.24.6867
4. Zhu MJ, Ford SP, Nathanielsz PW, Du M. Effect of maternal nutrient restriction in sheep on the development of fetal skeletal muscle. *Biol Reprod*. 2004;71:1968-1973. doi:10.1095/biolreprod.104.034561
5. Du M, Tong J, Zhao J, et al. Fetal programming of skeletal muscle development in ruminant animals. *J Anim Sci*. 2010;88:E51-E60. doi:10.2527/jas.2009-2311
6. Bordbar F, Jensen J, du M, et al. Identification and validation of a novel candidate gene regulating net meat weight in Simmental beef cattle based on imputed next-generation sequencing. *Cell Prolif*. 2020;53:e12870. doi:10.1111/cpr.12870
7. Qi X, Hu M, Xiang Y, et al. LncRNAs are regulated by chromatin states and affect the skeletal muscle cell differentiation. *Cell Prolif*. 2020;53:e12879. doi:10.1111/cpr.12879
8. Lu H, Shi X, Wu G, et al. FGF13 regulates proliferation and differentiation of skeletal muscle by down-regulating Spry1. *Cell Prolif*. 2015;48:550-560. doi:10.1111/cpr.12200
9. Kassar-Duchossoy L, Giacone E, Gayraud-Morel B, Jory A, Gomès D, Tajbakhsh S. Pax3/Pax7 mark a novel population of primitive myogenic cells during development. *Genes Dev*. 2005;19:1426-1431. doi:10.1101/gad.345505
10. Stern HM, Brown AM, Hauschka SD. Myogenesis in paraxial mesoderm: preferential induction by dorsal neural tube and by cells expressing Wnt-1. *Development*. 1995;121:3675-3686.
11. Petropoulos H, Skerjanc IS. Beta-catenin is essential and sufficient for skeletal myogenesis in P19 cells. *J Biol Chem*. 2002;277:15393-15399. doi:10.1074/jbc.M112141200
12. Cai S, Zhu Q, Guo C, et al. MLL1 promotes myogenesis by epigenetically regulating Myf5. *Cell Prolif*. 2020;53:e12744. doi:10.1111/cpr.12744
13. Rawls A, Valdez MR, Zhang W, Richardson J, Klein WH, Olson EN. Overlapping functions of the myogenic bHLH genes MRF4 and MyoD revealed in double mutant mice. *Development*. 1998;125:2349-2358.
14. Zhang W, Behringer RR, Olson EN. Inactivation of the myogenic bHLH gene MRF4 results in up-regulation of myogenin and rib anomalies. *Genes Dev*. 1995;9:1388-1399. doi:10.1101/gad.9.11.1388
15. Halstead MM, Ma X, Zhou C, Schultz RM, Ross PJ. Chromatin remodeling in bovine embryos indicates species-specific regulation of genome activation. *Nat Commun*. 2020;11:4654. doi:10.1038/s41467-020-18508-3
16. Fang L, Liu S, Liu M, et al. Functional annotation of the cattle genome through systematic discovery and characterization of chromatin states and butyrate-induced variations. *BMC Biol*. 2019;17:68. doi:10.1186/s12915-019-0687-8
17. Ramachandran K, Senagolage MD, Sommars MA, et al. Dynamic enhancers control skeletal muscle identity and reprogramming. *PLoS Biol*. 2019;17:e3000467. doi:10.1371/journal.pbio.3000467
18. Yue J, Hou X, Liu X, et al. The landscape of chromatin accessibility in skeletal muscle during embryonic development in pigs. *J Anim Sci Biotechnol*. 2021;12:56. doi:10.1186/s40104-021-00577-z
19. Giuffra E, Tuggle CK, FAANG Consortium. Functional annotation of animal genomes (FAANG): current achievements and roadmap. *Annu Rev Anim Biosci*. 2019;7:65-88. doi:10.1146/annurev-animal-020518-114913
20. Halstead MM, Kern C, Saelao P, et al. A comparative analysis of chromatin accessibility in cattle, pig, and mouse tissues. *BMC Genomics*. 2020;21:698. doi:10.1186/s12864-020-07078-9
21. Alexandre PA, Naval-Sánchez M, Menzies M, et al. Chromatin accessibility and regulatory vocabulary across indicine cattle tissues. *Genome Biol*. 2021;22:273. doi:10.1186/s13059-021-02489-7
22. Kern C, Wang Y, Xu X, et al. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nat Commun*. 2021;12:1821. doi:10.1038/s41467-021-22100-8
23. Cao X, Cheng J, Huang Y, Lan X, Lei C, Chen H. Comparative enhancer map of cattle muscle genome annotated by ATAC-Seq. *Front Vet Sci*. 2021;8:782409. doi:10.3389/fvets.2021.782409
24. Guan L, Hu X, Liu L, et al. Bta-miR-23a involves in adipogenesis of progenitor cells derived from fetal bovine skeletal muscle. *Sci Rep*. 2017;7:43716. doi:10.1038/srep43716
25. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29:15-21. doi:10.1093/bioinformatics/bts635
26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15:550. doi:10.1186/s13059-014-0550-8
27. Ignatiadis N, Klaus B, Zaugg JB, Huber W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat Methods*. 2016;13:577-580. doi:10.1038/nmeth.3885
28. Langmead B, Salzberg SL. Fast gapped-read alignment with bowtie 2. *Nat Methods*. 2012;9:357-359. doi:10.1038/nmeth.1923
29. Zhang Y, Liu T, Meyer CA, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137. doi:10.1186/gb-2008-9-9-r137
30. Corces MR, Buenrostro JD, Wu B, et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat Genet*. 2016;48:1193-1203. doi:10.1038/ng.3646
31. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30:923-930. doi:10.1093/bioinformatics/btt656
32. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*. 2015;31:2382-2383. doi:10.1093/bioinformatics/btv145
33. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*. 2016;32:2847-2849. doi:10.1093/bioinformatics/btw313
34. Ramirez F, Ryan DP, Grüning B, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*. 2016;44(W1):W160-W165. doi:10.1093/nar/gkw257

35. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol*. 2011;29:24-26. doi:10.1038/nbt.1754

36. Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods*. 2017;14:975-978. doi:10.1038/nmeth.4401

37. Xie W, Schultz MD, Lister R, et al. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell*. 2013;153:1134-1148. doi:10.1016/j.cell.2013.04.022

38. Wu J, Huang B, Chen H, et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*. 2016;534:652-657. doi:10.1038/nature18606

39. Heinz S, Benner C, Spann N, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576-589. doi:10.1016/j.molcel.2010.05.004

40. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37:1-13. doi:10.1093/nar/gkn923

41. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009;4:44-57. doi:10.1038/nprot.2008.211

42. Bu D, Luo H, Huo P, et al. KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Res*. 2021;49:W317-W325. doi:10.1093/nar/gkab447

43. Starks RR, Biswas A, Jain A, Tuteja G. Combined analysis of dissimilar promoter accessibility and gene expression profiles identifies tissue-specific genes and actively repressed networks. *Epigenetics Chromatin*. 2019;12:16. doi:10.1186/s13072-019-0260-2

44. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet*. 2013;29:569-574. doi:10.1016/j.tig.2013.05.010

45. Smedley D, Haider S, Durinck S, et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res*. 2015;43:W589-W598. doi:10.1093/nar/gkv350

46. Zhu B, Niu H, Zhang W, et al. Genome wide association study and genomic prediction for fatty acid composition in Chinese Simmental beef cattle using high density SNP array. *BMC Genomics*. 2017;18:464. doi:10.1186/s12864-017-3847-7

47. Niu Q, Zhang T, Xu L, et al. Integration of selection signatures and multi-trait GWAS reveals polygenic genetic architecture of carcass traits in beef cattle. *Genomics*. 2021;113:3325-3336. doi:10.1016/j.ygeno.2021.07.025

48. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88:76-82. doi:10.1016/j.ajhg.2010.11.011

49. Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL. Advantages and pitfalls in the application of mixed-model association methods. *Nat Genet*. 2014;46:100-106. doi:10.1038/ng.2876

50. Kent WJ, Sugnet CW, Furey TS, et al. The human genome browser at UCSC. *Genome Res*. 2002;12:996-1006. doi:10.1101/gr.229102

51. Cai W, Li C, Li J, Song J, Zhang S. Integrated small RNA sequencing, transcriptome and GWAS data reveal microRNA regulation in response to Milk protein traits in Chinese Holstein cattle. *Front Genet*. 2021;12:726706. doi:10.3389/fgene.2021.726706

52. Zebedee Z, Hara E. Id proteins in cell cycle control and cellular senescence. *Oncogene*. 2001;20:8317-8325. doi:10.1038/sj.onc.1205092

53. Meixner A, Karreth F, Kenner L, Penninger JM, Wagner EF. Jun and JunD-dependent functions in cell proliferation and stress response. *Cell Death Differ*. 2010;17:1409-1419. doi:10.1038/cdd.2010.22

54. Song JY, Pineault KM, Dones JM, Raines RT, Wellik DM. Hox genes maintain critical roles in the adult skeleton. *Proc Natl Acad Sci USA*. 2020;117:7296-7304. doi:10.1073/pnas.1920860117

55. Huang NF, Patel S, Thakar RG, et al. Myotube assembly on nanofibrous and micropatterned polymers. *Nano Lett*. 2006;6:537-542. doi:10.1021/nl060060o

