The landscape of chromatin accessibility in skeletal muscle during embryonic development in pigs

Jingwei Yue

Chinese Academy of Agricultural Sciences Institute of Animal Science

Xinhua Hou

Chinese Academy of Agricultural Sciences Institute of Animal Science

Xin Liu

Chinese Academy of Agricultural Sciences Institute of Animal Science

Ligang Wang

Chinese Academy of Agricultural Sciences Institute of Animal Science

Hongmei Gao

Chinese Academy of Agricultural Sciences Institute of Animal Science

Fuping Zhao

Chinese Academy of Agricultural Sciences Institute of Animal Science

Lijun Shi

Chinese Academy of Agricultural Sciences Institute of Animal Science

Liangyu Shi

Chinese Academy of Agricultural Sciences Institute of Animal Science

Hua Yan

Chinese Academy of Agricultural Sciences Institute of Animal Science

Tianyu Deng

Chinese Academy of Agricultural Sciences Institute of Animal Science

Jianfei Gong

Chinese Academy of Agricultural Sciences Institute of Animal Science

Lixian Wang

Chinese Academy of Agricultural Sciences Institute of Animal Science

Longchao ZHANG (✉ zhlchias@163.com)

Chinese Academy of Agricultural Sciences Institute of Animal Science <https://orcid.org/0000-0002-4520-3165>

Research

Keywords: Chromatin accessibility, Embryo, Pig, Skeletal muscle, Transcriptome

Posted Date: January 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-95713/v2>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Version of Record: A version of this preprint was published at Journal of Animal Science and Biotechnology on May 3rd, 2021. See the published version at <https://doi.org/10.1186/s40104-021-00577-z>.

Abstract

Background: The development of skeletal muscle during the embryonic stage in pigs is precisely regulated by transcriptional mechanisms, which depends on chromatin accessibility. However, the landscape of chromatin accessibility in skeletal muscle during embryonic development in pigs has not been reported. To gain insight into the landscape of chromatin accessibility and the associated genome-wide transcriptome during embryonic muscle development, we performed ATAC-seq and RNA-seq on skeletal muscle of pig embryos at 45, 70 and 100 days post coitus (dpc).

Results: In total, 21638, 35447 and 60181 unique regions (or peaks) were found across 45 dpc (LW45), 70 dpc (LW70) and 100 dpc (LW100) embryos, respectively. More than 91% of peaks were annotated within -1 kb to 100 bp of transcription start sites (TSSs). First, widespread increases in specific accessible chromatin regions (ACRs) from 45 to 100 dpc embryos suggested that the regulatory mechanisms became increasingly complicated during embryonic development. Second, the findings of integrated ATAC-seq and RNA-seq analyses showed that not only the numbers but also the peak intensities of ACRs could control the expression of associated genes. Finally, motif screening of stage-specific ACRs revealed some transcription factors that regulated muscle development-related genes, such as MyoD, Mef2c, and Mef2d. Motif screening of DPI of common peaks detected that a potential transcriptional repressor, namely CTCF, was identified among those genes that exhibited different change trends between the ATAC-seq and RNA-seq data.

Conclusions: This work indicates that chromatin accessibility plays an important regulatory role in the embryonic muscle development of pigs and regulates the temporal and spatial expression patterns of key genes in muscle development by influencing the binding of transcription factors. Our results contribute to a better understanding of the regulatory dynamics of genes involved in pig embryonic skeletal muscle development.

Background

The growth of skeletal muscle has received considerable attention since its dysfunction can cause debilitating musculoskeletal disorders [1]. Previous research has shown that the development of skeletal muscle is a complex process including the formation of embryonic muscle fibers, the expansion of postnatal muscle fibers, and the regeneration of adult muscles [2]. The number of muscle fibers is essentially fixed during the embryonic period [3]. Additionally, postnatal fiber hypertrophy depends on the total number of muscle fibers within a muscle [4]. Therefore, the embryonic muscle development process is extremely important. Research on the genetic mechanisms affecting muscle development, especially during embryonic stages, will be beneficial for improvement of the methods of pork production and for expansion of pig breeding strategies. Moreover, pigs are more closely related to humans in terms of their size, anatomy, genome, and physiology than other non-primate species (e.g., traditional rodent models); thus, pigs are more suitable than other species for human health research [5, 6].

Muscle development in pig embryos takes place in two growth waves: the first occurs during days 35–60 of the embryonic stage, forming mainly primary fibers, and the second occurs during days 54–90 of the embryonic period, forming mainly secondary fibers [7, 8]. The primary fibers which were the initially muscle fibers were formed by cell fusion. Then myoblasts attach to the surface of primary fibers and fuse to form secondary fibers. And the morphology between primary fibers and secondary fibers could be recognized easily before prenatal 80 days. Primary fiber, which center had the nucleus or myofibril-free region, had a tubular appearance. The second fibers which generally surround the primary fibers have a solid section. In addition, the volume of primary fibers was 2-3 times that of secondary fibers for most of the prenatal period [3]. This process is regulated by multiple mechanisms at the epigenetic, transcriptional, and posttranscriptional levels [2, 9]. As one of the most common types of regulatory factors, transcription factors (TFs), which can bind with target DNA sequences by their DNA-binding domains to promote or inhibit mRNA transcription, play crucial roles in skeletal muscle development [10–12]. In addition, studies have shown that some TFs called pioneer factors can establish chromatin accessibility by replacing or binding nucleosome DNA or by recruiting chromatin-remodeling agents [13, 14]. Quantitative trait locus (QTL) analysis and 3D genome assembly have shown that changes in chromatin accessibility ultimately alter the long-range effects of TFs [15]. Therefore, it is especially important to study the interaction mechanism between TFs and chromatin accessibility during muscle development.

Chromatin accessibility, an important component of epigenomics, can directly reflect the effects of chromatin structural modification on gene transcription. In eukaryotic lineages, the binding of TFs results in transcriptional activation that is closely related to the disruption of nucleosome assembly at promoters, enhancers, insulators, and locus control regions. Thus, regulatory DNA affects the openness or accessibility of a genomic locus of remodeled chromatin [16]. Several methods have been used to profile chromatin accessibility, such as deoxyribonuclease I (DNase I)-hypersensitive site sequencing (DNase-seq), assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq), and formaldehyde-assisted isolation of regulatory elements sequencing (FAIRE-seq) [16]. Among these methods, ATAC-seq is preferred due to its strong advantages: it requires a small input of cells, has a shorter sample-processing period than the other techniques, and has been applied in a variety of studies [17–19]. Although several studies have recently performed ATAC-seq on numerous tissues, such as the liver, frontal cortex, lung and longissimus dorsi muscle, the landscape of chromatin accessibility in skeletal muscle during embryonic stages remains poorly demonstrated [20–22].

