Zhu et al. *Zool. Res.* 2023, 44(1): 53−62
https://doi.org/10.24272/j.issn.2095-8137.2022.228

ZR | Zoological Research

Article Open Access

Chicken chromatin accessibility atlas accelerates epigenetic annotation of birds and gene fine-mapping associated with growth traits

Xiao-Ning Zhu[1], Yu-Zhe Wang[1,2,*], Chong Li[1], Han-Yu Wu[1,2], Ran Zhang[1], Xiao-Xiang Hu[1,2,*]

[1] *State Key Laboratory of Agrobiotechnology*, *College of Biological Sciences*, *China Agricultural University*, *Beijing* 100193, *China*
[2] *National Research Facility for Phenotypic and Genotypic Analysis of Model Animals (Beijing)*, *China Agricultural University*, *Beijing* 100193, *China*

ABSTRACT

The development of epigenetic maps, such as the ENCODE project in humans, provides resources for gene regulation studies and a reference for research of disease-related regulatory elements. However, epigenetic information, such as a bird-specific chromatin accessibility atlas, is currently lacking for the thousands of bird species currently described. The major genomic difference between birds and mammals is their shorter introns and intergenic distances, which seriously hinders the use of humans and mice as a reference for studying the function of important regulatory regions in birds. In this study, using chicken as a model bird species, we systematically compiled a chicken chromatin accessibility atlas using 53 Assay of Transposase Accessible Chromatin sequencing (ATAC-seq) samples across 11 tissues. An average of 50 796 open chromatin regions were identified per sample, cumulatively accounting for 20.36% of the chicken genome. Tissue specificity was largely reflected by differences in intergenic and intronic peaks, with specific functional regulation achieved by two mechanisms: recruitment of several sequence-specific transcription factors and direct regulation of adjacent functional genes. By integrating data from genome-wide association studies, our results suggest that chicken body weight is driven by different regulatory variants active in growth-relevant tissues. We propose *CAB39L* (active in the duodenum), *RCBTB1* (muscle and liver), and novel long non-coding RNA *ENSGALG00000053256* (bone) as candidate genes regulating chicken body weight. Overall, this study demonstrates the value of epigenetic data in fine-mapping functional variants and provides a compendium of resources for further research on the epigenetics and evolution of birds and mammals.

Keywords: Chicken; Chromatin accessibility atlas; ATAC-seq; Tissue-specific OCRs; GWAS; Growth traits

INTRODUCTION

An increasing number of complex features, such as human diseases and agricultural production traits, are driven by non-coding variants that presumably affect gene regulation (Boyle et al., 2017). Indeed, significant mutations are more abundant in highly active chromatin regions comprised of various

Received: 29 August 2022; Accepted: 26 October 2022; Online: 26 October 2022
Foundation items: This study was supported by the National Natural Science Foundation of China (U2002205, 32272862)
*Corresponding authors, E-mail: yuzhe891@cau.edu.cn; huxx@cau.edu.cn

regulatory elements (such as enhancers, promoters, and repressors) in relevant cell or tissue types (Roadmap Epigenomics Consortium et al., 2015). The development of epigenetics and associated analytical tools has provided powerful strategies for recognizing and interpreting the function of non-coding regions. Several epigenetic maps have been reported for humans and mice (e.g., ENCODE). These spatiotemporal epigenome maps provide resources for the study of gene regulation in tissue and organ development and provide a reference for studying regulatory elements related to human diseases (The ENCODE Project Consortium, 2004; The ENCODE Project Consortium et al., 2020).

The Functional Annotation of Animal Genomes (FAANG) consortium focuses on farm animal genome-wide datasets, including data on gene expression, methylation, chromatin modification, chromatin accessibility, and interactions (Foissac et al., 2019; Giuffra et al., 2019). Recent large-scale analysis of multiple epigenomes in cattle (*Bos taurus*), pigs (*Sus scrofa*), and chickens (*Gallus gallus*) has provided new insights into the evolutionary properties of avian and mammalian epigenomes (Kern et al., 2021). Furthermore, the dynamic epigenetic landscape of different pig breeds has been systematically described across tissues based on functional annotation of different chromatin states (Pan et al., 2021; Zhao et al., 2021).

As the first agricultural species to be sequenced, genome-level studies of chickens continue to expand. Recently, Wang et al. (2020a) analyzed 863 genomes from a worldwide sampling and found that domestic chickens were likely derived from the red jungle fowl (RJF) subspecies *Gallus gallus spadiceus*, thus helping to resolve the geographic and temporal origins of chicken domestication. Analysis of 20 *de novo* assembled genomes revealed unique pan-genome patterns in chickens and further updated knowledge regarding the evolutionary rates in birds (Li et al., 2022). To date, however, no large-scale multi-tissue Assay of Transposase Accessible Chromatin sequencing (ATAC-seq) data have been applied to map and characterize open chromatin regions (OCRs) in birds or assist in functional gene identification of important agricultural traits. The reduced genome size in avian species (~1 Gb) compared to mammals (~2.5–3.0 Gb) is largely due to the shorter introns and intergenic distances (Zhang et al., 2014), which severely hinder the use of humans or mice as a reference for studying the function of important regulatory regions in birds. Given the great evolutionary distance in non-coding regions between birds and mammals, it is necessary to establish a chromatin accessibility atlas of chickens as a bird model organism.

In this study, we systematically compiled a chromatin accessibility atlas of 53 ATAC-seq samples across 11 tissues. The functional characteristics of open elements in different chicken tissues were analyzed, revealing two distinct regulatory modes that contribute to tissue-specific functions. By integrating available data from genome-wide association studies (GWAS), we demonstrated the role of epigenetic data (ATAC-seq) in fine-mapping functional variants and genes of complex traits. This new benchmark resource for chicken epigenetics should facilitate studies on the evolution of non-coding regulatory regions in birds and mammals.

MATERIALS AND METHODS

Ethics statement

The State Key Laboratory Animal Welfare Committee approved all animal care and experimental procedures for agrobiotechnology conducted at China Agricultural University (approval number SKLAB-2014-06-04). The chickens were sacrificed according to local animal welfare standards.

Data sources and sample collection

The duodenal samples collected in this study were derived from an advanced intercross line (AIL) based on the Lingnan yellow chicken line A03 (HQLA) and local Chinese Huiyang bearded chicken (HB) (Wang et al., 2020b). Six duodenal samples were collected from 7-week-old chickens. Each sample was quickly frozen in liquid nitrogen and stored at –80 °C until nuclear extraction. The sequencing data for other tissues were obtained from previous reports (Foissac et al., 2019; Halstead et al., 2020a; Lai et al., 2018; Patoori et al., 2020; Rothstein & Simoes-Costa, 2020; Sackton et al., 2019; Young et al., 2019). All samples used in this study and their sources are listed in Supplementary Table S1.

