Journal of Integrative Agriculture 2022, 21(7): 2065–2075

Available online at www.sciencedirect.com

ScienceDirect

RESEARCH ARTICLE

Integration of genome-wide association study and selection signatures reveals genetic determinants for skeletal muscle production traits in an F_2 chicken population

LI Yu-dong[1, 2, 3], BAI Xue[1, 2, 3], LIU Xin[1, 2, 3], WANG Wei-jia[1, 2, 3], LI Zi-wei[1, 2, 3], WANG Ning[1, 2, 3], XIAO Fan[4], GAO Hai-he[4], GUO Huai-shun[4], LI Hui[1, 2, 3], WANG Shou-zhi[1, 2, 3]

[1] *Key Laboratory of Chicken Genetics and Breeding, Ministry of Agriculture and Rural Affairs, Harbin 150030, P.R.China*

[2] *Key Laboratory of Animal Genetics, Breeding and Reproduction, Education Department of Heilongjiang Province, Harbin 150030, P.R.China*

[3] *School of Animal Sciences and Technology, Northeast Agricultural University, Harbin 150030, P.R.China*

[4] *Fujian Sunnzer Biotechnology Development Co., Ltd., Guangze 354100, P.R.China*

Abstract

Improving the production of broiler chicken meat has been a goal of broiler breeding programs worldwide for many years. However, the genetic architectures of skeletal muscle production traits in chickens have not yet been fully elucidated. In the present study, a total of 519 F_2 birds, derived from a cross of Arbor Acres broiler and Baier layer, were re-sequenced (26 F_0 individuals were re-sequenced at a 10-fold depth; 519 F_2 individuals were re-sequenced at a 3-fold depth) and the coupling of genome-wide association study (GWAS) and selection signatures (F_{ST} (fixation index) and θ_π (nucleotide diversity)) was carried out to pinpoint the associated loci and genes that contribute to pectoral muscle weight (PMW) and thigh muscle weight (TMW). A total of 7 890 258 single nucleotide polymorphisms (SNPs) remained to be analyzed after quality control and imputation. The integration of GWAS and selection signature analyses revealed that genetic determinants responsible for skeletal muscle production traits were mainly localized on chromosomes 1 (168.95–172.43 Mb) and 4 (74.37–75.23 Mb). A total of 17 positional candidate genes (PCGs) (*LRCH1*, *CDADC1*, *CAB39L*, *LOC112531568*, *LOC112531569*, *FAM124A*, *FOXO1*, *NBEA*, *GPALPP1*, *RUBCNL*, *ARL11*, *KPNA3*, *LHFP*, *GBA3*, *LOC112532426*, *KCNIP4*, and *SLIT2*) were identified in these regions. In particular, *KPNA3* and *FOXO1* were the most promising candidates for meat production in chickens. These findings will help enhance our understanding of the genetic architecture of chicken muscle production traits, and the significant SNPs identified could be promising candidates for integration into practical breeding programs such as genome-wide selection (GS) to improve the meat yield of chickens.

Received 6 June, 2021 Accepted 20 July, 2021

LI Yu-dong, E-mail: 710827104@qq.com, liyudong886@163.com; Correspondence WANG Shou-zhi, Tel: +86-451-55191495, E-mail: shouzhiwang@neau.edu.cn

doi: 10.1016/S2095-3119(21)63805-4

Keywords: chicken, muscle growth and development, GWAS, selection signature

1. Introduction

With the improvement of living standards and the enhancement of healthy consumption consciousness, consumers prefer meat products with high protein, low fat and low cholesterol contents. Chicken products with rich nutrients and lower calories meet people's demands (Mir *et al*. 2017), which has led to a noticeable increase in poultry consumption among all animal products. Therefore, it has become the goal of producers and scientists to obtain commercial broilers with rapid growth and high feed efficiency (Pampouille *et al*. 2018). Over the past few decades, the body weight of commercial broilers has been dramatically increased, and the time to market has been shortened by half (Petracci and Cavani 2012). These developments have led to a significant increase in chicken meat production, especially breast muscle and drumstick meat yields, which represent the most valuable portion for most product operators and consumers (Baldi *et al*. 2019).

In the recent two decades, a number of the quantitative trait loci (QTL) affecting chicken meat production traits have been identified based on the candidate gene approach, marker-QTL linkage analysis, and SNP-chip based genome wide association study (GWAS). Le Mignon *et al*. (2009) confirmed one QTL at the distal end of *Gallus gallus* chromosomes (GGA) 5, which influences breast muscle (BM) weight in chickens, by performing multiple-trait and multi-QTL analyses of the whole available data set from two F_2 populations. Ankra-Badu *et al*. (2010) found several sex-specific and sex-antagonistic QTL related to breast meat and drumstick meat yield. Sato *et al*. (2012) confirmed that polymorphisms in the Insulin-like growth factor-1 (*IGF1*) promoter region were significantly associated with breast muscle weight (BMW) in a chicken F_2 population. Myostatin has been widely reported to be closely related to the growth and muscle development of chickens (Zhang *et al*. 2015; Liu L X *et al*. 2016; Dou *et al*. 2018). Chen *et al*. (2015) identified *FOXO3*, as a candidate gene affecting the growth of chest and leg muscles in chickens through RNA-sequencing. Godoy *et al*. (2015) found SNPs and indels associated with breast muscle deposition in a QTL region on chicken chromosome 2 by low-density genome-wide sequencing. Xie *et al*. (2012) identified a narrow region on chromosome 1 (173.5–175 Mb) that is strongly associated with chicken breast muscle weight (BMW) and leg muscle weight (LMW) using a 60K SNP Illumina iSelect chicken array based on GWAS in an F_2 chicken population.

Scientists have made significant progress in using GWAS to identify the genetic association between genotype and phenotype over the past decade. However, it is still a major challenge to fine-map the markers and genes responsible for the potential phenotypic variations of quantitative traits in livestock. Generally speaking, the genomic region screened by GWAS still includes many genes due to high linkage disequilibrium (LD) between markers and QTL. In this case, it is necessary to combine GWAS based on whole-genome resequencing with other strategies, such as selective signature and multi-omics methods, to improve the efficiency and accuracy of the gene mapping (Zhou *et al*. 2018).

Domesticated animal species have undergone intensive selection due to breeding and domestication, which has given rise to various phenotypes when compared with their wild counterparts (Liu Z *et al*. 2016). In the process of long-term artificial or natural selection, some traits of the animal population change in a specific direction, leaving obvious selection characteristics in the genome, which is the signature of selection (Grossman *et al*. 2010; Mariadassou *et al*. 2020). F_{ST} (fixation index) and θ_{π} (nucleotide diversity) are commonly used to reflect the genomic changes caused by selection from different angles, so as to reveal the domestication history in the process of animal breeding and better reveal the underlying genetic basis for the formation of economically important traits.