In this study, to investigate the dynamics of chromatin accessibility during muscle development in pig embryos, we used ATAC-seq and RNA-seq to analyze the chromatin accessibility and transcriptome of *longissimus dorsi* tissue at different embryonic stages (45, 70, and 100 days post coitus (dpc)) in Large White (LW) pigs (LW45, LW70, and LW100, respectively). The results of this work will provide a theoretical basis for comparison of the molecular mechanisms of embryonic muscle development among different stages, which will broaden our knowledge of epigenetics during muscle development.

Methods And Materials

Ethics statement

All experiments on pigs were performed under the guidance of the Chinese Academy of Sciences and the Institute of Animal Science, Chinese Academy of Agricultural Sciences (CAAS), China.

Sample description

All Large White purebred pigs in this study were obtained from an experimental pig farm of the Institute of Animal Science, CAAS (Beijing, China). Three full-sib sows were slaughtered at 45, 70 and 100 dpc, which approximately correspond to the primary fiber establishment stage, the secondary fiber development stage and total number of fibers fixed stage, respectively. At each stage, four full-sib embryos (2 males and 2 females) were selected, and fresh *longissimus dorsi* muscle tissue was isolated from between the 5th and 6th ribs of each embryo (Fig. 1A). The tissue was immediately placed in liquid nitrogen for storage.

ATAC-seq and data analysis

A total of 12 samples were used to construct libraries for ATAC-seq. First, approximately 50000 fresh cells were collected from each sample. After several centrifugations, 50 μ L of tagmentation reaction mix (25 μ L of 2 $\times$ reaction buffer from a Nextera kit, 2.5 μ L of Nextera Tn5 transposase from the Nextera kit, and 22.5 μ L of nuclease-free H₂O) was added to each sample, and the reactions were immediately incubated for 30 min at 37°C before undergoing 8 cycles of PCR amplification. A MinElute PCR purification kit (Qiagen) was used to purify the libraries, and an Agilent Bioanalyzer 2000 was used to assess the quality of the libraries. The ATAC-seq DNA was sequenced in 150 bp paired-end sequencing mode on a NovaSeq 6000 platform.

The raw fragments were trimmed with Trimmomatic to eliminate reads containing sequence adapters as well as low-quality base pairs [23]. After trimming, the quality of the fragments was evaluated with FastQC. The reads were aligned to the swine reference genome (*Sus scrofa* 11.1.94) using BWA-MEM with the parameter $-v$ 3 [24]. Duplicated fragments, fragments with a mapping quality of less than 30, fragments mapped to the Y chromosome and mitochondrial DNA were removed with SAMtools [25]. The accessible chromatin regions (ACRs) in each individual were called by MACS 2.0 with the parameter $-f$ BAMPE [26]. To further verify the most representative accessible regions of the genome among samples, BEDTools was used to merge ACRs that were shared by all samples at the same embryonic age [27]. DeepTools was used to convert the BAM files to BigWig files for visualization of the genome-wide peaks in Integrative Genomics Viewer as well as for investigation of the signal distribution in and around the gene bodies [28]. The HOMER annotatePeaks function was used to annotate peaks using *Sus scrofa* 11.1.94 [29]. ATAC-seq peaks that were within -2 kb to +100 bp of transcription start sites (TSSs) were selected as promoter peaks. To evaluate the enrichment of peaks in different genomic regions (observed region/random region), we used bedtools to extract random positions on random chromosomes [27].

Then, those random regions were annotated in genome with a custom script. With process being repeated 5000 times, the average value of annotation result of different genomic regions was regarded as random region. To investigate the relationship between the number of peaks and chromosome length, we normalized the length of chromosome (normalized chromatin length=total peak number*chromatin length/total chromatin length), and then calculate the Pearson correlation coefficient between the normalized chromatin length and the number of peaks. The specific and common peaks among different embryonic ages were identified with BEDTools [27]. DEseq2 was used to perform differential peak intensity (DPI) analysis among different embryonic stages [30]. An adjusted P -value<0.05 and a $\log_2$ |fold change (FC)|>1 were defined as the thresholds of significant DPI. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses were performed with the R package clusterProfiler [31]. The significance cutoff for GO terms and pathways identification were a p -value of 0.01. For investigation of *cis*-regulation elements (CREs) in genomic regions, a 200 bps region from -100 to +100 bp relative from peak center was screened for motifs with the HOMER findMotifsGenome function with the default parameters [29]. The significance cutoff for motif identification was a p -value of 0.01.

RNA-seq library preparation, sequencing and data processing

RNA was extracted from all samples using a TruSeq Stranded Total RNA Ribo-Zero H/M/R Kit (Illumina RS-122-2201) and then subjected to quality assessment using 1.5% agarose gel electrophoresis (to verify integrity) and a NanoDrop instrument (to estimate purity and concentration). High-quality RNA samples were used to prepare the RNA-seq libraries, which were sequenced as paired-end 150 bp sequences on a HiSeq X platform.

For all samples, the raw reads were first trimmed in Trimmomatic to remove adapters and low-quality base pairs [23]. Then, the trimmed reads were aligned to the reference genome (*Sus scrofa* 11.1.94) using STAR [32]. The alignment results were processed with SAMtools to remove unmapped reads [25]. Picard was used to remove optical reads and PCR duplicates (<http://broadinstitute.github.io/picard/>). The read counts within exons were calculated using HTSeq [33]. The methods used to determine the differentially expressed genes (DEGs) and the significance threshold were the same as those described for ATAC-seq. The fragments per kilobase per million mapped reads (FPKM) values were determined to measure the gene expression levels using Cufflinks; this is one of the most common methods at present [34]. All downstream analyses were based on the genes whose FPKM value greater than 1 at least 3 samples in one of embryonic stages.