Nuclear extraction, library construction, and sequencing

Native nuclei were purified from the duodenal samples as described previously (Corces et al., 2017). A Nextera DNA Library Preparation Kit (Illumina, USA) was used to perform transposition following the manufacturer's instructions. In total, 50 000 nuclei were pelleted and resuspended in transposase for 30 min at 37 °C. The transposed DNA fragments were immediately purified using a MinElute PCR Purification Kit (Qiagen, Germany). Samples were amplified by polymerase chain reaction (PCR) using 1×NEBNext High-Fidelity PCR Master Mix (New England Biolabs, USA). Subsequent libraries were purified using a MinElute PCR Purification Kit (Qiagen, Germany) and sequenced on the Illumina NovaSeq 6000 platform (Illumina, USA) using the PE150 model.

ATAC-seq data analysis

Low-quality bases and residual adapter sequences were trimmed from the raw sequencing data using Trim Galore (v0.6.6) (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), a wrapper tool around Cutadapt (v2.10) (Martin, 2011), to retain trimmed reads at least 20 bp in length and with a Phred quality score greater than 25. Trimmed reads were aligned to the chicken reference genome (GRCg6a version) using Bowtie2 (v2.4.2) (Langmead & Salzberg, 2012) with the options "-p 5 --very-sensitive -X 2000". Duplicate alignments were removed using Sambamba (v0.8.0), and mitochondrial and low-quality (MAPQ < 30) alignments were removed using SAMtools (v1.9) (Li et al., 2009).

Insert fragments were counted, and BAM file coordinates were transformed using ATACseqQC (Ou et al., 2018). Narrow peaks were called using MACS2 (v2.2.7.1) (Zhang et al., 2008) with the options "-f BAMPE -nomodel -q 0.05 --keep-dup 1 -B --SPMR". The annotatePeaks.pl scripts in HOMER (v4.11) (Heinz et al., 2010) were used to annotate narrow peaks. All tissue peaks were merged into a standard peak. We counted the number of raw reads mapped to each standard peak using the intersecting function in BEDTools

(v2.30.0) (Quinlan & Hall, 2010). The raw count matrix was normalized to reads per million mapped reads (RPM). Pearson correlation coefficients between technical and biological replicates across tissues were calculated based on the $\log_{10}$ RPM matrix. The non-redundant fraction (NRF) was calculated using the available reads obtained from the BAM file prior to conversion of coordinates divided by the total reads in the comparison (excluding available reads mapped to the peak region of the mitochondrial reads), representing library complexity. The fraction of reads in peaks (FRiP) was calculated by dividing the available reads mapped to the peak region by the total available reads. We used the bamCoverage function in deepTools (v3.5.0) (Ramírez et al., 2016) to normalize the whole-gene ATAC-seq signals in 50 bp windows with bins per million mapped reads (BPM), similar to TPM in RNA-sequencing. The bw files were obtained using the computeMatrix function to analyze enrichment near the transcription start site (TSS) and gene body. Data were visualized using plotHeatmap and plotProfile (Ramírez et al., 2016).

Identification and annotation of tissue-specific chromatin-accessible regions

Using a previously described Shannon entropy-based method (Shen et al., 2012), tissue specificity indices for each peak were calculated (Schug et al., 2005; Shen et al., 2012), with entropy values closer to zero indicating higher tissue specificity of the peak, and vice versa. Based on the entropy score distribution, peaks with scores less than 3.0 were selected as tissue-specific peaks. The findMotifsGenome.pl script in HOMER (v4.11) was used to search for transcription factor (TF) motifs in tissue-specific chromatin-accessible regions (Heinz et al., 2010). A motif enrichment matrix was then generated, with each row representing the *P*-value of a motif and each column representing a tissue. Simultaneously, tissue-specific chromatin-accessible regions of different species were annotated. Genes were annotated using Gene Ontology (GO).

Visualization of chromatin accessibility peaks and conservation of chromatin-accessible regions among different species

We used the CyVerse website (https://www.cyverse.org/) to generate URL links for the bw files and generated a web link for visualization of sample results via the UCSC website. To study the conservation of chromatin-accessible regions between chickens and mammals, we used the mouse GRCm39 genome as a reference, taking intestinal tissue as an example. After indexing the GRCm39 genome (main chromosomes) using lastdb (http://last.cbrc.jp/), we used the lastal program and the last-split program to project chicken intestinal chromatin-accessible regions onto the mouse genome. We further identified open regions in both chicken and mouse intestinal tissues based on available mouse intestinal OCR information (Liu et al., 2019).

Chromatin accessibility peak distribution of conserved non-coding elements in chickens

Based on conserved non-protein-coding elements (CNEs) in chickens and avian-specific highly conserved elements (ASHCEs) identified in previous studies (Groß et al., 2020; Seki et al., 2017), we evaluated the relationship between conserved elements and open regions. ASHCEs were obtained from the galGal3 chicken reference genome and converted to the GRCg6a version using the UCSC LiftOver tool (http://www.genome.ucsc.edu/cgi-bin/hgLiftOver). BEDTools (Quinlan & Hall, 2010) was used to analyze the relationship between chromatin-accessible regions and non-coding elements. R was used to perform *t*-tests to assess differences in non-coding elements between open and non-open regions.

Genome-wide association study

We employed a highly accurate, low-coverage Tn5-based sequencing method (BaseVar-Stitch pipeline) to obtain whole-genome high-density single-nucleotide polymorphism (SNP) markers for 554 AIL F9 individuals. (Yang et al., 2021). Phenotypic data, body weight at 8 weeks (BW8), and duodenum length (DL) were previously reported by Wang et al. (2020c). Briefly, a mixed linear model (MLM) was applied for genome-wide association analysis using the GCTA tool (Jiang et al., 2019; Yang et al., 2011). The model used to analyze BW8 and DL data included sex and batch size as discrete covariates. A quantile-quantile (Q-Q) plot generated in R (v3.0.2) was used to assess the potential impact of population stratification on genetic association studies. The linkage disequilibrium (LD) correlation (r^2) between genotypes was calculated using PLINK software (Purcell et al., 2007).

Dual-luciferase reporter assay

Sequences (170 525 591–170 526 591) containing two alleles (A/G) of chromosome 1 at bp position 170 526 091 were synthesized and cloned into the pGL3-basic luciferase reporter vector (Promega, USA). DF-1 cells (chicken fibroblast cell line) were cultured in Dulbecco's modified Eagle medium (Gibco, USA) supplemented with 10% fetal bovine serum (Gibco, USA), 100 IU/ mL penicillin, and 100 μg/mL streptomycin (Gibco, USA). Lipofectamine 3000 reagent (Invitrogen, USA) was used for transient transfection following the manufacturer's protocols. The recombinant plasmid was transfected into the DF-1 cells together with the PRL-TK plasmid (Promega, USA). The DF-1 cells were then cultured in 24-well culture plates (Thermo Scientific, USA) at 37 °C and 5% CO_2 for 48 h. Firefly and Renilla luciferase activities were measured at 48 h post-transfection using a Dual-Luciferase Assay System Kit (Promega, USA) according to the manufacturer's instructions. Luminescence was detected using a microplate reader (Tecan, Switzerland) and firefly luciferase activities were normalized to Renilla luminescence in each well.