Pectoral muscle weight (PMW) and thigh muscle weight (TMW) are important commercial traits. However, the genetic architectures underlying these complex traits have yet to be uncovered. We hypothesized that some variants and genes would be associated with skeletal muscles and these traits would leave similarly selected footprints on the genome because of the long-term artificial selection of meat production traits in chickens. The objective of this study was to dissect the genetic basis of skeletal muscle growth and development by the integration of GWAS and genome-wide detection of selection signatures using an F_2 chicken resource population. The findings will be essential for obtaining an in-depth understanding of the genetic architecture of chicken skeletal muscle.

2. Materials and methods

2.1. Experimental populations and phenotypic measurements

A chicken F_2 population, the Northeast Agricultural University Resource Population, China (NEAURP), was used in the current study. The population was constructed by crossing broiler cocks derived from the fat line with high abdominal fat content (Leng *et al*. 2009) and Baier layer dams (a Chinese native breed). More details of this population have been described in previous reports (Liu *et al*. 2008; Zhang *et al*. 2010, 2011). A total of 519 F_2 individuals (263 male chickens and 256 female chickens) from 12 half-sib families were used in this study. In the process of feeding, all F_2 birds had free access to feed and water. These birds were kept under the same environmental conditions. The birds were raised in hatch and housed in pens. The temperature in the chicken coop was kept at 18 to 25°C, and the air humidity was maintained at 60 to 65%. The commercial diets provided were based on corn and soybeans and in line with all NRC (1994) requirements. From hatch to 3 weeks of age, the birds received a starter feed (metabolizable energy (ME), 3 000 kcal kg^{-1}; crude protein (CP), 210 g kg^{-1}), and from 4 weeks of age to slaughter the birds were fed a grower diet (ME, 3 100 kcal kg^{-1}; CP, 190 g kg^{-1}). All birds were euthanized by intramuscular injection of pentobarbital (Sigma, St. Louis, MO, USA) (0.04 g kg^{-1}) under deep anesthesia and exsanguination from the jugular vein at the age of 12 weeks. PMW and TMW were measured after slaughter at the age of 12 weeks. Only the samples that could successfully provide the chicken pectoralis and thigh muscles were included. During the experiment, fixed personnel were responsible for slaughtering, sample collection, weighing, preservation, and data recording.

2.2. Genotypes and quality control

Total genomic DNA was extracted from the blood of each sample using the reagent test kit. A single individual was used for genome sequencing on the Illumina HiSeq PE150 Platform (26 F_0 individuals were re-sequenced at 10-fold depth; 519 F_2 individuals were re-sequenced at 3-fold depth). Each individual was the experimental unit. Library construction and sample indexing were done according to the standard protocol of Illumina (Zhang *et al*. 2020). After alignment, we performed SNP calling on a population scale in the package SAMtools (Li *et al*. 2009). Then we calculated genotype likelihoods from reads for each individual at each genomic location and the allele frequencies in the sample. The 'mpileup' command was used to identify SNPs with the parameters of '–q 1 –C 50 –S –D –m 2 –F 0.002 –u'. Then, to exclude SNP calling errors caused by incorrect mapping, only high-quality SNPs (coverage depth ≥2, root mean square (RMS) mapping quality ≥20, miss ≤0.3) were kept for subsequent analysis (Huang *et al*. 2019). A total of 10 889 955 SNPs were left after filtering from 15 868 916 raw SNPs. The missing genotype was imputed using the F_0 generation 10-fold cross-validation in 519 sequencing individuals of the F_2 generation. Imputation was performed using BEAGLE 4.0 (Browning and Browning 2009) with default parameter settings. It was assumed that there was no relationship between each individual and that the genotypes were unphased. Imputation accuracy (*r*) was calculated per SNP by the correlation between the observed and imputed genotypes. A total of 7 890 258 SNPs were left after the imputed 10 889 955 SNPs were filtered by MAF ≥0.05 and miss ≤0.2 for the 519 individuals.

2.3. Single-marker GWAS

There were 519 samples in our association panel. A total of 7 890 258 SNPs were used in GWAS for meat production traits. Association analysis was conducted using the GEMMA (Genome-wide Efficient Mixed-model Association) Software Package (Zhou and Stephens 2012). For the MLM (mixed linear model) analysis, the equation is as follows:

$$y=\mathbf{S}\beta+X\alpha+\mathbf{K}\mu+e$$

In this equation, y represents phenotype; **S** is the incidence matrix of fixed effects and β is the vector of corresponding coefficients including the intercept. Gender, BW_0 (body weight at hatch), and the top 10 PCs (principal components) used for population structure correction were included as covariates to build up the **S** matrix. X represents the vector of SNP genotype and α is the corresponding effect of the marker; **K** is the incidence matrix for μ and μ is the vector of random additive genetic effects following the multinormal distribution $N(0, \mathbf{G}\sigma_\mu^2)$, in which **G** is the genomic relationship matrix based on IBS (identity by state), and σ_μ^2 is the polygenetic additive variance. e represents the random residual with a distribution of $N(0, \mathbf{I}\sigma_e^2)$ (**I** is a *n* by *n* identity matrix and *n* is the number of the individual). Based on the Bonferroni correction method (Duggal *et al*. 2008), the genome-wide significance threshold value was set as 0.05/N (*P*-value=6.34E–9), where N is the number of informative SNPs.

2.4. Estimation of genetic parameters

The genetic parameters (heritability, genetic and

phenotypic correlations) for PMW and TMW were estimated by the Average Information Restricted Maximum Likelihood method based on an animal model using the WOMBAT Software (Meyer 2007). The animal model used for the genetic parameter estimation is described as follows:

$$Y_{li}=\mu+S_l+BW_0+a_i+e_i$$

where Y_{li} is the vector of observations for the PMW and TMW; μ is the value of the population mean; S_l is the fixed effect of sex; BW_0 is taken as a covariate in the heritability estimation of PMW and TMW; a_i is the random direct additive genetic effect of individual i, and e_i is the random residual effects. A single trait model was used to estimate the heritability of PMW and TMW. Bivariate analyses were performed to compute phenotypic and genetic correlations between PMW and TMW.

2.5. Genome-wide selection signatures test

Fifteen samples with the highest and lowest phenotypic values of PMW and TMW were selected respectively from 519 individuals and divided into two groups. Selection signature analysis was carried out between the two groups. To identify genome-wide selection signatures associated with the adaptation of chicken meat production traits, we calculated the genome-wide distribution of F_{ST} and θ_π for the defined group pairs with VCFtools (40-kb windows sliding in 10-kb steps). The θ_π ratios were $\log_2$-transformed. Subsequently, we estimated and ranked the empirical percentiles of F_{ST} and $\log_2(\theta_\pi$ ratio) in each window. We considered the windows with the top 5% of F_{ST} and $\log_2(\theta_\pi$ ratio) values simultaneously as candidate outliers under strong selective sweep (Li *et al*. 2013). All outlier windows were assigned to corresponding SNPs and genes. In other steps, the analysis of the allele frequency difference (ΔAF) between the two groups was realized by R and Perl.