Results

Genome-wide identification of ACRs during pig embryonic development

In order to examine the genome-wide ACRs involved in muscle development, we conducted ATAC-seq to profile chromatin accessibility at 45, 70 and 100 days during pig embryonic development. ATAC-seq was performed on 12 samples and a total of 152–170 million reads were uniquely mapped to the reference

genome (Additional file 1: Table S1). We first assessed the quality of the libraries based on the peak signal distributions and the lengths of the inserts. Detailed information on the high-quality libraries derived from each individual can be found in Additional file 6: Fig. S1 and Additional file 6: Fig. S2, which show that all libraries exhibited the expected fragment length, with abundant nucleosome-free and span-mononucleosomal fragments (Additional file 6: Fig. S1), and the highest peak signal across the whole gene body in the TSS (Additional file 6: Fig. S2). Numerous peaks with high confidence were obtained based on the 12 libraries (Additional file 1: Table S1). As shown in Table S1, more peaks were acquired in samples from older embryos, indicating that chromatin is globally more accessible in older embryos than in younger embryos. To further verify the most representative accessible regions of the genome among samples, we merged the peaks that were shared among all four samples in the same group. Ultimately, 21638, 35447 and 60181 unique regions (or peaks) with 0.50%, 0.89% and 1.89% coverage for the swine genome were found at 45 dpc, 70 dpc and 100 dpc, respectively. All downstream analyses were based on these regions.

We investigated the relationship between peak numbers and chromosome length and found that chromatin with longer chromosomes displayed more peaks ($r^2=0.84$ (LW45), $r^2=0.90$ (LW70), $r^2=0.89$ (LW100)) (Fig. 1B). To further profile the ACR distribution across the whole genome, we evaluated the relative positions of the peaks in the genome (Fig. 1C). At each stage, numerous peaks were annotated in intron and intergenic regions, accounting for approximately 2/3 of all peaks. Approximately 30%, 21%, and 14% of peaks were identified in the promoter region at 45 dpc, 70 dpc and 100 dpc, respectively. Among those peaks located in the promoter region, more than 91% of the peaks were annotated within -1 kb to 100 bp of the TSS. Few peaks were identified in transcriptional termination sites (TTSSs) and exonic regions. In addition, the percentage of peaks located in the proximal promoter region (-1 kb to 100 bp away from the TSS) gradually decreased as the embryo developed, whereas the percentages of peaks within intron and intergenic regions increased. The ACRs among all stages were relatively enriched in promoter and exon regions, especially in the proximal promoter, instead of intergenic regions (Fig 1D).

We further investigated the average lengths of peaks in different genomic regions (Fig. 2A-2C) and found that peaks within proximal promoter regions were longer than those in other regions. Those in intergenic regions were the shortest.

Coordinate regulation of gene expression by ACRs

RNA-seq was conducted using the same individuals as ATAC-seq to explore the effect of ACRs on gene expression. The basic information of each RNA library is listed in Additional file 2: Table S2. Averages of 32.69 million reads per sample were uniquely mapped to the swine reference genome finally (*Sus scrofa* 11.1.94). The results of cluster analysis showed that the four replicates of each stage were all grouped together, indicating that our experiment had good repeatability (Additional file 6: Fig. S3). To explore the role chromatin accessibility in regulating gene expression, we performed further analysis on those peaks located at the proximal promoter. More than 90% of genes annotated by those peaks contained only one ACR in the proximal promoter region (Fig. 3A). We then analyzed the relationship between ACR number

and gene expression level. The proportion of genes with high expression (FPKM>30) in the multiple-ACR group was greater than that in the one-ACR group (Fig. 3B). The genes that contained a single ACR in the proximal promoter region were sorted according to peak length and were then divided into 3 equal groups, namely, the top, middle and bottom groups, to investigate the connection between peak length and gene expression level. When several peak had the same length, we assigned the peak to the groups included the greatest numbers of peaks of similar length (i.g. Assuming that there are a total of 90 peaks. We divide them into 3 groups according their length. Thus, there should be 30 peaks in each group. However, when the length of the 29th and 30th peaks of the first group is equal to the first peak of the second group, we will assign the first peak in the second group to the first group). The results showed that the proportion of highly expressed genes with FPKM values greater than 30 gradually decreased from the top group to the bottom group in all of stages except in LW70 stage of one ACR group and LW45 stage of multiple group (based on total length), regardless of whether the genes have an ACR or multiple ACRs. And low-expressed genes with FPKM values lower than 5 of whole stages except in LW70 stage of multiple ACR group (based on max length) increased as the peak length declined (Fig. 4A-4C). These results demonstrate that genes with longer ACRs show comparatively higher expression levels.

Identification of stage-specific and differential ACRs during embryonic development

Although a large number of ACRs were detected at all stages examined in this research, differences were observed among stages (Fig. 5A) when we compared the genome-wide ACRs obtained from the different samples. In total, 1390, 2808 and 28153 stage-specific peaks were identified at LW45, LW70 and LW100, respectively (Fig. 5B). For example, *XRCC1*, which can manage a temporally responsive DNA repair process to advance the muscle differentiation program, contained an ACR in its proximal promoter region specifically at LW45 (Fig. 5C) [35], and *ENPP2*, which can regulate WNT/ β -Catenin signaling to control myogenic differentiation, displayed an LW70-specific ACR in the proximal promoter region (Fig. 5C) [36]. In contrast, *BCL6*, which is related to the process of terminal differentiation in muscle cells, showed an LW100-specific ACR (Fig. 5C) [37]. These results demonstrate that ACRs vary dynamically. The changes in ACRs in each stage were consistent with the expression levels of the genes shown in Fig. 5E. The highest expression levels of *XRCC1* and *ENPP2* were observed at the LW45 and LW70 stages, respectively, whereas the highest expression level of *BCL6* was observed at the LW100 stage, indicating that ACRs may regulate gene expression by altering the binding sites for TFs.

We also identified 17574 common peaks that were shared by all three embryonic stages in this study (Fig. 5B). However, the intensities of some of these peaks differed among the different groups. For example, in the gene body and promoter regions of *MYOD1*, an important regulator affecting muscle development, chromatin became more accessible from LW45 to LW100, especially in the LW100 stage whose peak intensity was significantly higher than other stages ($p<0.05$) (Fig. 5D) [38]. Correspondingly, the expression level of *MYOD1* gradually increased as the embryonic stage progressed (Fig. 5E). This result suggests that the changes in the intensity of ACRs associated with this gene may also affect its expression ($r^2=0.37$). Then, we performed DPI analysis for the common peaks among the different embryonic stages (Table 1). In total, 145, 1726 and 2687 ACRs were detected to have DPIs in the LW70 vs

LW45, LW100 vs LW70 and LW100 vs LW45 comparisons, respectively. As shown in Fig 5B and Table 1, more peaks with significantly different intensities were observed in the LW100 vs LW70 comparison or in stage-specific ACRs at the LW100 stage than in other comparisons or stages, respectively, suggesting that a widespread change in chromatin accessibility occurs primarily during the formation of secondary fibers.