RESULTS

Sample information and data quality control

A total of 53 ATAC-seq libraries derived from 11 tissue types (duodenum in this study and bone, bud, liver, lung, muscle, neural crest, retina, skin, somatopleure, and T cells from public datasets) (Foissac et al., 2019; Halstead et al., 2020a; Lai et al., 2018; Patoori et al., 2020; Rothstein & Simoes-

Costa, 2020; Sackton et al., 2019; Young et al., 2019) were collected (Table 1; Supplementary Table S1). Samples were analyzed for genome-wide chromatin accessibility using a standard pipeline (see details in the Materials and Methods) and passed through stringent quality filtering. A mean depth of 46.46±22.63 million usable reads per sample was obtained (Supplementary Table S1), sufficient to detect accessible regions. The duodenum libraries generated in this study compared favorably with available data, showing the lowest fraction of mitochondrial reads (2.48%±0.34%) and highest usable reads (93.23±16.31 M) compared with data downloaded for analysis (31.21%±15.46% and 40.49±15.10 M, respectively). The chromatin accessibility fragments showed size periodicity corresponding to integer multiples of nucleosomes, indicating high library quality and non-disrupted open chromatin state (Figure 1A; Supplementary Figure S1). The ATAC-seq signals were more enriched near the TSSs than the gene regions (Figure 1B).

We used NRF and FRiP to evaluate library quality, reflecting library complexity and degree of enrichment, respectively. Average NRF and FRiP scores of the 53 samples were 0.75±0.11 and 0.19±0.08 (Figure 1C), respectively, indicating that quality met the official standards established by ENCODE (https://www.encodeproject.org/atac-seq/). Thus, these results indicate that the library was qualified and could be used for downstream analysis.

General characteristics of OCRs and visualization

We obtained an average of 50 796 high-confidence OCRs per sample (Supplementary Table S2), with a total of 382 603 unique OCRs merged across all tissues, accounting for 20.36% of the genome. Most OCRs were annotated to non-coding regions, especially introns (42.20%) and intergenic regions (34.06%), followed by TSSs (16.70%), exons (4.28%), and TTS regions (2.75%) (Figure 2A). We further investigated genomic repetitive elements (provided in the UCSC genome browser) in the ATAC-seq dataset. In total, 29.00%±5.51% of the ATAC-seq signals overlapped with repetitive element regions, ranging from 23.43% in T cells to 47.62% in bone. We further employed two representative datasets, CNEs (Groß et al., 2020) and ASHCEs (Seki et al., 2017), to assess the relationship between OCRs and evolutionarily conserved elements. Results showed a significant difference in the proportion of OCRs and non-OCRs that overlapped with conserved regions (CNEs: P=9.9×10^{-9}, ASHCEs: P=3.7×10^{-4};

Table 1 ATAC-seq metadata and mapping statistics of 11 tissues

Tissue	Sample size	Clean reads (M)	Mapped reads (M)	Usable reads (M)*	Usable ratio (%)	NRF (%)	FRiP (%)
Bone	12	92.98±11.59	90.01±11.29	48.93±6.7	52.63±3.52	76.80±1.96	19.48±6.09
Bud	6	86.27±5.61	84.15±5.47	45.95±3.91	53.31±3.54	76.37±2.43	22.88±1.86
Duodenum	6	145.14±30.42	141.05±29.57	93.23±16.31	64.58±2.33	68.19±2.34	23.97±0.82
Liver	2	115.31±10.75	111.81±0.87	32.68±4.67	28.28±1.4	44.87±5.98	4.06±0.39
Lung	1	111.27	107	43	38.68	41.79	18.89
Muscle	6	98.50±4.63	95.04±4.71	59.60±2.24	60.58±2.65	77.03±3.16	9.03±2.57
Neural crest	4	35.77±12.87	32.38±12.75	15.47±7.24	43.94±15.03	83.06±2.42	27.70±12.33
Retina	2	62.03±0.47	53.80±1.56	39.96±2.7	64.44±4.84	88.17±0.91	22.69±10.85
Skin	2	38.43±0.74	37.73±0.73	31.63±1.07	82.29±1.20	90.00±0.08	34.37±0.41
Somatopleure	9	76.83±13.30	74.91±12.98	26.09±6.78	33.85±5.18	79.37±2.08	20.33±2.79
T cell	3	130.52±48.51	122.66±45.54	44.82±6.51	32.66±6.83	55.12±1.37	9.51±2.03

*Usable reads: Number of mapped read minus number of low-mapping quality, duplicate, and mitochondrial reads.

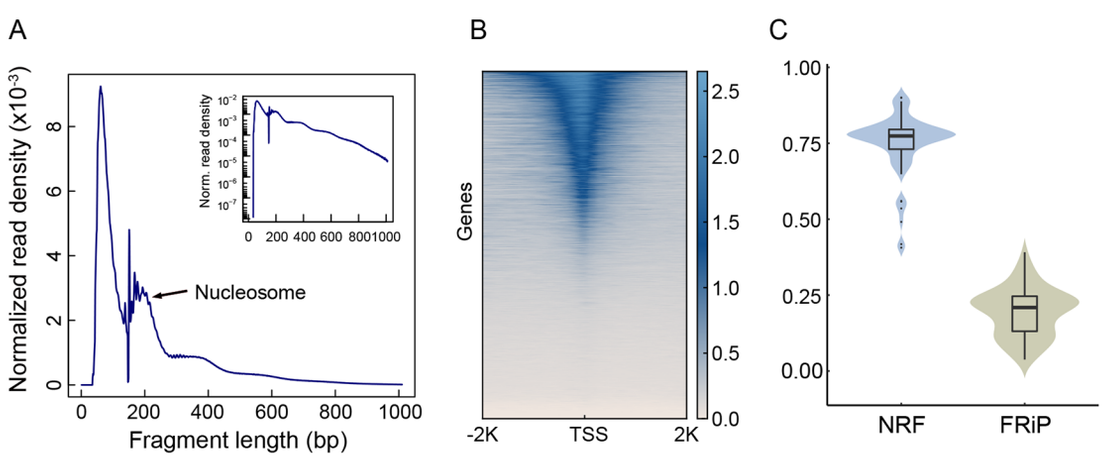

Figure 1 ATAC-seq data quality metrics

A: Insert-size distribution of ATAC-seq profiles for the duodenum (representative example). B: ATAC-seq signal enrichment around TSSs in duodenum. C: NRF and FRiP distributions for all samples.