56. Engler AJ, Griffin MA, Sen S, Bönnemann CG, Sweeney HL, Discher DE. Myotubes differentiate optimally on substrates with tissue-like stiffness: pathological implications for soft or stiff microenvironments. *J Cell Biol*. 2004;166:877-887. doi:10.1083/jcb.200405004

57. Chalaux E, Lopez-Rovira T, Rosa JL, Bartrons R, Ventura F. JunB is involved in the inhibition of myogenic differentiation by bone morphogenetic protein-2. *J Biol Chem*. 1998;273:537-543. doi:10.1074/jbc.273.1.537

58. Kataoka Y, Matsumura I, Ezoe S, et al. Reciprocal inhibition between MyoD and STAT3 in the regulation of growth and differentiation of myoblasts. *J Biol Chem*. 2003;278:44178-44187. doi:10.1074/jbc.M304884200

59. Miner JH, Wold BJ. C-myc inhibition of MyoD and myogenin-initiated myogenic differentiation. *Mol Cell Biol*. 1991;11:2842-2851. doi:10.1128/mcb.11.5.2842-2851.1991

60. Joshi S, Davidson G, le Gras S, et al. TEAD transcription factors are required for normal primary myoblast differentiation in vitro and muscle regeneration in vivo. *PLoS Genet*. 2017;13:e1006600. doi:10.1371/journal.pgen.1006600

61. Andreucci JJ, Grant D, Cox DM, et al. Composition and function of AP-1 transcription complexes during muscle cell differentiation. *J Biol Chem*. 2002;277:16426-16432. doi:10.1074/jbc.M110891200

62. Angel P, Karin M. The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. *Biochim Biophys Acta*. 1991;1072:129-157. doi:10.1016/0304-419x(91)90011-9

63. Pownall ME, Gustafsson MK, Emerson CP Jr. Myogenic regulatory factors and the specification of muscle progenitors in vertebrate embryos. *Annu Rev Cell Dev Biol*. 2002;18:747-783. doi:10.1146/annurev.cellbio.18.012502.105758

64. Hernandez-Hernandez JM, Garcia-Gonzalez EG, Brun CE, Rudnicki MA. The myogenic regulatory factors, determinants of muscle development, cell identity and regeneration. *Semin Cell Dev Biol*. 2017;72:10-18. doi:10.1016/j.semcdb.2017.11.010

65. Zammit PS. Function of the myogenic regulatory factors Myf5, MyoD, Myogenin and MRF4 in skeletal muscle, satellite cells and regenerative myogenesis. *Semin Cell Dev Biol*. 2017;72:19-32. doi:10.1016/j.semcdb.2017.11.011

66. Haberle V, Stark A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat Rev Mol Cell Biol*. 2018;19:621-637. doi:10.1038/s41580-018-0028-8

67. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet*. 2014;15:272-286. doi:10.1038/nrg3682

68. Beagrie RA, Pombo A. Gene activation by metazoan enhancers: diverse mechanisms stimulate distinct steps of transcription. *Bioessays*. 2016;38:881-893. doi:10.1002/bies.201600032

69. Lenhard B, Sandelin A, Carninci P. Metazoan promoters: emerging characteristics and insights into transcriptional regulation. *Nat Rev Genet*. 2012;13:233-245. doi:10.1038/nrg3163

70. Boyle AP, Davis S, Shulha HP, et al. High-resolution mapping and characterization of open chromatin across the genome. *Cell*. 2008;132:311-322. doi:10.1016/j.cell.2007.12.014

71. Bouwman AC, Daetwyler HD, Chamberlain AJ, et al. Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals. *Nat Genet*. 2018;50:362-367. doi:10.1038/s41588-018-0056-5

72. Xiang R, Berg I, MacLeod IM, et al. Quantifying the contribution of sequence variants with regulatory and evolutionary significance to 34 bovine complex traits. *Proc Natl Acad Sci U S A*. 2019;116:19398-19408. doi:10.1073/pnas.1904159116

73. ENCODE Project Consortium. The ENCODE (ENCyclopedia of DNA elements) project. *Science*. 2004;306:636-640. doi:10.1126/science.1105136

74. Pan Z, Yao Y, Yin H, et al. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nat Commun*. 2021;12:5848. doi:10.1038/s41467-021-26153-7

75. Foissac S, Djebali S, Munyard K, et al. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. *BMC Biol*. 2019;17:108. doi:10.1186/s12915-019-0726-5

76. Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat Rev Genet*. 2011;13:59-69. doi:10.1038/nrg3095

77. Levine M. Transcriptional enhancers in animal development and evolution. *Curr Biol*. 2010;20:R754-R763. doi:10.1016/j.cub.2010.06.070

78. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell*. 2011;144:327-339. doi:10.1016/j.cell.2011.01.024