Distinct regulation of genes at different stages during skeletal muscle development

To further elucidate the regulatory roles of genes with stage-specific ACRs as well as with DPIs during embryonic muscle development, we investigated the peak-related genes and their functions. A total of 11344 genes were found to have stage-specific ACRs among the three stages (Fig. 6A). To assess the biological functions associated with the stage-specific ACR-related genes, GO analyses were performed (Additional file 3: Table S3). The GO analysis results showed that genes with stage-specific ACRs were significantly enriched for many different muscle-related biological processes (BPs). For example, genes associated with the LW45 were significantly enriched for the actin cytoskeleton organization and actin filament organization BPs [39, 40]. Genes with stage-specific ACRs at the LW70 were enriched for BMP signaling pathway and positive regulation of Wnt signaling pathway [41, 42]. And genes in LW100 stages were enriched for the muscle hypertrophy, striated muscle hypertrophy and muscle tissue development BPs. The results of KEGG analysis were displayed in Additional file 3: Table S3. Only one pathway, namely axon guidance, was found in LW70 stage and no significant pathways were enriched in LW45 stage. Three pathways were found in LW100 stage. The most significant enrichment pathway was Leukocyte trans-endothelial migration. Notably, 223 genes with stage-specific ACRs overlapped among the three stages, suggesting that these genes may be differentially regulated during different stages of embryonic muscle development (Fig. 6B) [43]. These overlapping genes were significantly enriched for muscle-related terms, such as skeletal muscle tissue development, actin filament organization and actin cytoskeleton organization (Additional file 3: Table S3).

For common peaks, we annotated these DPI found before, and then performed functional analysis on the annotated genes. A total of 2378 genes were observed among all comparisons (Fig. 6C). Similar to the results obtained for genes with stage-specific ACRs, genes annotated by DPIs in each comparison except in LW70 vs LW45 group were significantly enriched for many muscle-related related BPs (Additional file 4: Table S4), such as muscle organ development, muscle system process, skeletal muscle tissue development and skeletal muscle tissue regeneration. A total of 14 terms were unmasked in LW70 vs LW45 group, most of which were significantly associated with cardiac-related development (Additional file 4: Table S4). Several metabolic related pathways, such as fatty acid metabolism were significantly enriched for the genes with DPIs in the LW70 vs LW45 comparison [44]. Moreover, some custom developmental pathways, such as Wnt signaling pathway, MAPK signaling pathway and AMPK signaling pathway have been proven to play essential roles in muscle development and were enriched significantly in LW100 vs LW70 and LW100 vs LW45 comparisons [42, 45, 46]. In addition, among those genes with DPIs, we identified 63 genes with continuous significant changes in their peak intensities (Fig. 6D). A total

of 10 terms were enriched significantly and related to cell fundamental functions (Additional file 4: Table S4).

Identification of regulatory DNA elements at different stages during embryonic development

The expression levels of genes were dramatically different at different stages during embryonic muscle development. We identified 2805 DEGs through pairwise differential expression analysis among the different stages. All of these DEGs could be divided into 6 clusters based on their expression patterns (Fig. 7A). The expression level of genes in cluster 1 decreased gradually from LW45 to LW100, while opposite trends of genes expression was detected in cluster 5. The genes in cluster 2 were highly expressed at the LW70 stage but displayed sharp changes at the LW45 and LW100 stages. Genes involved in cluster 3 were low expressed at LW45 and LW70 stage, while increased sharply in the LW100 stage. The expression level of cluster 4 was high in the LW45 and LW70 stages, while the expression level was the lowest in the LW100 stage. Genes in cluster 6 had a peak expression at LW45 and bottom expression at LW70 (Fig 7B).

For genes with stage-specific peaks in their proximal promoter regions, we integrated those genes with the DEGs revealed by RNA-seq and found that genes containing stage-specific ACRs in LW45 were most enriched in cluster 1, which had the highest expression level in the LW45 stage (Table 2). Genes with LW100 stage-specific ACRs were significantly enriched in cluster 3 and cluster 5, which displayed the highest expression level at LW100 (Table 2). Interestingly, genes associated with LW70 stage-specific ACRs were significantly enriched in cluster 4, which exhibited the peak expression level at LW45 (Table 2). These findings suggested that the alterations in gene expression may relate to the regulation of ACRs.

We further investigated the CREs within ACRs at the proximal promoter in cluster 1 and cluster 3-5 with HOMER software. Several known motifs associated with muscle development were significantly enriched in all stages (Additional file 6: Fig. S4-S7). For example, MEF2c and MyoG, which are well-known fundamental regulators of the skeletal muscle lineage during the embryonic period, were significantly enriched at the LW45 specific ACR in cluster 1 (Additional file 6: Fig. S4) [38, 47]. In addition, muscle development-related TFs, such as Mef2d and Mef2c, members of the MEF2 family of TFs, were significantly identified at the LW100 specific ACR in cluster 3 and 4 (Additional file 6: Fig. S6-S7) [47]. These results suggest that the identified TFs regulate a considerable proportion of genes and may play important roles during embryonic muscle development. The Nf1 being important roles in skeletal muscle development was detected at the LW70 stage (Additional file 6: Fig. S5) [48, 49]. These results indicate that the regulation of genes by TFs is dynamic during embryonic muscle development.

For genes with common peaks, we integrated the DPIs-related gens in ATAC-seq with the DEGs detected in RNA-seq (Table 3). A total of 637 differential peaks corresponding to 506 DEGs were revealed. Among those peaks mapped to DEGs, that approximately 80% showed the same expression trend (i.e genes containing DPIs in ATAC-seq were upregulated and the same genes in DEGs found by RNA-seq were also upregulated in the same comparison and vice versa.), suggesting ACRs might play important roles in regulating the expression of genes.

To further explore the functions of these genes in detail, GO and KEGG pathway analyses were performed (Additional file 5: Table S5). The comparison of LW100 vs LW70 and LW100 vs LW45 showed that several muscle-related terms were enriched, such as striated muscle contraction and muscle contraction. While the significantly enriched GO terms in LW70 vs LW45 comparison were mainly associated with synaptic-related functions, such as regulation of presynapse assembly and regulation of presynapse organization BPs. No pathways were significantly enriched in LW70 vs LW45 comparison. And two pathways, AMPK signaling pathway and calcium signaling pathway were simultaneously exposed in LW100 vs LW70 and LW100 vs LW45. And calcium signaling pathway was meanwhile the most enriched pathway in LW100 vs LW70 and LW100 vs LW45.