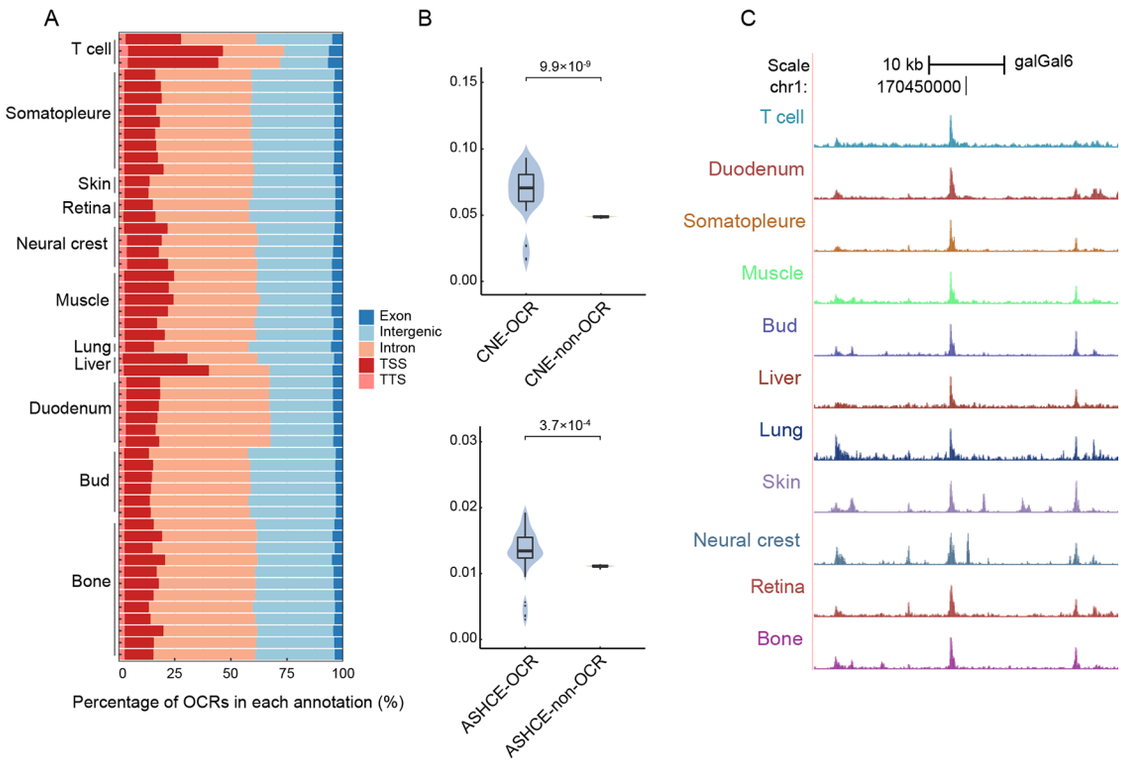

Figure 2 General characteristics of chicken OCRs

A: Percentage of OCRs per sample in different genomic regions. B: Proportion of OCRs and non-OCRs overlapping with CNEs in chickens and ASHCEs. Numbers above black line represent *P*-values (*t*-test). C: Local visualization of chicken chromatin accessibility atlas.

Figure 2B), demonstrating the important role of OCRs in evolution. Using the duodenum results, we compared the OCRs of chickens and mice and found that OCRs common to both species accounted for only 23.21% in mice, with most located in the exons and TSS regions. The collinear alignment results showed that only 15.00%±0.74% of chicken sequences could be aligned to the mouse genome sequences. These results suggest that the regulatory element sequences between chickens and mice are not conserved, especially the various non-coding regions.

To better display the chicken chromatin accessibility atlas, we used CyVerse to generate URL links to peak information (BW files), and further released an open visual link on the UCSC genome browser (https://data.cyverse.org/dav-anon/iplant/home/zhuxiaoning/chicken_ATAC-seq_2021/Myhub/hub.txt) (available from the blue navigation bar “My Data”, then “Track Hubs” to reach the Public Track Hubs page). This enables direct access to the genome-wide chromatin accessibility peaks of all samples, allowing comparison of differences between tissues (see Figure 2C as an example).

Identification and correlation analysis of tissue-specific chromatin accessibility

We first employed a consensus set of 382 603 OCRs by merging the peaks called in all individuals. Heatmap clustering of Pearson correlation coefficients from the comparisons of the 53 data points revealed that all samples were clustered according to tissue type (Figure 3A). The somatopleure and bud from different experiments clustered together, indicating similarities in the early stages of development. The liver, lung, and duodenum exhibited similar global chromatin accessibility patterns. We used a Shannon entropy-based strategy to compute tissue-specific peaks. Peaks with an entropy value less than 3.0 were selected as tissue-specific regions and clustered using *t*-distributed stochastic neighbor embedding (*t*-SNE) (Figure 3B). Results also showed strong correlations within the same tissue.

The number of tissue-specific peaks was highly variable among tissues, ranging from 2 965 (somatopleure) to 8 962 (liver), accounting for approximately 2.60%–48.87% (mean 11.01%) of common OCRs for each tissue (Supplementary Table S3). Only a few tissue-specific OCRs (2.51%) were within 1 kb of the TSS (Figure 3C). Considering the positional relationship between the promoter and TSS, this result suggests that a considerable proportion of promoters may be conserved across tissues.

Different regulatory model of tissue-specific OCRs

Our analysis revealed two distinct regulatory models contributing to tissue-specific functions. First, we identified tissue-specific TF motifs as drivers of tissue development and tissue identity maintenance. We observed high enrichment of tissue-specific TF motifs, such as the LHX family (retina), SOX family (neural crest), and TCF/LEF family (skin), which play important roles in ocular defects, nervous system development, and cutaneous squamous cell carcinoma, respectively (Figure 3D) (Bikle, 2020; Hutton & Pevny, 2011;