2.6. Functional annotation and enrichment analysis of the candidate genes

SNP annotation was performed according to the GCF_000002315.6_GRCg6a reference genome using the package ANNOVAR (Wang *et al*. 2010). Based on the genome annotation, SNPs were categorized in either exon regions (overlapping with a coding exon), intronic regions (overlapping with an intron), splicing sites (within 2 bp of a splicing junction), upstream and downstream regions (within a 1-kb region upstream or downstream), or intergenic regions. Only the high-quality SNPs were annotated. We identified candidate genes according to their physical location on the chromosomes and biological functions. According to the analysis results of linkage disequilibrium attenuation distance based on PopLDdecay Software, candidate genes were screened in the 40 kb region upstream and downstream of each top SNP.

Functional enrichment of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) were performed using OmicShare tools (www.omicshare.com/tools). The significant *P*-value (<0.05) was calculated as:

$$P=1-\sum_{i=0}^{m-1}\frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

where N is the number of genes with a GO/KEGG annotation in all genes; n is the number of predicted candidate genes in N; M is the number of genes annotated as a specific GO term/KEGG pathway in all genes; and m is the number of candidate genes annotated as a specific GO term/KEGG pathway. The corrected *P*-value by FDR is 0.05. The GO term/KEGG pathway satisfying this condition is defined as the one significantly enriched in the candidate target gene.

3. Results

3.1. Descriptive statistics and genetic parameter analysis

The descriptive statistics and genetic parameter analysis are shown in Tables 1 and 2, respectively. The heritabilities of PMW and TMW were 0.976 and 0.937, respectively. There were significant (P<0.01) and positive phenotypic correlations (0.831) and genetic correlations (0.972) between PMW and TMW.

Table 1 Number of animals (N), mean (M), standard deviation (SD), minimum (MIN), maximum (MAX), and coefficient of variation (CV) of meat production traits of F_2 chickens

Trait[1)]	N	M	SD	MIN	MAX	CV (%)
PMW	519	236.99	54.18	118.00	405.00	22.86
TMW	519	349.01	80.69	184.00	606.00	23.12

[1)] PMW, pectoral muscle weight; TMW, thigh muscle weight.

Table 2 Phenotypic and genetic correlations of pectoral muscle weight (PMW) and thigh muscle weight (TMW)

	PMW	TMW
PMW	0.976±0.014[1)]	0.831±0.014[**2)]
TMW	0.972±0.007[**3)]	0.937±0.016

[1)] Heritability of traits (on diagonal).
[2)] Phenotypic correlation coefficients with standard error (above diagonal).
[3)] Genetic correlation coefficients with standard errors (below diagonal).
**, significant correlations of PMW and TMW (P<0.01).

3.2. Genome-wide association of meat production traits

The results show that 247 SNPs were associated with PMW and 98 SNPs were associated with TMW at the Bonferroni corrected genome-wide significance level. Most of the detected SNPs were clustered closely on chromosomes 1, 4, and 27 (Fig. 1, left). Function annotation was performed in the 40 kb region (genome distance spans 40 kb when the r^2 drops to 0.1) upstream and downstream of each top SNP according to the analysis results of the linkage disequilibrium attenuation distance (Fig. 2). There were 70 significant SNPs on chromosome 1 overlapping in the association study of the two traits, including the top SNP (SNP with the most significant association with phenotype) located in the 171 411 019 bp of the intronic region of *SERPINE3* on chromosome 1. A large number of loci explained relatively little genetic variation individually. The top SNP explained 11.4 and 9.66% of the phenotypic variance of PMW and TMW, respectively (data not shown), demonstrating higher genetic contributions than other SNPs. A Q–Q plot was generated to estimate the difference between observed and expected chi-square statistic values of quantitative traits (Fig. 1, right), and it indicated that the potential candidate loci related to the traits were not caused by population stratification and the statistical model was reasonable.

3.3. Analysis of the selected signature regions

The selection signature regions of 36 chromosomes in F_2 population during natural selection were identified by the combination of F_{ST} and θ_π ratios statistics. Each was divided into two distinct selection regions (high and low phenotypic value) at a 5% empirical distribution. The selected regions were usually accompanied by a decrease in population polymorphisms and an increase in the genetic differentiation rates among subpopulations. The distribution of the selection regions shared by F_{ST}