Several peaks mapped to DEGs exhibited different change trends (i.e genes containing DPIs in ATAC-seq were upregulated and the same genes in DEGs found by RNA-seq were downregulated in the same comparison and vice versa.) between the ATAC-seq and RNA-seq data shown in Table 3. We sought to identify whether these differentially altered ACRs were enriched for some particular transcriptional repression factor (Additional file 6: Fig. S8–S10). Ultimately, a known transcriptional repression factors CTCF, was discovered in both the LW100 vs LW70 and LW100 vs LW45 comparisons, suggesting that some of the genes in these two comparisons may be regulated by the same TFs during embryonic muscle development (Additional file 6: Fig. S9-S10).

Discussion

The development of skeletal muscle during the embryonic period determines muscle growth [50, 51]. Pigs undergo primary and secondary fiber formation during embryonic muscle development via a series of complex regulatory mechanisms [52]. Genome-wide chromatin accessibility affects BPs at different developmental stages, including embryonic muscle development, by regulating TF activity. In the present study, we analyzed the chromatin accessibility and transcriptomes of *longissimus dorsi* from Large White pig embryos (at 45 dpc, 70 dpc, 100 dpc). To our knowledge, this is the first systematic study on the chromatin accessibility of pig embryonic skeletal muscle by ATAC-seq in conjunction with transcriptomic analysis.

Although ATAC-seq has become the preferred approach quickly for the study of chromatin accessibility, there are still several limitations. The method is performed based on the activity of Tn5 transposase which exhibit tiny preference for specific DNA sequence [53, 54]. And, the limitation or inappropriateness of generalized methods may challenge data analysis from ATAC-seq [16]. Moreover, it is increasingly appreciated that the interactions between proteins and DNA is highly dynamic, which may not be capable to disclose by the current methods for profiling chromatin accessibility. The particular regions we obtained are the stable relatively, open parts of chromatin [55]. In addition, accurate peak call was essential for downstream analysis in ATAC-seq. However, the analysis of repeat regions, usually called blacklist, may lead to false positive peaks. Thus, reads that fall into blacklist regions were removed from further analyses [56]. At present, blacklist had been built for some species, such as humans, mice, and fly.

And some software had been developed to detect those regions, but it has not been applied cross-species or validated on ChIP-Seq results [56-58].

The crucial ATAC-seq technique involves library construction using the hyperactive transposase Tn5. Therefore, we analyzed the insert size distribution and peak signal enrichment and observed a clear pattern in the fragment distribution: the nucleosome-depleted and span-mononucleosomal regions accounted for half of the total reads. In addition, the ACRs were concentrated at the proximal TSSs, which is consistent with the fact that chromatin around TSSs throughout the genome is more accessible than that in surrounding genomic regions. Overall, our results are in excellent agreement with those of many studies [59, 60]. Through ATAC-seq genome-wide identification of peaks in muscle tissue from embryos of different stages, we found that the number of peaks increased with the age of the embryo, indicating that chromatin accessibility is involved in the regulation of embryonic muscle development. During the period of secondary fiber proliferation and differentiation (70–100 dpc), chromatin accessibility in LW pigs increased significantly, indicating that compared with those of primary fibers, the regulatory mechanisms of proliferation and differentiation of secondary fibers may be more complicated.

The landscapes of chromatin accessibility have been profiled in many species, such as humans, mice and plants [61-63], and the resulting data have provided references for deep genome research. Based on an early theory, the hierarchical compaction and organization of nucleosomes in eukaryotes can divide the genome into inactive regions and active regions, such as promoters and enhancers, that participate in subsequent transcription processes [64]. Based on this theory, the locations of ACRs are enriched in promoters in many studies [65, 66]. Consistent with our results, the majority of peaks have been mapped to intergenic regions and introns, followed by promoters and exons in approximately equal proportions, and the distributions of peaks in the genome have been similar among different species despite the different genome sizes and degrees of genome annotation [21, 67, 68]. We speculate that this similarity may reflect a little of conservation in the distribution of open chromatin among various species or tissues.

In our study, the ACR number and ACR length, which were associated with gene expression levels, varied among different stages. We found that ACRs in proximal promoter regions were significantly longer than those in other regions, whereas those in intergenic regions were the shortest. Moreover, the genes with longer ACRs or multiple ACRs in their proximal promoter regions tended to exhibit higher expression levels. These observations suggest that the chromatin accessibility of promoters can alter the expression levels of associated genes. Similar results have been found in many previous studies that can verify our results to some extent [69, 70]. These findings may indicate a conserved regulatory pattern for gene expression mediated by chromatin accessibility among different species.

ACRs vary dramatically among different tissues and cell types during the developmental stage [66, 71, 72]. For example, widespread decreases in chromatin accessibility have been demonstrated to occur in age-related macular degeneration, and a few common ACRs have been found to be shared among different tissues [73]. In the current study, both specific and shared ACRs were uncovered during

embryonic muscle development. Genes associated with stage-specific or common ACRs were enriched in a series of classic muscle signaling pathways and various muscle-related GO terms. Notably, several genes harbored dynamic ACRs or exhibited DPIs in different stages. These results show that ATAC-seq can be used to effectively analyze the effects of epigenetics on muscle development.

Moreover, through joint analysis of chromatin accessibility and transcriptome data, we further investigated the interactions between TFs and ACRs and analyzed the reasons for the specific spatiotemporal expression of genes during embryonic muscle development in pigs. In our study, some genes annotated with DPIs exhibited different change patterns between ATAC-seq and RNA-seq data. Hence, we speculated that some potential transcriptional repressors might play important roles in embryonic muscle development. Through motif analysis, a key TF CTCF, was found in both the LW100 vs LW70 and LW100 vs LW45 comparisons. CTCF, which is a highly conserved zinc-finger DNA-binding protein, has been found to play a role as a transcriptional repressor of the *Myc* gene and to be involved in the occurrence of various cardiovascular diseases [74, 75]. CTCF has a very large number of binding sites in the mammalian genome (40000–80000), which are mainly concentrated in intergenic regions and introns and overlap with enhancer and promoter sequences. According to previous studies, the binding of CTCF to promoter or enhancer regions often exerts an inhibitory effect [76, 77]. Therefore, we speculate that the inconsistent trend may have been caused by the binding of CTCF to the promoter or enhancer regions of some genes.