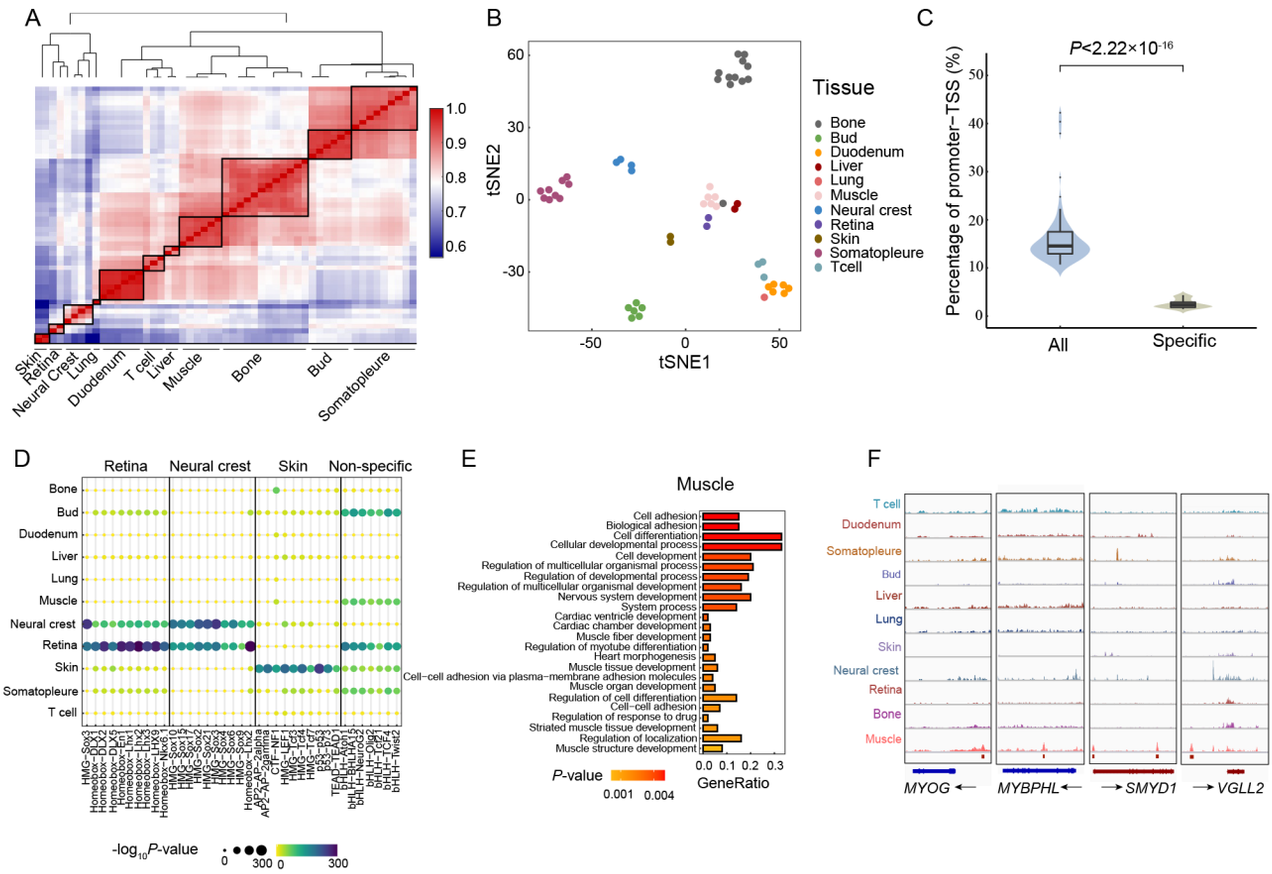

Figure 3 Characteristics of tissue-specific OCRs and their different tissue-specific regulatory models

A: Heatmap clustering of correlation coefficients across all tissue profiles using all peaks. B: t-SNE plot of all 53 ATAC-seq profiles, based on tissue-specific peaks. C: Significant difference (P<2.22×10^{-16}) in proportion of cells localized in the promoter-TSS region for all peaks and tissue-specific peaks. D: Enrichment of indicated TF motifs in the retina, neural crest, and skin. Non-specific is a comparison result, indicating no individual tissue-specific TF enrichment. Color and size of points represent motif enrichment P-values. E: GO enrichment analysis of genes closest to muscle OCRs. F: Genome browser views of ATAC-seq signals for muscle-associated genes. Red block at the bottom represents muscle-specific OCRs. Arrows indicate transcription direction.

Pérez et al., 2012). Furthermore, examination of various tissues, e.g., bud, muscle, and somatopleure, showed tight clustering, likely due to their similar functions during tissue development.

Second, we speculated that genes adjacent to these specific OCRs may also have tissue-specific functions. GO annotation of these genes revealed that many tissues were enriched in biological processes associated with the corresponding tissue type. For example, muscle-related biological processes, including muscle fiber development, regulation of myotube differentiation, striated muscle tissue development, and muscle structure development, were enriched in the muscle (P<0.005, Figure 3E). Various core genes known to play important roles in different stages of muscle development were also identified, including *MYBPHL*, *MYOG*, *SMYD1*, and *VGLL2* (Figure 3F).

Overlap of fine-mapping duodenum length and body weight variants with OCRs

To demonstrate the auxiliary role of ATAC-seq in gene fine-mapping of complex traits, we integrated the sequencing data with the GWAS results of the AIL, which has accumulated nine generations of recombination data. This is a powerful experimental design for identifying quantitative trait loci (QTLs). We first obtained the genome-wide distribution of 7 969 074 SNP markers (~121 bp distance/SNP, Supplementary Table S4) for 554 individuals in the F_9 generation of the AIL population using low-coverage sequencing genotyping. We then measured BW8 and DL for GWAS analysis (Supplementary Figure S2).

A major-effect QTL (chr1: 169 894 149–171 223 058 bp) was mapped at the distal end of chromosome 1, explaining 12.6% of the genetic variation in DL. Based on the condition of GWAS P<1.0×10^{-7}, we identified 174 extremely significant SNPs spanning chr1: 170.17–171.12 Mb (Figure 4A). Notably, none of these sites were located in exonic regions, and the top 50 loci (chr1: 170.17–170.65 Mb) were in a high LD state (average LD r^2=0.73±0.19), making it difficult to determine which loci played more important roles based on GWAS only. Subsequently, we integrated the GWAS results of the DL traits and ATAC-seq of the duodenum and found that only 5.17% of the SNPs (n=9) were located in the OCR. The most significant