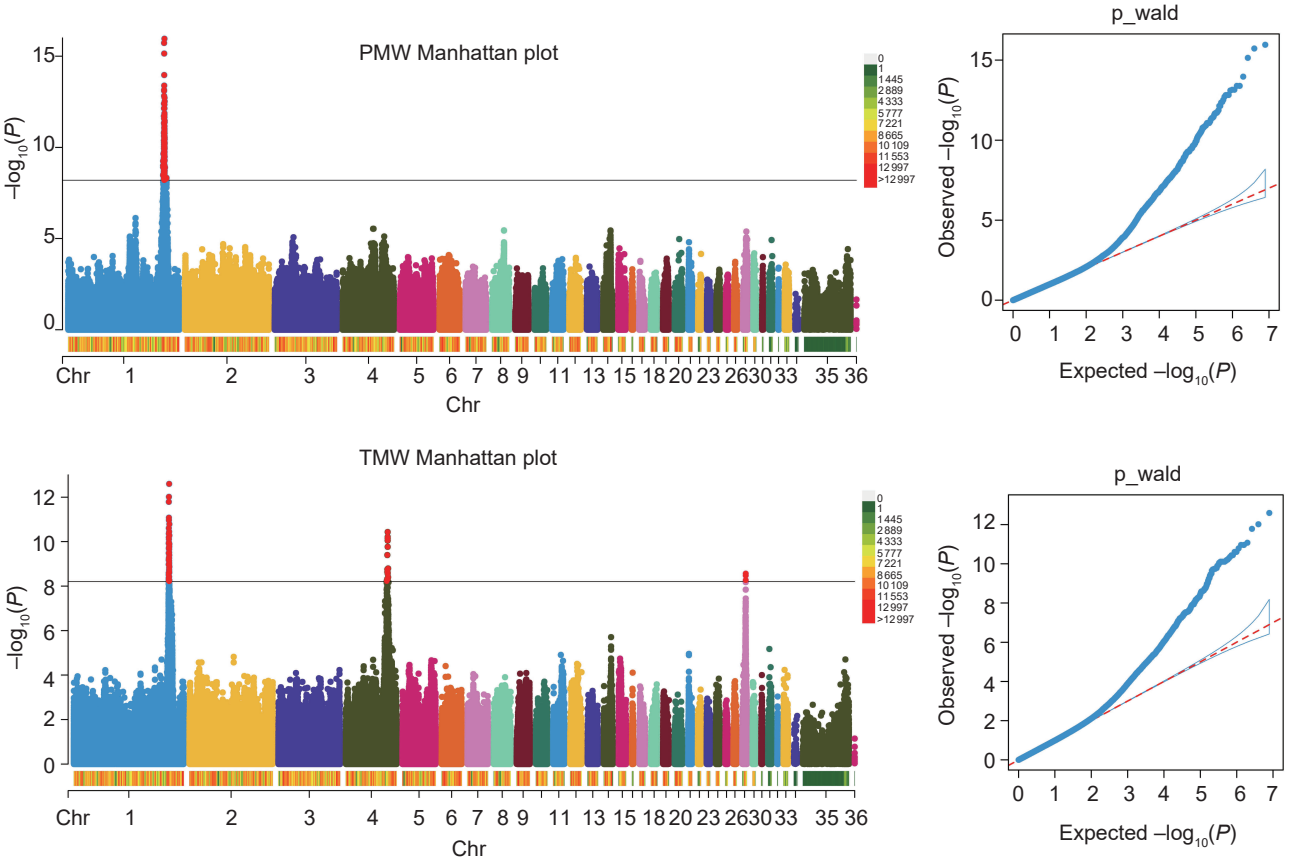

Fig. 1 Manhattan plot with marker density information and quantile–quantile (Q–Q) plot for the association analyses of pectoral muscle weight (PMW) and thigh muscle weight (TMW). In the Manhattan plots (left), $-\log_{10}(P\text{-value})$ of the filtered high-quality single nucleotide polymorphisms (SNPs) (*y*-axis) are plotted against their genomic positions (*x*-axis); SNPs on different chromosomes (1 to 36) are denoted by different colors. Marker density is shown at the bottom of the Manhattan plots. The horizontal black lines represent significant genome-wide association thresholds. Q–Q plots (right) are displayed as scatter plots of observed and expected log *P*-values.

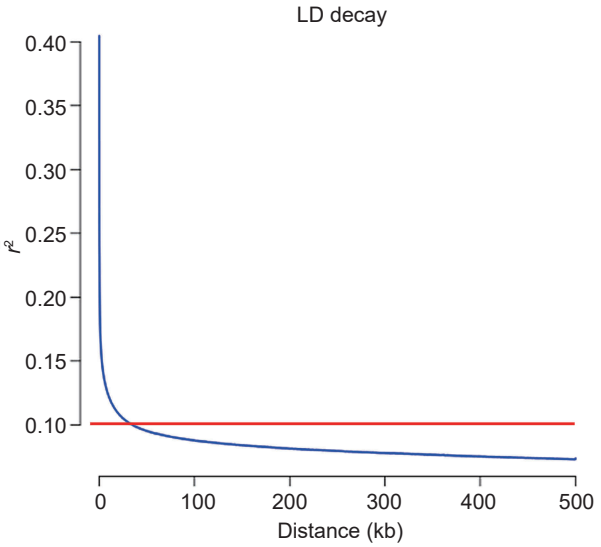

Fig. 2 Linkage disequilibrium (r^2) plot of the chicken genome. The horizontal red line indicates the critical value of r^2. The LD decay distance is 40 kb.

and $\log_2(\theta_\pi$ ratio) on the chromosome is shown in Fig. 3. Among the candidate regions screened by the F_{ST} and θ_π method, more regions of high meat yield traits (green data points) were scanned than low meat yield traits (blue data points), suggesting that higher meat yield tends to be a positive selection based on artificial action. There were 448 candidate genes in the selected region of HPMW ($\log_2(\theta_\pi$ ratio) (θ_π LPMW/θ_π HPMW)≥0.38, F_{ST}≥0.09) and 317 candidate genes in the selected regions of LPMW ($\log_2(\theta_\pi$ ratio) (θ_π LPMW/θ_π HPMW)≤–0.34, F_{ST}≥0.09). There were 432 candidate genes in the selected regions of HTMW ($\log_2(\theta_\pi$ ratio) (θ_π LTMW/θ_π HTMW)≥0.33, F_{ST}≥0.08) and 359 candidate genes in the selected regions of LTMW ($\log_2(\theta_\pi$ ratio) (θ_π LTMW/θ_π HTMW)≤–0.33, F_{ST}≥0.08).

3.4. Identification of candidate genes and functional enrichment analysis

We checked the distribution of selection signature regions and their overlap with the GWAS results within the genomic windows to reveal the selected genes of chickens in the process of domestication. A total of 17 underlying candidate genes were identified, including three uncharacterized genes (Table 3). It is worth noting that 6 genes, *FOXO1*, *CAB39L*, *FAM124A*, *LOC112531568*, *LOC112531569*, and *LOC112532426*, were uniformly mapped in low meat-producing regions (LPMW and LTMW). All underlying candidate genes were analyzed for enrichment. These genes were mainly enriched in the GO term carbohydrate derivative catabolic process (Fig. 4-A). Additionally, they were also involved in 14 biological process categories, 6 molecular functional categories, and 7 cellular components (Appendix A). KEGG analysis suggested that the candidate genes

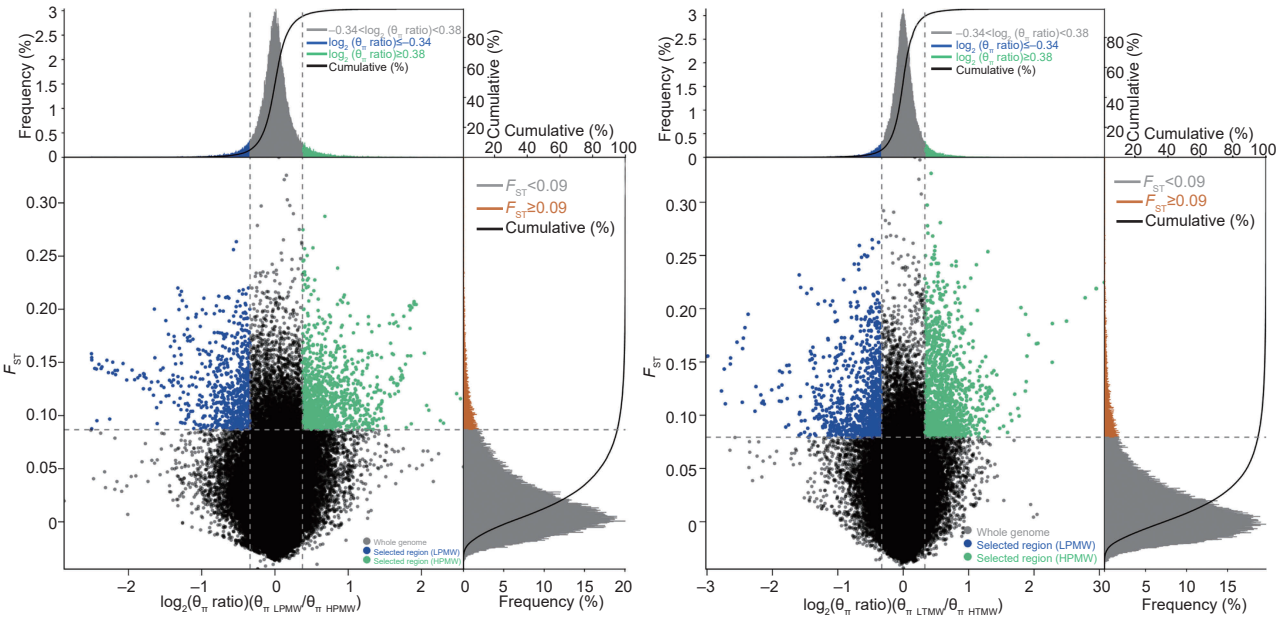

Fig. 3 Intersection of the two methods, fixation index and nucleotide diversity (F_{ST} and π) used to identify high-quality selection regions. Data points (blue and green) are located on both sides of the left and right vertical dashed lines. LTMW, low TMW selection region; HTMW, high TMW selection region. Different colors represent different intersection regions. The frequencies of the two methods' values are distributed in the right and top of the X and Y lines.