79. van der Bosch J, Schudt C, Pette D. Influence of temperature, cholesterol, dipalmitoyllecithin and Ca2+ on the rate of muscle cell fusion. *Exp Cell Res*. 1973;82:433-438. doi:10.1016/0014-4827(73)90362-5

80. Hirayama E, Sasao N, Yoshimasu S, Kim J. K252a, an indrocarbazole derivative, causes the membrane of myoblasts to enter a fusion-capable state. *Biochem Biophys Res Commun*. 2001;285:1237-1243. doi:10.1006/bbrc.2001.5327

81. Nakanishi M, Hirayama E, Kim J. Characterisation of myogenic cell membrane: II. Dynamic changes in membrane lipids during the differentiation of mouse C2 myoblast cells. *Cell Biol Int*. 2001;25:971-979. doi:10.1006/cbir.2001.0750

82. Sekiya T, Takenawa T, Nozawa Y. Reorganization of membrane cholesterol during membrane fusion in myogenesis in vitro: a study using the filipin-cholesterol complex. *Cell Struct Funct*. 1984;9:143-155. doi:10.1247/csf.9.143

83. Possidonio AC, Miranda M, Gregoracci GB, Thompson FL, Costa ML, Mermelstein C. Cholesterol depletion induces transcriptional changes during skeletal muscle differentiation. *BMC Genomics*. 2014;15:544. doi:10.1186/1471-2164-15-544

84. Hsieh V, Kim MJ, Gelissen IC, et al. Cellular cholesterol regulates ubiquitination and degradation of the cholesterol export proteins ABCA1 and ABCG1. *J Biol Chem*. 2014;289:7524-7536. doi:10.1074/jbc.M113.515890

85. Yvan-Charvet L, Wang N, Tall AR. Role of HDL, ABCA1, and ABCG1 transporters in cholesterol efflux and immune responses. *Arterioscler Thromb Vasc Biol*. 2010;30:139-143. doi:10.1161/ATVBAHA.108.179283

86. Li G, Gu HM, Zhang DW. ATP-binding cassette transporters and cholesterol translocation. *IUBMB Life*. 2013;65(6):505-512. doi:10.1002/iub.1165

87. Stein O, Stein Y, Lefevre M, Roheim PS. The role of apolipoprotein A-IV in reverse cholesterol transport studied with cultured cells and liposomes derived from an ether analog of phosphatidylcholine. *Biochim Biophys Acta*. 1986;878:7-13. doi:10.1016/0005-2760(86)90337-1

88. Vedhachalam C, Duong PT, Nickel M, et al. Mechanism of ATP-binding cassette transporter A1-mediated cellular lipid efflux to apolipoprotein a-I and formation of high density lipoprotein particles. *J Biol Chem*. 2007;282:25123-25130. doi:10.1074/jbc.M704590200

89. Przybylski RJ, MacBride RG, Kirby AC. Calcium regulation of skeletal myogenesis. I. Cell content critical to myotube formation. *In Vitro Cell Dev Biol*. 1989;25:830-838. doi:10.1007/BF02623667

90. Constantin B, Cognard C, Raymond G. Myoblast fusion requires cytosolic calcium elevation but not activation of voltage-dependent calcium channels. *Cell Calcium*. 1996;19:365-374. doi:10.1016/s0143-4160(96)90109-8

91. Przewozniak M, Czaplicka I, Czerwińska AM, et al. Adhesion proteins - an impact on skeletal myoblast differentiation. *PLoS ONE*. 2013;8(5):e61760. doi:10.1371/journal.pone.0061760

92. Cifuentes-Diaz C, Nicolet M, Mege RM. Cell adhesion and development of skeletal muscle. *C R Seances Soc Biol Fil*. 1994;188:505-525.

93. Maartens AP, Brown NH. The many faces of cell adhesion during drosophila muscle development. *Dev Biol*. 2015;401:62-74. doi:10.1016/j.ydbio.2014.12.038

94. Kim JH, Jin P, Duan R, Chen EH. Mechanisms of myoblast fusion during muscle development. *Curr Opin Genet Dev*. 2015;32:162-170. doi:10.1016/j.gde.2015.03.006

95. Ampuja M, Rantapero T, Rodriguez-Martinez A, et al. Integrated RNA-seq and DNase-seq analyses identify phenotype-specific BMP4 signaling in breast cancer. *BMC Genomics*. 2017;18:68. doi:10.1186/s12864-016-3428-1

96. Heintzman ND, Ren B. Finding distal regulatory elements in the human genome. *Curr Opin Genet Dev*. 2009;19:541-549. doi:10.1016/j.gde.2009.09.006

97. Buniello A, MacArthur JAL, Cerezo M, et al. The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res*. 2019;47:D1005-D1012. doi:10.1093/nar/gky1120

98. Tak YG, Farnham PJ. Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome. *Epigenetics Chromatin*. 2015;8:57. doi:10.1186/s13072-015-0050-4

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.