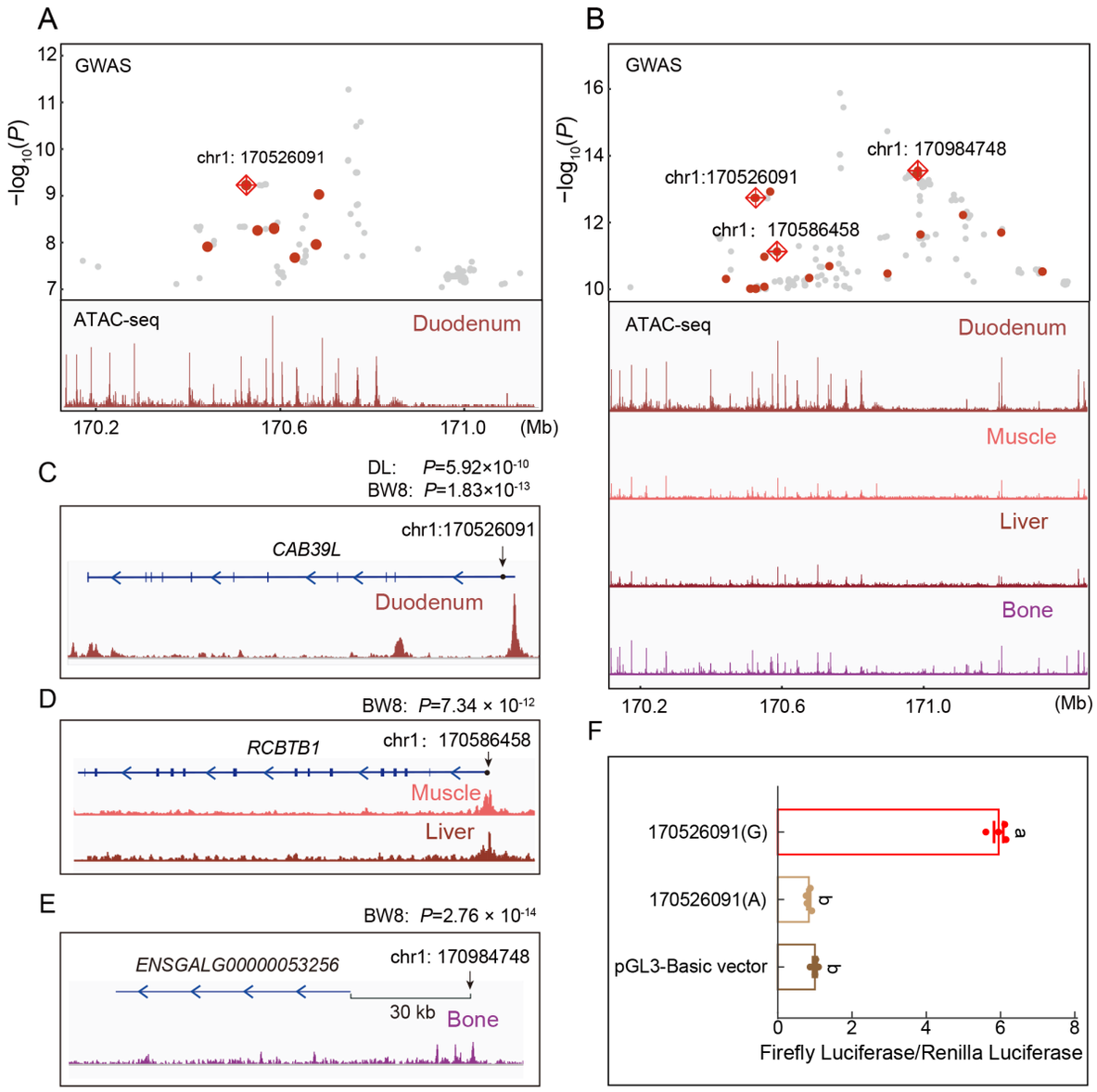

Figure 4 Integrative GWAS analysis of two traits and OCRs of four tissues

A: Scatter plot (above) illustrating 174 GWAS-significant (P<1.0×10^{-7}) SNPs in QTL region (chr1: 170.17–171.12 Mb) for DL in advanced intercross line. Red sites represent loci that intersect with OCRs of the duodenum (below), with the most significant site marked with diamonds and annotated with genomic coordinates. B: Scatter plot (above) illustrating 306 GWAS-significant SNPs (P<1.0×10^{-10}) in QTL region (chr1: 170.17–171.41 Mb) for body weight at 8 weeks (BW8). Red sites represent loci that intersect with merged OCRs of the duodenum, muscle, liver, or bone (below), with most significant site in each tissue marked with diamonds and annotated with genomic coordinates. C–E: Enlarged illustration near the three most significant loci in the duodenum (C), muscle/liver (D), and bone (E), respectively. Blue bars represent exons and arrows indicate transcription direction. F: Comparison of *CAB39L* transcriptional activity of different alleles of a candidate SNP (position 170 526 091 bp) in chicken DF-1 cells, where 170 526 091(A) is the reference and 170 526 091(G) is the variant. Analysis of variance (ANOVA) and multiple comparisons with Duncan's test were performed. Data are mean±standard error of the mean (*SEM*). Letters indicate significant differences at P<0.0001. Values followed by the same letter were not significantly different.

of these was located on chromosome 1 at position 170 526 091 bp within the *CAB39L* promoter region (P=5.92×10^{-10}; ranked equal 19th out of 174 significant sites, Figure 4A). Therefore, we suspect that this locus plays a critical role in regulating DL.

As body weight is a comprehensive growth trait, we integrated the GWAS results of body weight with OCRs from multiple tissues (bone, duodenum, muscle, and liver). Based on the GWAS condition P<1.0×10^{-10}, we discovered 306 significant SNPs spanning chr1: 170.17–171.41 Mb (Figure 4B). The most significant BW8 site in the duodenal OCR was the same as that found for duodenal traits

(P=1.83×10^{-13}, rank: 114/306, Figure 4C). In muscle and liver, the most significant site was located in the *RCBTB1* promoter region (chr1: 170 586 458, P=7.34×10^{-12}, rank: 181/306; Figure 4D). The most significant SNP in bone tissue was located in the intergenic region, and the closest gene was a novel long non-coding RNA (*ENSGALG00000053256*, chr1: 170 984 748, P=2.76×10^{-14}, rank: 10/306; Figure 4E). These results suggest differences in the genetic mechanisms affecting body weight in different tissues. In other words, the GWAS results for BW8 in this region were generated by the joint influence of multiple tissues from different regulatory pathways.

To verify the effects of two alleles (ref:A/alt:G) on the transcriptional activity of *CAB39L* on chromosome 1 at 170 526 091 bp, we reverse-cloned the sequence (170 525 591–170 526 591) into the pGL3-basic vector, with subsequent transfection into DF1 cells. At 48 h after transfection, 170 526 091(A) luciferase activity did not change significantly compared with the empty vector, but 170 526 091(G) activity increased significantly compared with that of 170 526 091(A) (P<0.0001, ~5.9-fold difference). These results suggest that this locus may play a role in regulating DL and BW8 by affecting the transcriptional activity of *CAB39L*, further affecting complex growth phenotypes (Figure 4F).

DISCUSSION

In this study, we characterized OCRs in chickens. Most OCRs were found in the intergenic or intronic regions, which are critical for comprehensive annotation of the chicken noncoding genome. To the best of our knowledge, this is the first chicken chromatin accessibility map obtained using integrated ATAC-seq data, enriching epigenetic annotation of chickens, and laying a foundation for follow-up study of functional genes. In mammals, certain gene regulatory properties are highly conserved, especially in promoters and genetic enhancers; however, large differences have been shown in specific sequences and genomic positions of functional regulatory elements (Halstead et al., 2020b; Yue et al., 2014). In our study, this difference was even more pronounced between birds and mammals, consistent with their evolutionary distance. Although researchers have identified thousands of conserved promoters and enhancers across all five amniotes (including chickens) (Kern et al., 2021), this ratio is very low for the total number of chicken regulatory elements. Our ability to interpret the functional importance of non-protein coding element variants is limited by current genomic annotations (Groß et al., 2020). Given the differences between mammals and birds and conservation across bird species, the chicken chromatin accessibility atlas is crucial for studying avian-specific gene regulation patterns and for providing a more comprehensive picture of the evolution of regulatory elements and networks.