Table 3 Overview of the significant single nucleotide polymorphisms (SNPs) of genome-wide association study (GWAS) and selective signatures associated with pectoral muscle weight (PMW) and thigh muscle weight (TMW)

Trait	The position of lead SNP in GWAS	*P*-value	Candidate gene	Gene position (bp)	Selection direction of traits[1)]
PMW	Intergenic region between *RUBCNL* (dist=40 651) and *LRCH1* (dist=25 848)	2.25E–10	*LRCH1*	Chr1: 169 520 469–169 644 616	LPMW
PMW	Exonic	2.06E–10	*CDADC1*	Chr1: 170 447 991–170 463 866	LPMW
PMW	Intronic	1.92E–12	*CAB39L*	Chr1: 170 465 092–170 526 727	LPMW
PMW	Exonic	5.03E–10	*LOC112531568*	Chr1: 171 008 615–171 039 942	LPMW
PMW	Intergenic region between *LOC112531569* (dist=17 093) and *DLEU7* (dist=25 823)	4.50E–11	*LOC112531569*	Chr1: 171 083 270–171 100 945	LPMW
PMW	Intronic	4.52E–13	*FAM124A*	Chr1: 171 336 721–171 377 902	LPMW
PMW	Intronic	3.04E–13	*FOXO1*	Chr1: 171 900 263–171 963 540	LPMW
PMW	Intronic	5.28E–09	*NBEA*	Chr1: 174 373 491–174 843 850	LPMW
PMW	Intergenic region between *GPALPP1* (dist=14 156) and *GTF2F2* (dist=8 915)	5.64E–10	*GPALPP1*	Chr1: 168 950 859–168 965 646	HPMW
PMW	Intergenic region between *RUBCNL* (dist=40 647) and *LRCH1* (dist=25 852)	2.25E–10	*RUBCNL*	Chr1: 169 433 550–169 453 972	HPMW
PMW	3′ UTR	6.88E–10	*ARL11*	Chr1: 170 586 604–170 597 668	HPMW
PMW	Intronic	4.02E–13	*KPNA3*	Chr1: 170 597 160–170 650 244	HPMW
PMW	Intronic	1.28E–09	*LHFP*	Chr1: 172 287 762–172 427 229	HPMW
TMW	Intronic	2.57E–09	*CAB39L*	Chr1: 170 465 092–170 526 727	LTMW
TMW	Exonic	9.88E–10	*LOC112531568*	Chr1: 171 008 615–171 039 942	LTMW
TMW	Intergenic region between *LOC112531569* (dist=17 093) and *DLEU7* (dist=25 823)	4.95E–09	*LOC112531569*	Chr1: 171 083 270–171 100 945	LTMW
TMW	Intronic	3.59E–11	*FAM124A*	Chr1: 171 336 721–171 377 902	LTMW
TMW	Intronic	1.65E–11	*FOXO1*	Chr1: 171 900 263–171 963 540	LTMW
TMW	Intronic	4.83E–09	*GBA3*	Chr4: 74 366 408–74 429 486	LTMW
TMW	Intergenic region between *ADGRA3L* (dist=41 418) and *LOC112532426* (dist=17 941)	4.01E–10	*LOC112532426*	Chr4: 74 546 421–74 555 772	LTMW
TMW	Upstream	2.29E–09	*KCNIP4*	Chr4: 74 568 444–74 948 433	HTMW
TMW	Intronic	3.69E–11	*SLIT2*	Chr4: 74 981 753–75 225 786	HTMW

[1)] LPMW, low PMW selection region; HPMW, high PMW selection region; HTMW, high TMW selection region; LTMW, low TMW selection region.

were involved in the insulin signaling pathway and FOXO signaling pathway (Fig. 4-B).

4. Discussion

We performed a feasible combination strategy that integrates single-marker GWAS methodology and selection signature analysis to explore the genetic architectures of skeletal muscle traits in an F_2 chicken population. The study revealed that the genetic determinants responsible for skeletal muscle production traits were mainly localized on chromosomes 1 (168.95–172.43 Mb) and 4 (74.37–75.23 Mb). A total of 17 candidate genes were identified and six of them were uniformly mapped in low meat-producing regions (LPMW and LTMW). Thus, these results have confirmed our hypothesis that some variants and genes would be associated with skeletal muscles and these traits would leave similarly selected footprints on the genome because of long-term artificial selection of meat production traits in chickens.

The long-term artificial selection of traits with high premiums, such as chicken breast and drumstick meat yields, caters to the needs of consumers around the world for meat products. To further mine the key regulatory genes of phenotypic differences in chickens, GWAS based on linkage disequilibrium between SNPs markers and causal loci is considered to be an effective method to identify genetic links between phenotypes and genotypes (Hirschhorn and Daly 2005; McCarthy *et al*. 2008; Zhang *et al*. 2019). However, precisely distinguishing the potential causal variations related to the traits of interest from nearby neutral loci is one of the daunting challenges in genetic research. Due to the genetic hitch-hiking effect, some variants coexist with the selected loci on the LD block. These variants may produce signals similar to the actually selected loci that stay in LD. Given the above description, we adopted a feasible combination strategy that integrates single-marker GWAS methodology and selection signature analysis to identify those significant