Conclusions

Taken together, the findings of this study demonstrate the dynamic changes in chromatin accessibility that occur during embryonic muscle development. Integrated analysis of ATAC-seq and RNA-seq data showed that modulation of ACRs can significantly alter the associated genes and identified numerous valuable CREs and potential transcriptional repressors for the regulation of muscle development. This study can therefore be used as a reference for future research on muscle development in mammals.

Declarations

Consent for publication:

Not applicable.

Availability of data and materials:

The sequencing datasets supporting the conclusions of this article are available in the BIG Data Center (<http://bigd.big.ac.cn/>) with the accession code CRA003275

Competing interests:

The authors declare that they have no competing interests.

Funding:

This research was supported by the Agricultural Science and Technology Innovation Program (ASTIP-IAS02).

Abbreviations

dpc:days post-coitus; TF: transcription factors; QTL: quantitative trait locus; DNase-seq: Deoxyribonuclease I (DNase I)-Hypersensitive Site sequencing; ATAC-seq: transposase-accessible chromatin with high-through sequencing; FAIRE-seq: Formaldehyde-Assisted Isolation of Regulatory Elements sequencing; D/LW: Large White; ACR: accessible chromatin regions; TSS: transcription start sites; TTS: transcriptional termination sites; BP: biological processes; DPI:differential peak intensity; DEGs: differential expressed genes;

Acknowledgements

We thank all the researchers at our laboratories for their help with samples collection.

Authors' contributions

YJW ZLC designed the experiment and performed the analyses. YJW drafted the manuscript. YJW, WLX and ZLC improved the experiments. All authors read and approved the final manuscript.

Ethics approval and consent to participate

All the experiments on pigs were performed under the guidance of the Chinese Academy of Sciences and Institute of Animal Science, CAAS, China.

References

1. Chal J, Pourquie O. Making muscle: skeletal myogenesis in vivo and in vitro. *Development*. 2017;144(12):2104-22.
2. Bentzinger CF, Wang YX, Rudnicki MA. Building muscle: molecular regulation of myogenesis. *Cold Spring Harb Perspect Biol*. 2012;4(2):a008342.
3. Wigmore PM, Stickland NC. Muscle development in large and small pig fetuses. *J Anat*. 1983;137 (Pt 2)(Pt 2):235-45.
4. Rehfeldt C, Henning M, Fiedler I. Consequences of pig domestication for skeletal muscle growth and cellularity. *Livestock Science*. 2008;116(1-3):30-41.
5. Meurens F, Summerfield A, Nauwynck H, Saif L, Gerds V. The pig: a model for human infectious diseases. *Trends Microbiol*. 2012;20(1):50-7.
6. Wernersson R, Schierup MH, Jorgensen FG, Gorodkin J, Panitz F, Staerfeldt HH, et al. Pigs in sequence space: a 0.66X coverage pig genome survey based on shotgun sequencing. *BMC*

- Genomics. 2005;6:70.
7. Lefaucheur L, Ecolan P. Pattern of muscle fiber formation in Large White and Meishan pigs. *Archiv fur Tierzucht*. 2005;48:117-22.
 8. Picard B, Berri C, Lefaucheur L, Molette C, Sayd T, Terlouw C. Skeletal muscle proteomics in livestock production. *Brief Funct Genomics*. 2010;9(3):259-78.
 9. Baar K. Epigenetic control of skeletal muscle fibre type. *Acta Physiol (Oxf)*. 2010;199(4):477-87.
 10. Meadows SM, Warkman AS, Salanga MC, Small EM, Krieg PA. The myocardin-related transcription factor, MASTR, cooperates with MyoD to activate skeletal muscle gene expression. *Proc Natl Acad Sci U S A*. 2008;105(5):1545-50.
 11. Wu H, Olson EN. Activation of the MEF2 transcription factor in skeletal muscles from myotonic mice. *J Clin Invest*. 2002;109(10):1327-33.
 12. Smith NC, Matthews JM. Mechanisms of DNA-binding specificity and functional gene regulation by transcription factors. *Curr Opin Struct Biol*. 2016;38:68-74.
 13. Zhu F, Farnung L, Kaasinen E, Sahu B, Yin Y, Wei B, et al. The interaction landscape between transcription factors and the nucleosome. *Nature*. 2018;562(7725):76-81.
 14. Spitz FO, Furlong EEM. Transcription factors: from enhancer binding to developmental control. *Nat Rev Genet*. 2012;13(9):613-26.
 15. Tehranchi A, Hie B, Dacre M, Kaplow I, Pettie K, Combs P, et al. Fine-mapping cis-regulatory variants in diverse human populations. *Elife*. 2019;8:e39595.
 16. Tsompana M, Buck MJ. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin*. 2014;7(1):33.
 17. Buenrostro JD, Wu B, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*. 2015;523(7561):486-90.
 18. Mezger A, Klemm S, Mann I, Brower K, Mir A, Bostick M, et al. High-throughput chromatin accessibility profiling at single-cell resolution. *Nat commun*. 2018;9(1):3647.
 19. Chen X, Miragaia RJ, Natarajan KN, Teichmann SA. A rapid and robust method for single cell chromatin accessibility profiling. *Nat Commun*. 2018;9(1):5345.
 20. Tissières V, Geier F, Kessler B, Wolf E, Zeller R, Lopez-Rios J. Gene Regulatory and Expression Differences between Mouse and Pig Limb Buds Provide Insights into the Evolutionary Emergence of Artiodactyl Traits. *Elsevier Sponsored Documents*. 31(1):107490.
 21. Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, et al. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. *BMC Biol*. 2019;17(1):108.
 22. Halstead MM, Kern C, Saelao P, Wang Y, Chanthavixay G, Medrano JF, et al. A comparative analysis of chromatin accessibility in cattle, pig, and mouse tissues. *BMC Genomics*. 2020;21(1):698.
 23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-20.

24. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60.
25. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078-9.
26. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9(9):R137.
27. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841-2.
28. Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160-5.
29. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576-89.
30. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.
31. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284-7.
32. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2013;29(1):15-21.
33. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166-9.
34. Roberts A, Trapnell C, Donaghey J, Rinn JL, Pachter L. Improving RNA-Seq expression estimates by correcting for fragment bias. *Genome Biol*. 2011;12(3):R22.
35. Al-Khalaf MH, Blake LE, Larsen BD, Bell RA, Brunette S, Parks RJ, et al. Temporal activation of XRCC1-mediated DNA repair is essential for muscle differentiation. *Cell Discov*. 2016;2:15041.
36. Sah JP, Hao NTT, Han X, Tran TTT, McCarthy S, Oh Y, et al. Ectonucleotide pyrophosphatase 2 (ENPP2) plays a crucial role in myogenic differentiation through the regulation by WNT/beta-Catenin signaling. *Int J Biochem Cell Biol*. 2020;118:105661.
37. Kumagai T, Miki T, Kikuchi M, Fukuda T, Miyasaka N, Kamiyama R, et al. The proto-oncogene Bcl6 inhibits apoptotic cell death in differentiation-induced mouse myogenic cells. *Oncogene*. 1999;18(2):467-75.
38. Yamamoto M, Legendre NP, Biswas AA, Lawton A, Yamamoto S, Tajbakhsh S, et al. Loss of MyoD and Myf5 in Skeletal Muscle Stem Cells Results in Altered Myogenic Programming and Failed Regeneration. *Stem Cell Reports*. 2018;10(3):956-69.
39. Kepser L-J, Damar F, Cicco TD, Chaponnier C, Prószyński TJ, Pagenstecher A, et al. CAP2 deficiency delays myofibril actin cytoskeleton differentiation and disturbs skeletal muscle architecture and function. *Proc Natl Acad Sci U S A*. 2019;116(17):8397-420.

40. Robaszkiewicz K, Moraczewska J. [Congenital myopathies - skeletal muscle diseases related to disorder of actin filament structure and functions]. *Postepy Hig Med Dosw (Online)*. 2011;65(13):347-56.
41. Sartori R, Schirwis E, Blaauw B, Bortolanza S, Zhao J, Enzo E, et al. BMP signaling controls muscle mass. *Nat Genet*. 2013;45(11):1309-18.
42. Girardi F, Le Grand F. Wnt Signaling in Skeletal Muscle Development and Regeneration. *Prog Mol Biol Transl Sci*. 2018;153:157-79.
43. Heberle H, Meirelles GV, Silva FRd, Telles GP, Minghim R. InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*. 2015;16(1):169.
44. Frayn KN, Arner P, Yki-Järvinen H. Fatty acid metabolism in adipose tissue, muscle and liver in health and disease. *Essays Biochem*. 2006;42:89-103.
45. Keren A, Tamir Y, Bengal E. The p38 MAPK signaling pathway: a major regulator of skeletal muscle development. *Mol Cell Endocrinol*. 2006;252(1-2):224-30.
46. Thomson DM. The Role of AMPK in the Regulation of Skeletal Muscle Size, Hypertrophy, and Regeneration. *Int J Mol Sci*. 2018;19(10).
47. Taylor MV, Hughes SM. Mef2 and the skeletal muscle differentiation program. *Semin Cell Dev Biol*. 2017;72:33-4.
48. Sullivan K, El-Hoss J, Quinlan KG, Deo N, Garton F, Seto JT, et al. NF1 is a critical regulator of muscle development and metabolism. *Hum Mol Genet*. 2014;23(5):1250-9.
49. Kossler N, Stricker S, Rödelberger C, Robinson PN, Kim J, Dietrich C, et al. Neurofibromin (Nf1) is required for skeletal muscle development. *Human Molecular Genetics*. 2011;20(14):2697-709.
50. Picard B, Lefaucheur L, Berri C, Duclos MJ. Muscle fibre ontogenesis in farm animal species. *Reprod Nutr Dev*. 2002;42(5):415-31.
51. Ashmore CR, Addis PB, Doerr L. Development of muscle fibers in the fetal pig. *J Anim Sci*. 1973;36(6):1088-93.
52. Picard B, Berri C, Lefaucheur L, Molette C, Sayd T, Terlouw C. Skeletal muscle proteomics in livestock production. *Brief Funct Genomics* 9:259-278. *Brief Funct Genomics*. 2010;9(3):259-78.
53. Goryshin IY, Miller JA, Kil YV, Lanzov VA, Reznikoff WS. Tn5/IS50 target recognition. *Proc Natl Acad Sci U S A*. 1998;95(18):10716-21.
54. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol*. 2010;11(12):R119.
55. Shashikant T, Etensohn CA, *Genome-wide analysis of chromatin accessibility using ATAC-seq*. Vol. 151. 2018.
56. Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep*. 2019;9(1):9354.
57. Wimberley CE, Heber S. PeakPass: Automating ChIP-Seq Blacklist Creation. *J Comput Biol*. 2019.

58. Dolgalev I, Sedlazeck F, Busby B. DangerTrack: A scoring system to detect difficult-to-assess regions. *F1000Res*. 2017;6:443.
59. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for multimodal regulatory analysis and personal epigenomics. *Nat Methods*. 2013;10(12):1213-8.
60. Dechassa ML, Tryndyak V, de Conti A, Xiao W, Beland FA, Pogribny IP. Identification of chromatin-accessible domains in non-alcoholic steatohepatitis-derived hepatocellular carcinoma. *Mol Carcinog*. 2018;57(8):978-87.
61. Su Y, Shin J, Zhong C, Wang S, Roychowdhury P, Lim J, et al. Neuronal activity modifies the chromatin accessibility landscape in the adult brain. *Nat Neurosci*. 2017;20(3):476-83.
62. Preissl S, Fang R, Huang H, Zhao Y, Raviram R, Gorkin DU, et al. Single-nucleus analysis of accessible chromatin in developing mouse forebrain reveals cell-type-specific transcriptional regulation. *Nat Neurosci*. 2018;21(3):432-9.
63. Qiu Z, Li R, Zhang S, Wang K, Xu M, Li J, et al. Identification of Regulatory DNA Elements Using Genome-wide Mapping of DNase I Hypersensitive Sites during Tomato Fruit Development. *Mol Plant*. 2016;9(8):1168-82.
64. Gross DS, Garrard WT. Nuclease hypersensitive sites in chromatin. 1988;57:159-87.
65. Wu J, Huang B, Chen H, Yin Q, Liu Y, Xiang Y, et al. The landscape of accessible chromatin in mammalian preimplantation embryos. *Nature*. 2016;534(7609):652-7.
66. Shan X, Roberts C, Lan Y, Percec I. Age Alters Chromatin Structure and Expression of SUMO Proteins under Stress Conditions in Human Adipose-Derived Stem Cells. *Sci Rep*. 2018;8(1):11502.
67. Ackermann AM, Wang Z, Schug J, Naji A, Kaestner KH. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol Metab*. 2016;5(3):233-44.
68. Hu S, Yang S, Lu Y, Deng Y, Li L, Zhu J, et al. Dynamics of the Transcriptome and Accessible Chromatin Landscapes During Early Goose Ovarian Development. *Front Cell Dev Biol*. 2020;8:196.
69. 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(1):16.
70. Frank CL, Liu F, Wijayatunge R, Song L, Biegler MT, Yang MG, et al. Regulation of chromatin accessibility and Zic binding at enhancers in the developing cerebellum. *Nat Neurosci*. 2015;18(5):647-56.
71. Ming H, Sun J, Pasquariello R, Gatenby L, Herrick JR, Yuan Y, et al. The landscape of accessible chromatin in bovine oocytes and early embryos. *Epigenetics*. 2020:1-13.
72. Liu Y, Chang JC, Hon CC, Fukui N, Tanaka N, Zhang Z, et al. Chromatin accessibility landscape of articular knee cartilage reveals aberrant enhancer regulation in osteoarthritis. *Sci Rep*. 2018;8(1):15499.
73. Wang J, Zibetti C, Shang P, Sripathi SR, Zhang P, Cano M, et al. ATAC-Seq analysis reveals a widespread decrease of chromatin accessibility in age-related macular degeneration. *Nat Commun*.