This study is the first to focus on the general characteristics of tissue-specific OCRs in chickens. Tissue specificity is more reflected in the differences in peaks in intergenic and intronic regions (considered candidate enhancer peaks) than in conserved TSS region peaks (considered promoter peaks), similar to comparisons among different domesticated animals (Foissac et al., 2019). Specifically, tissue-specific regulation was achieved through two mechanisms. The first was to recruit several sequence-specific TFs. TFs are known drivers of tissue development and identity maintenance (Uhlen et al., 2015). We observed high enrichment of the LHX family motif in the retina, which plays an important role in pituitary hormone deficiency associated with ocular defects (Pérez et al., 2012; Xu et al., 2018). The second mechanism is the direct regulation of adjacent tissue-specific functional genes. Through GO annotation, we identified various genes (e.g., *MYBPHL*, *MYOG*, *SMYD1*, *VGLL2*) (Barefield et al., 2017; Hitachi et al., 2019; Luo et al., 2015; Xue et al., 2017) and biological processes associated with muscle development in or near muscle-specific OCRs. Thus, the generation of a tissue-specific atlas of OCRs enabled gene regulation exploration in chickens with previously unattained details. Further research will focus on collecting samples from different tissues or stages of embryonic development to explore dynamic chromatin landscapes (Gorkin et al., 2020). It is important to note that biological tissues consist of heterogeneous assemblies of cell types that can differentially impact complex phenotypes. Recent studies have shown substantial heterogeneity in chromatin accessibility among different cell types (Carter & Zhao, 2021). Cellular variations in chromatin accessibility likely arise from a combination of asynchronicity in the cell cycle stage and differences in TF expression and/or binding (Buenrostro et al., 2015). Therefore, this tissue-based chromatin accessibility atlas only represents the first stage of chicken epigenome annotation. Application of finer resolution, such as single-cell ATAC-seq, would allow for more in-depth investigation of spatial variation.

Using a major QTL study on chicken chromosome 1, we demonstrated the potential role of chromatin openness data in gene fine-mapping. Previous studies have deduced multiple causal loci with differential effects on chicken body weight, likely due to polymorphic segregation at multiple, tightly linked regulatory mutation loci in this QTL region (Wang et al., 2020c). Here, we focused on this interval with a higher density of SNPs (7.9 M) and performed GWAS of two complex traits. Although numerous significant loci were not protein-coding mutations and were in high LD with each other, we still identified several candidate functional loci located in the OCRs that may contribute to genetic effects on body weight in different tissues. We reconfirmed the importance of the *CAB39L* gene in DL and identified a new candidate regulatory locus. Additionally, we identified a new candidate gene, *RCBTB1*, whose mutation in the promoter region plays an important role in muscle and liver development (Guo et al., 2004). For bone tissue, the most significant locus was 30 kb from the nearest novel lncRNA gene (*ENSGALG00000053256*), associated with chicken body size (Wu et al., 2021). We speculate that this may be a potential enhancer regulator, although few functional studies of this lncRNA have been performed.

In conclusion, we generated a multi-tissue chromatin accessibility atlas of the chicken using ATAC-seq data, providing a compendium of resources for further studies on epigenetics and the evolution of birds and mammals. We also highlighted the potentially important role of analyzing OCRs to

facilitate identification of causative mutations. This new strategy for gene fine-mapping studies provides an improved understanding of the genetic architecture of complex traits.

DATA AVAILABILITY

All new sequence reads were deposited in the National Center for Biotechnology Information (NCBI) sequence read archive (SRA) under BioProjectID PRJNA847569, in the Genome Sequence Archive under Accession No. CRA008349, and in the Science Data Bank under DOI: 10.57760/sciencedb.02970. All data generated in this study are available within the article and its Supplementary Data files.

SUPPLEMENTARY DATA

Supplementary data to this article can be found online.

COMPETING INTERESTS

The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS

Y.Z.W. and X.X.H. conceived and designed the study and conducted the primary analyses. X.N.Z. collected samples and performed the experiments. C.L. carried out experimental verification. X.X.H., H.Y.W., and R.Z. helped analyze and interpret the data. Y.Z.W. wrote the initial manuscript. X.N.Z. and X.X.H. were responsible for statistical analyses and manuscript revision. All authors read and approved the final version of the manuscript.

ACKNOWLEDGMENTS

We thank Ding-Ming Shu, Hao Qu, and Cheng-Long Luo for initiating the intercross experiment with Xiao-Xiang Hu, Jian Ji, Xin-Chun Xu, and Jing-Yi He are acknowledged for their valuable contributions during sample collection.

REFERENCES

Barefield DY, Puckelwartz MJ, Kim EY, Wilsbacher LD, Vo AH, Waters EA, et al. 2017. Experimental modeling supports a role for MyBP-HL as a novel myofilament component in arrhythmia and dilated cardiomyopathy. *Circulation*, **136**(16): 1477−1491.

Bikle DD. 2020. The vitamin D receptor as tumor suppressor in skin. *Advances in Experimental Medicine and Biology*, **1268**: 285−306.

Boyle EA, Li YI, Pritchard JK. 2017. An expanded view of complex traits: from polygenic to omnigenic. *Cell*, **169**(7): 1177−1186.

Buenrostro JD, Wu BJ, Litzenburger UM, Ruff D, Gonzales ML, Snyder MP, et al. 2015. Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, **523**(7561): 486−490.

Carter B, Zhao KJ. 2021. The epigenetic basis of cellular heterogeneity. *Nature Reviews Genetics*, **22**(4): 235−250.

Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S, et al. 2017. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nature Methods*, **14**(10): 959−962.

Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, et al. 2019. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. *BMC Biology*, **17**(1): 108.

Giuffra E, Tuggle CK, FAANG Consortium. 2019. Functional annotation of animal genomes (FAANG): current achievements and roadmap. *Annual Review of Animal Biosciences*, **7**: 65−88.

Gorkin DU, Barozzi I, Zhao Y, Zhang YX, Huang H, Lee AY, et al. 2020. An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature*, **583**(7818): 744−751.

Groß C, Bortoluzzi C, de Ridder D, Megens HJ, Groenen MAM, Reinders M, et al. 2020. Prioritizing sequence variants in conserved non-coding elements in the chicken genome using chCADD. *PLoS Genetics*, **16**(9): e1009027.

Guo DF, Tardif V, Ghelima K, Chan JSD, Ingelfinger JR, Chen XM, et al. 2004. A novel angiotensin II type 1 receptor-associated protein induces cellular hypertrophy in rat vascular smooth muscle and renal proximal tubular cells. *Journal of Biological Chemistry*, **279**(20): 21109−21120.

Halstead MM, Kern C, Saelao P, Chanthavixay G, Wang Y, Delany ME, et al. 2020a. Systematic alteration of ATAC-seq for profiling open chromatin in cryopreserved nuclei preparations from livestock tissues. *Scientific Reports*, **10**(1): 5230.

Halstead MM, Kern C, Saelao P, Wang Y, Chanthavixay G, Medrano JF, et al. 2020b. A comparative analysis of chromatin accessibility in cattle, pig, and mouse tissues. *BMC Genomics*, **21**(1): 698.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. 2010. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Molecular Cell*, **38**(4): 576−589.

Hitachi K, Inagaki H, Kurahashi H, Okada H, Tsuchida K, Honda M. 2019. Deficiency of *Vgll2* gene alters the gene expression profiling of skeletal muscle subjected to mechanical overload. *Frontiers in Sports and Active Living*, **1**: 41.