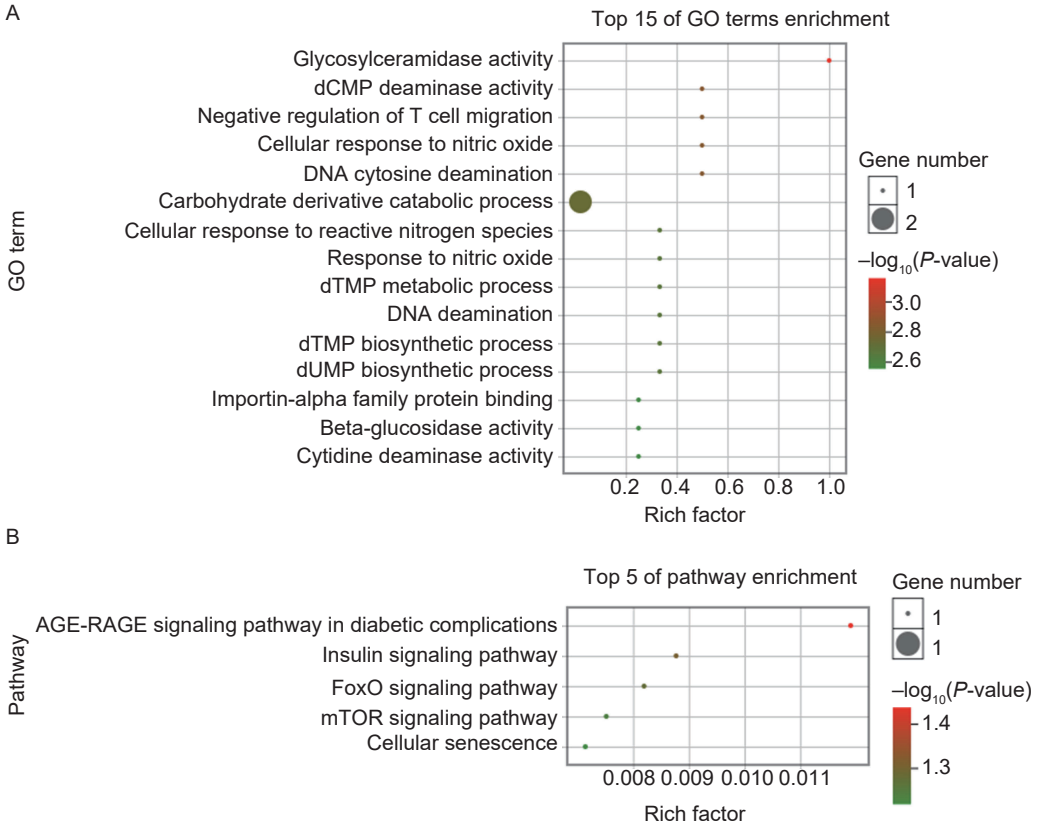

Fig. 4 GO and KEGG enrichment analysis. A, the top 15 enriched GO terms for meat production traits. B, the top 5 pathways enriched for meat production traits. Rich factor refers to the ratio of the number of genes with the indicated term entry with respect to the total number of genes in that term. The larger the Rich factor, the higher the degree of enrichment. The size of the bubble indicates the number of genes, and the color of the bubble indicates the level of significance.

SNPs associated with meat production traits at the genome-wide level and to explore the regions of the genome under selection and the candidate genes that exist.

In this study, we found that most of the significant SNPs explained the small phenotypic variance, which is consistent with the fact that complex traits in chickens are controlled by multiple genes (Andersson and Georges 2004). The GWAS top SNP in the current study, located in the 171 411 019 bp of the intronic region of *SERPINE3* on chromosome 1, was the signal with the strongest association for two meat production traits at the genome-wide level. There were significant (P<0.01) phenotypic and genetic correlation coefficients between PMW and TMW (Table 2). It is reasonable to speculate that the same gene or variation has multiple effects on PMW and TMW, known as pleiotropy, which is pervasive in chickens (Wright *et al*. 2010). The significant SNPs identified in this study are promising candidates for application to the whole genome selection array to improve the meat yield of chickens.

There were 17 candidate genes co-mapped by GWAS and selection signature analysis. Some of these genes were enriched in a GO term, carbohydrate derivative catabolic process, and KEGG pathways, insulin signaling, and FOXO signaling pathways (Fig. 4). The process and pathways are considered to be related to meat production traits. Kong *et al*. (2017) showed that the cellular and physiological functions of differentially expressed genes such as carbohydrate metabolism are tightly associated with rapid growth and differential muscle fiber contents in modern broiler lines by an RNA sequencing (RNA-seq) method. Insulin and FOXO signaling pathways play a crucial role in both the metabolism of carbohydrates and lipid (Lee and Dong 2017) and the development and growth of muscle (O'Neill *et al*. 2016). In addition, some of the identified genes have been reported to be related to muscle development or muscle cell differentiation in previous literature. *KPNA3* had significant effects on some growth traits such as chest muscle weight and leg muscle weight in chickens (Xie *et al*. 2012; Abdalhag *et al*. 2015). The copy number variation in *LHFP*

(*LHFPL6*) plays an essential role in the daily weight gain of beef cattle (Xu *et al*. 2019). Data obtained in mice and humans suggested that variation of *NBEA* abundance or activity critically affects body weight, presumably by influencing the activity of feeding-related neural circuits (Olszewski *et al*. 2012). *SLIT2* is located in the QTL of chicken chromosome 4, affecting growth and muscle quality (Pertille *et al*. 2015; Lyu *et al*. 2017). An important candidate gene *FOXO1* was also identified in our study. *FOXO1* has been widely confirmed to be closely related to skeletal muscle growth and myogenic metabolism in animals, such as chicken (Jia *et al*. 2017), pig (Liu *et al*. 2017), and mice (Kamei *et al*. 2004). Notably, in addition to known genes, an array of uncharacterized genes, *LOC112531568*, *LOC112531569*, and *LOC112532426*, might be involved in muscle growth and contribute to the decrease of chicken meat production. The candidate genes identified in our study are worth investigating with respect to their functions and mechanisms in chicken muscle growth and development in the future.

The integration of GWAS and selection signature analyses revealed that the QTLs responsible for skeletal muscle production traits are mainly localized on chromosomes 1 (168.95–172.43 Mb) and 4 (74.37–75.23 Mb). Some of the genomic regions we identified in this study have previously been annotated in two different chicken populations. One was detected in an F_2 population derived from a reciprocal cross between White Recessive Rock and Xinghua chickens (Xie *et al*. 2012), and the other was mapped in an advanced intercross line generated from an F_2 inbred New Hampshire and White Leghorn (WL77) lines (Lyu *et al*. 2017). In total, five detected candidate genes overlapped with six published QTL mapped (Appendix B) for meat production traits in the Gallus_gallus-5.0 assembly QTL database (Hu *et al*. 2019). Interestingly, we checked the gene expression of chicken tissues (from 259 samples) at different developmental stages in the Animal Omics Database, and found that *KPNA3* was highly expressed in breast muscle and leg muscle (Appendix C). An RNA-seq transcriptomic study from the rat also strongly supports our results. *KPNA3* is widely expressed in different periods of rat muscle development and shows a biased expression in muscle tissue (RPKM 393.2) (Yu *et al*. 2014). Also, the mechanism of *KPNA3* in the growth and development of chicken skeletal muscle needs to be investigated in the future.

5. Conclusion

In this study, we integrated GWAS data and selection signature regions to refine the list of positional candidate genes related to meat production traits in a chicken F_2 population. The results showed that an array of genes, including *LRCH1*, *CDADC1*, *CAB39L*, *LOC112531568*, *LOC112531569*, *FAM124A*, *FOXO1*, *NBEA*, *GPALPP1*, *RUBCNL*, *ARL11*, *KPNA3*, *LHFP*, *GBA3*, *LOC112532426*, *KCNIP4*, and *SLIT2*, are candidates for PMW and TMW. In particular, the available evidence suggested *KPNA3* and *FOXO1* are the most promising causative candidates for skeletal muscle growth and development.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (31572394), the China Agriculture Research System of MOF and MARA (CARS-41), and the White Feather Broiler Breeding Joint Project of the Ministry of Agriculture and Rural Affairs of China (19190526).