2018;9(1):1364.

74. Klenova EM, Nicolas RH, Paterson HF, Carne AF, Heath CM, Goodwin GH, et al. CTCF, a conserved nuclear factor required for optimal transcriptional activity of the chicken c-myc gene, is an 11-Zn-finger protein differentially expressed in multiple forms. *Mol Cell Biol.* 1993;13(12):7612-24.
75. Zeng Z, Huang N, Zhang Y, Wang Y, Su Y, Zhang H, et al. CTCF inhibits endoplasmic reticulum stress and apoptosis in cardiomyocytes by upregulating RYR2 via inhibiting S100A1. *Life Sci.* 2020;242:117158.
76. Chen H, Tian Y, Shu W, Bo X, Wang S. Comprehensive identification and annotation of cell type-specific and ubiquitous CTCF-binding sites in the human genome. *PLoS One.* 2012;7(7):e41374.
77. Recillas-Targa F, Pikaart MJ, Burgess-Beusse B, Bell AC, Litt MD, West AG, et al. Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities *Proc Natl Acad Sci U S A.* 2002;99(10):6883-8.

Tables

Table 1. Results of DPI analysis of common peaks

group	down	up	total
LW70 vs LW45	127	18	145
LW100 vs LW70	1412	314	1726
LW100 vs LW45	2224	463	2687

Table 2. Significance of the overlap between specific ACR-related genes and DEGs in six clusters

	Gene associated with specific ACRs in LW45	Gene associated with specific ACRs in LW70	Gene associated with specific ACRs in LW100
Cluster1	0.05	0.40	0.99
Cluster2	0.08	0.08	0.95
Cluster3	1	0.84	0.03
Cluster4	0.42	0.03	0.97
Cluster5	1	1	0.02
Cluster6	1	1	0.41

Table 3. The overlap between DPI-related genes in common peaks and DEGs in RNA-seq

group	down-down ^a	up-up ^b	down-up ^c	up-down ^d
LW70 vs LW45	8	0	3	0
LW100 vs LW70	129	48	15	5
LW100 vs LW45	269	101	99	11

^a: the overlap of down-regulated genes in ATAC-seq and down-regulated genes in RNA-seq; ^b: the overlap of up-regulated genes in ATAC-seq and up-regulated genes in RNA-seq; ^c: the overlap of down-regulated genes in ATAC-seq and up-regulated genes in RNA-seq; ^d: the overlap of up-regulated genes in ATAC-seq and down-regulated genes in RNA-seq.

Figures

Overview of the ATAC-seq libraries. (A) Schematic representation of the experimental. (B) Distributions of ACR numbers across different chromosomes. (C) Percentages of ACRs in different genomic regions.

Figure 2

Overview of the ACR lengths in different genomic regions. (A-C): Distributions of ACR lengths in different genomic regions in LW45, LW70 and LW100 stages, respectively. Different letters above violins indicate significant differences between the groups based on Tukey's honestly significant difference test ($P < 0.05$)

Figure 3

Expression levels of genes associated with ACRs numbers. (A) Statistics for ACR numbers in the proximal promoter regions of genes. (B) Heatmap of the percentages of ACR-associated gene expression levels in different groups. “One ACR” and “> one ACR” correspond to the groups in which genes contained one ACR and multiple ACRs in their proximal promoter regions, respectively. The expression levels of genes were divided into six groups based on the FPKM values: 0–1, 1–2, 2–5, 5–10, 10–30, and >30.

Figure 4

Expression levels of genes associated with ACRs length. (A-C) Percentages of ACR-associated gene expression levels in different groups. The genes were grouped according to whether they contained a single ACR (upper panel) or multiple ACRs (middle and lower panel) in their proximal promoters according to the peak length (middle panel: based on longest peak length; lower panel: based on total length) and were then divided into 3 equal groups (the top, middle and bottom groups).

Figure 5

ACR changes during muscle development. (A) Visualization of ACRs in the LW45, LW70 and LW100 groups within a region on Chr1. (B) Statistics for common and specific ACRs in the LW45, LW70 and LW100 groups. (C) Visualization of specific ACRs at different stages. (D) Visualization of common ACRs in MYOD1 at different stages. (E) Heatmap of the expression of XRCC1, ENPP2, BCL6 and MYOD1 at the LW45, LW70 and LW100 stages.

Figure 6

Functional analysis of genes associated with stage-specific ACRs. (A) Numbers of genes related to stage-specific ACRs. (B) Numbers of shared and stage-specific genes associated with specific ACRs in the LW45, LW70 and LW100 groups. (C) Numbers of genes with DPIs in ACRs shared among stages. (D) Numbers of shared and group-specific genes associated with common ACRs in the LW70 vs LW45, LW100 vs LW70 and LW100 vs LW45 groups.

Figure 7

Identification of regulatory DNA elements at different stages. (A) Heatmap of the expression levels of DEGs at different developmental stages. (B) Expression profiles of all DEGs in six clusters.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile1.TableS1.xlsx](#)
- [Additionalfile2.TableS2.xlsx](#)
- [Additionalfile3.TableS3.xlsx](#)
- [Additionalfile4.TableS4.xlsx](#)
- [Additionalfile5.TableS5.xlsx](#)
- [Additionalfile6.docx](#)