Hutton SR, Pevny LH. 2011. *SOX2* expression levels distinguish between neural progenitor populations of the developing dorsal telencephalon. *Developmental Biology*, **352**(1): 40−47.

Jiang LD, Zheng ZL, Qi T, Kemper KE, Wray NR, Visscher PM, et al. 2019. A resource-efficient tool for mixed model association analysis of large-scale data. *Nature Genetics*, **51**(12): 1749−1755.

Kern C, Wang Y, Xu XQ, Pan ZY, Halstead M, Chanthavixay G, et al. 2021. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nature Communications*, **12**(1): 1821.

Lai YC, Liang YC, Jiang TX, Widelitz RB, Wu P, Chuong CM. 2018. Transcriptome analyses of reprogrammed feather / scale chimeric explants revealed co-expressed epithelial gene networks during organ specification. *BMC Genomics*, **19**(1): 780.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, **9**(4): 357−359.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**(16): 2078−2079.

Li M, Sun CJ, Xu NY, Bian PP, Tian XM, Wang XH, et al. 2022. De novo assembly of 20 chicken genomes reveals the undetectable phenomenon for thousands of core genes on microchromosomes and subtelomeric regions. *Molecular Biology and Evolution*, **39**(4): msac066.

Liu CY, Wang MY, Wei XY, Wu L, Xu JS, Dai X, et al. 2019. An ATAC-seq atlas of chromatin accessibility in mouse tissues. *Scientific Data*, **6**(1): 65.

Luo W, Li EX, Nie QH, Zhang XQ. 2015. Myomaker, regulated by MYOD, MYOG and miR-140-3p, promotes chicken myoblast fusion. *International Journal of Molecular Sciences*, **16**(11): 26186−26201.

Martin M. 2011. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet. Journal*, **17**(1): 10−12.

Ou JH, Liu HB, Yu J, Kelliher MA, Castilla LH, Lawson ND, et al. 2018. ATACseqQC: a Bioconductor package for post-alignment quality assessment of ATAC-seq data. *BMC Genomics*, **19**(1): 169.

Pan ZY, Yao YL, Yin HW, Cai ZX, Wang Y, Bai LJ, et al. 2021. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. *Nature Communications*, **12**(1): 5848.

Patoori S, Jean-Charles N, Gopal A, Sulaiman S, Gopal S, Wang B, et al. 2020. Cis-regulatory analysis of Onecut1 expression in fate-restricted retinal progenitor cells. *Neural Development*, **15**(1): 5.

Pérez C, Dastot-Le Moal F, Collot N, Legendre M, Abadie I, Bertrand AM, et al. 2012. Screening of *LHX2* in patients presenting growth retardation with posterior pituitary and ocular abnormalities. *European Journal of Endocrinology*, **167**(1): 85−91.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**(3): 559−575.

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, **26**(6): 841−842.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, **44**(W1): W160−W165.

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature*, **518**(7539): 317−330.

Rothstein M, Simoes-Costa M. 2020. Heterodimerization of TFAP2 pioneer factors drives epigenomic remodeling during neural crest specification. *Genome Research*, **30**(1): 35−48.

Sackton TB, Grayson P, Cloutier A, Hu ZR, Liu JS, Wheeler NE, et al. 2019. Convergent regulatory evolution and loss of flight in paleognathous birds. *Science*, **364**(6435): 74−78.

Schug J, Schuller WP, Kappen C, Salbaum JM, Bucan M, Stoeckert CJ Jr. 2005. Promoter features related to tissue specificity as measured by Shannon entropy. *Genome Biology*, **6**(4): R33.

Seki R, Li C, Fang Q, Hayashi S, Egawa S, Hu J, et al. 2017. Functional roles of Aves class-specific *cis*-regulatory elements on macroevolution of bird-specific features. *Nature Communications*, **8**: 14229.

Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. 2012. A map of the *cis*-regulatory sequences in the mouse genome. *Nature*, **488**(7409): 116−120.

The ENCODE Project Consortium. 2004. The ENCODE (ENCyclopedia of DNA Elements) project. *Science*, **306**(5696): 636−640.

The ENCODE Project Consortium, Snyder MP, Gingeras TR, Moore JE, Weng ZP, Gerstein MB, et al. 2020. Perspectives on ENCODE. *Nature*, **583**(7818): 693−698.

Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. 2015. Proteomics. Tissue-based map of the human proteome. *Science*, **347**(6220): 1260419.

Wang MS, Thakur M, Peng MS, Jiang Y, Frantz LAF, Li M, et al. 2020a. 863 genomes reveal the origin and domestication of chicken. *Cell Research*, **30**(8): 693−701.

Wang YZ, Bu LN, Cao XM, Qu H, Zhang CY, Ren JL, et al. 2020b. Genetic dissection of growth traits in a unique chicken advanced intercross line. *Frontiers in Genetics*, **11**: 894.

Wang YZ, Cao XM, Luo CL, Sheng ZY, Zhang CY, Bian C, et al. 2020c. Multiple ancestral haplotypes harboring regulatory mutations cumulatively contribute to a QTL affecting chicken growth traits. *Communications Biology*, **3**(1): 472.

Wu Z, Bortoluzzi C, Derks MFL, Liu LQ, Bosse M, Hiemstra SJ, et al. 2021. Heterogeneity of a dwarf phenotype in Dutch traditional chicken breeds revealed by genomic analyses. *Evolutionary Applications*, **14**(4): 1095−1108.

Xu M, Xie XL, Dong XH, Liang GQ, Gan L. 2018. Generation and characterization of *Lhx3*^{GFP} reporter knockin and *Lhx3*^{loxP} conditional knockout mice. *Genesis*, **56**(4): e23098.

Xue Q, Zhang GX, Li TT, Ling JJ, Zhang XQ, Wang JY. 2017. Transcriptomic profile of leg muscle during early growth in chicken. *PLoS One*, **12**(3): e0173824.

Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, **88**(1): 76−82.

Yang RF, Guo XL, Zhu D, Tan C, Bian C, Ren JL, et al. 2021. Accelerated deciphering of the genetic architecture of agricultural economic traits in pigs using a low-coverage whole-genome sequencing strategy. *GigaScience*, **10**(7): giab048.

Young JJ, Grayson P, Edwards SV, Tabin CJ. 2019. Attenuated Fgf signaling underlies the forelimb heterochrony in the Emu *Dromaius novaehollandia*e. *Current Biology*, **29**(21): 3681−3691.e5.

Yue F, Cheng Y, Breschi A, Vierstra J, Wu WS, Ryba T, et al. 2014. A comparative encyclopedia of DNA elements in the mouse genome. *Nature*, **515**(7527): 355−364.

Zhang GJ, Li C, Li QY, Li B, Larkin DM, Lee C, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, **346**(6215): 1311−1320.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, **9**(9): R137.

Zhao YX, Hou Y, Xu YY, Luan Y, Zhou HH, Qi XL, et al. 2021. A compendium and comparative epigenomics analysis of *cis*-regulatory elements in the pig genome. *Nature Communications*, **12**(1): 2217.