Declaration of competing interest

The authors declare that they have no conflict of interest.

Ethical approval

This study was approved by the Laboratory Animal Management Committee of Northeast Agricultural University. All studies involving animals were carried out following the guidelines for the Care and Use of Experimental Animals established by the Ministry of Science and Technology of the People's Republic of China (approval number 2006–398).

Appendices associated with this paper are available on http://www.ChinaAgriSci.com/V2/En/appendix.htm

References

Abdalhag M A, Zhang T, Fan Q C, Zhang X Q, Zhang G X, Wang J Y, Wei Y, Wang Y J. 2015. Single nucleotide polymorphisms associated with growth traits in Jinghai yellow chickens. *Genetics and Molecular Research* (GMR), **14**, 16169–16177.

Andersson L, Georges M. 2004. Domestic-animal genomics: Deciphering the genetics of complex traits. *Nature Reviews Genetics*, **5**, 202–212.

Ankra-Badu G A, Shriner D, Le Bihan-Duval E, Mignon-Grasteau S, Pitel F, Beaumont C, Duclos M J, Simon J, Porter T E, Vignal A, Cogburn L A, Allison D B, Yi N, Aggrey S E. 2010. Mapping main, epistatic and sex-specific QTL for body composition in a chicken population divergently selected for low or high growth rate. *BMC Genomics*, **11**,

107.

Baldi G, Soglia F, Laghi L, Tappi S, Rocculi P, Tavaniello S, Prioriello D, Mucci R, Maiorano G, Petracci M. 2019. Comparison of quality traits among breast meat affected by current muscle abnormalities. *Food Research International*, **115**, 369–376.

Browning B L, Browning S R. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *American Journal of Human Genetics*, **84**, 210–223.

Chen B, Xu J, He X, Xu H, Li G, Du H, Nie Q, Zhang X. 2015. A genome-wide mRNA screen and functional analysis reveal *FOXO3* as a candidate gene for chicken growth. *PLoS ONE*, **10**, e0137087.

Dou T, Li Z, Wang K, Liu L, Rong H, Xu Z, Huang Y, Gu D, Chen X, Hu W, Zhang J, Zhao S, Jois M, Li Q, Ge C, Te Pas M F W, Jia J. 2018. Regulation of myostatin expression is associated with growth and muscle development in commercial broiler and DMC muscle. *Molecular Biology Reports*, **45**, 511–522.

Duggal P, Gillanders E M, Holmes T N, Bailey-Wilson J E. 2008. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics*, **9**, 516.

Godoy T F, Moreira G C, Boschiero C, Gheyas A A, Gasparin G, Paduan M, Andrade S C, Montenegro H, Burt D W, Ledur M C, Coutinho L L. 2015. SNP and INDEL detection in a QTL region on chicken chromosome 2 associated with muscle deposition. *Animal Genetics*, **46**, 158–163.

Grossman S R, Shlyakhter I, Karlsson E K, Byrne E H, Morales S, Frieden G, Hostetter E, Angelino E, Garber M, Zuk O, Lander E S, Schaffner S F, Sabeti P C. 2010. A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science*, **327**, 883–886.

Hirschhorn J N, Daly M J. 2005. Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, **6**, 95–108.

Hu Z L, Park C A, Reecy J M. 2019. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Research*, **47**, D701–D710.

Huang Y, Xiao L, Zhang Z, Zhang R, Wang Z, Huang C, Huang R, Luan Y, Fan T, Wang J, Shen C, Zhang S, Wang X, Randall J, Zheng B, Wu J, Zhang Q, Xia G, Xu C, Chen M, *et al*. 2019. The genomes of pecan and Chinese hickory provide insights into Carya evolution and nut nutrition. *Gigascience*, **8**, giz036.

Jia X, Ouyang H, Abdalla B A, Xu H, Nie Q, Zhang X. 2017. MiR-16 controls myoblast proliferation and apoptosis through directly suppressing *BCL2* and *FOXO1* activities. *Biochimica et Biophysica Acta Gene Regulatory Mechanisms*, **1860**, 674–684.

Kamei Y, Miura S, Suzuki M, Kai Y, Mizukami J, Taniguchi T, Mochida K, Hata T, Matsuda J, Aburatani H, Nishino I, Ezaki O. 2004. Skeletal muscle *FOXO1* (*FKHR*) transgenic mice have less skeletal muscle mass, down-regulated Type I (slow twitch/red muscle) fiber genes, and impaired glycemic control. *The Journal of Biological Chemistry*, **279**, 41114–41123.

Kong B W, Hudson N, Seo D, Lee S, Khatri B, Lassiter K, Cook D, Piekarski A, Dridi S, Anthony N, Bottje W. 2017. RNA sequencing for global gene expression associated with muscle growth in a single male modern broiler line compared to a foundational Barred Plymouth Rock chicken line. *BMC Genomics*, **18**, 82.

Lee S, Dong H H. 2017. FoxO integration of insulin signaling with glucose and lipid metabolism. *The Journal of Endocrinology*, **233**, R67–R79.

Leng L, Wang S, Li Z, Wang Q, Li H. 2009. A polymorphism in the 3′-flanking region of insulin-like growth factor binding protein 2 gene associated with abdominal fat in chickens. *Poultry Science*, **88**, 938–942.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.

Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, Wang T, Yeung C K, Chen L, Ma J, Zhang J, Jiang A, Li J, Zhou C, Zhang J, Liu Y, Sun X, Zhao H, Niu Z, Lou P, *et al*. 2013. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nature Genetics*, **45**, 1431–1438.

Liu L X, Dou T F, Li Q H, Rong H, Tong H Q, Xu Z Q, Huang Y, Gu D H, Chen X B, Ge C R, Jia J J. 2016. Myostatin mRNA expression and its association with body weight and carcass traits in Yunnan Wuding chicken. *Genetics and Molecular Research*, **15**, doi: 10.4238/gmr15048967.

Liu X, Zhang H, Li H, Li N, Zhang Y, Zhang Q, Wang S, Wang Q, Wang H. 2008. Fine-mapping quantitative trait loci for body weight and abdominal fat traits: Effects of marker density and sample size. *Poultry Science*, **87**, 1314–1319.

Liu Y, Yang X, Jing X, He X, Wang L, Liu Y, Liu D. 2017. Transcriptomics analysis on excellent meat quality traits of skeletal muscles of the chinese indigenous Min Pig compared with the Large White breed. *International Journal of Molecular Sciences*, **19**, 21.

Liu Z, Sun C J, Qu L, Wang K H, Yang N. 2016. Genome-wide detection of selective signatures in chicken through high density SNPs. *PLoS ONE*, **11**, e0166146.

Lyu S, Arends D, Nassar M K, Brockmann G A. 2017. Fine mapping of a distal chromosome 4 QTL affecting growth and muscle mass in a chicken advanced intercross line. *Animal Genetics*, **48**, 295–302.

Mariadassou M, Ramayo-Caldas Y, Charles M, Femenia M, Renand G, Rocha D. 2020. Detection of selection signatures in Limousin cattle using whole-genome resequencing. *Animal Genetics*, **51**, 815–819.

McCarthy M I, Abecasis G R, Cardon L R, Goldstein D B, Little J, Ioannidis J P, Hirschhorn J N. 2008. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. *Nature Reviews Genetics*, **9**, 356–369.

Meyer K. 2007. WOMBAT: A tool for mixed model analyses

in quantitative genetics by restricted maximum likelihood (REML). *Journal of Zhejiang University Science* (B), **8**, 815–821.

Le Mignon G, Pitel F, Gilbert H, Le Bihan-Duval E, Vignoles F, Demeure O, Lagarrigue S, Simon J, Cogburn L A, Aggrey S E, Douaire M, Le Roy P. 2009. A comprehensive analysis of QTL for abdominal fat and breast muscle weights on chicken chromosome 5 using a multivariate approach. *Animal Genetics*, **40**, 157–164.

Mir N A, Rafiq A, Kumar F, Singh V, Shukla V. 2017. Determinants of broiler chicken meat quality and factors affecting them: A review. *Journal of Food Science and Technology*, **54**, 2997–3009.

NRC (National Research Council). 1994. *Nutrient Requirements of Poultry*. 9th ed. National Academies Press, Washington, D.C.

Olszewski P K, Rozman J, Jacobsson J A, Rathkolb B, Strömberg S, Hans W, Klockars A, Alsiö J, Risérus U, Becker L, Hölter S M, Elvert R, Ehrhardt N, Gailus-Durner V, Fuchs H, Fredriksson R, Wolf E, Klopstock T, Wurst W, Levine A S, *et al*. 2012. Neurobeachin, a regulator of synaptic protein targeting, is associated with body fat mass and feeding behavior in mice and body-mass index in humans. *PLoS Genetics*, **8**, e1002568.

O'Neill B T, Lee K Y, Klaus K, Softic S, Krumpoch M T, Fentz J, Stanford K I, Robinson M M, Cai W, Kleinridders A, Pereira R O, Hirshman M F, Abel E D, Accili D, Goodyear L J, Nair K S, Kahn C R. 2016. Insulin and *IGF-1* receptors regulate FoxO-mediated signaling in muscle proteostasis. *The Journal of Clinical Investigation*, **126**, 3433–3446.

Pampouille E, Berri C, Boitard S, Hennequet-Antier C, Beauclercq S A, Godet E, Praud C, Jégo Y, Le Bihan-Duval E. 2018. Mapping QTL for white striping in relation to breast muscle yield and meat quality traits in broiler chickens. *BMC Genomics*, **19**, 202.

Pertille F, Zanella R, Felicio A M, Ledur M C, Peixoto J O, Coutinho L L. 2015. Identification of polymorphisms associated with production traits on chicken (*Gallus gallus*) chromosome 4. *Genetics and Molecular Research*, **14**, 10717–10728.

Petracci M, Cavani C. 2012. Muscle growth and poultry meat quality issues. *Nutrients*, **4**, 1–12.

Sato S, Ohtake T, Uemoto Y, Okumura Y, Kobayashi E. 2012. Polymorphism of insulin-like growth factor 1 gene is associated with breast muscle yields in chickens. *Animal Science Journal*, **83**, 1–6.

Wang K, Li M Y, Hakonarson H. 2010. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, **38**, e164.

Wright D, Rubin C J, Martinez Barrio A, Schütz K, Kerje S, Brändström H, Kindmark A, Jensen P, Andersson L. 2010. The genetic architecture of domestication in the chicken: Effects of pleiotropy and linkage. *Molecular Ecology*, **19**, 5140–5156.

Xie L, Luo C, Zhang C, Zhang R, Tang J, Nie Q, Ma L, Hu X, Li N, Da Y, Zhang X. 2012. Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. *PLoS ONE*, **7**, e30910.

Xu L, Yang L, Wang L, Zhu B, Chen Y, Gao H, Gao X, Zhang L, Liu G E, Li J. 2019. Probe-based association analysis identifies several deletions associated with average daily gain in beef cattle. *BMC Genomics*, **20**, 31.

Yu Y, Fuscoe J C, Zhao C, Guo C, Jia M, Qing T, Bannon D I, Lancashire L, Bao W, Du T, Luo H, Su Z, Jones W D, Moland C L, Branham W S, Qian F, Ning B, Li Y, Hong H, Guo L, *et al*. 2014. A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nature Communications*, **5**, 3230.

Zhang G X, Zhang T, Wei Y, Ding F X, Zhang L, Wang J Y. 2015. Functional identification of an exon 1 substitution in the myostatin gene and its expression in breast and leg muscle of the Bian chicken. *British Poultry Science*, **56**, 639–644.

Zhang H, Liang Q, Wang N, Wang Q, Leng L, Mao J, Wang Y, Wang S, Zhang J, Liang H, Zhou X, Li Y, Cao Z, Luan P, Wang Z, Yuan H, Wang Z, Zhou X, Lamont S J, Da Y, *et al*. 2020. Microevolutionary dynamics of chicken genomes under divergent selection for adiposity. *Science*, **23**, 101193.

Zhang H, Liu S H, Zhang Q, Zhang Y D, Wang S Z, Wang Q G, Wang Y X, Tang Z Q, Li H. 2011. Fine-mapping of quantitative trait loci for body weight and bone traits and positional cloning of the *RB1* gene in chicken. *Journal of Animal Breeding and Genetics*, **128**, 366–375.

Zhang H, Zhang Y D, Wang S Z, Liu X F, Zhang Q, Tang Z Q, Li H. 2010. Detection and fine mapping of quantitative trait loci for bone traits on chicken chromosome one. *Journal of Animal Breeding and Genetics*, **127**, 462–468.

Zhang Z, Chen Z, Ye S, He Y, Huang S, Yuan X, Chen Z, Zhang H, Li J. 2019. Genome-wide association study for reproductive traits in a duroc pig population. *Animals (Basel)*, **9**, 732.

Zhou X, Stephens M. 2012. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, **44**, 821–824.

Zhou Z, Li M, Cheng H, Fan W, Yuan Z, Gao Q, Xu Y, Guo Z, Zhang Y, Hu J, Liu H, Liu D, Chen W, Zheng Z, Jiang Y, Wen Z, Liu Y, Chen H, *et al*. 2018. An intercross population study reveals genes associated with body size and plumage color in ducks. *Nature Communications*, **9**, 2648.

Executive Editor-in-Chief LUO Xu-gang
Managing Editor ZHANG Juan