

Integrating Genomic and Transcriptomic Data to Reveal Genetic Mechanisms Underlying Piao Chicken Rumpless Trait

Yun-Mei Wang^{1,2,3,#,a}, Saber Khederzadeh^{1,2,#,b}, Shi-Rong Li^{1,2,c}, Newton Otieno Otecko^{1,2,d}, David M Irwin^{4,e}, Mukesh Thakur^{1,5,f}, Xiao-Die Ren^{1,2,g}, Ming-Shan Wang^{1,2,*,h}, Dong-Dong Wu^{1,2,*,i}, Ya-Ping Zhang^{1,2,*,j}

¹State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan, 650223, China

²Kunming College of Life Science, University of the Chinese Academy of Sciences, Kunming, Yunnan, 650223, China

³Center for Neurobiology and Brain Restoration, Skolkovo Institute of Science and Technology, Moscow 143026, Russia

⁴Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada

⁵Zoological Survey of India, New Alipore Kolkata-700053, West Bengal, India

Equal contribution.

* Corresponding authors.

E-mail: wangmingshan@mail.kiz.ac.cn (Wang MS), wudongdong@mail.kiz.ac.cn (Wu DD), zhangyp@mail.kiz.ac.cn (Zhang YP).

Running title: Wang YM et al / Genetic Mechanisms of Chicken Rumpless Trait

^aORCID: 0000-0001-8165-1320.

^bORCID: 0000-0002-0115-8710.

^cORCID: 0000-0002-6478-4364.

^dORCID: 0000-0002-9149-4776.

^eORCID: 0000-0001-6131-4933.

^fORCID: 0000-0003-2609-7579.

^gORCID: 0000-0002-5408-7127.

^hORCID: 0000-0001-9847-9803.

ⁱORCID: 0000-0001-7101-7297.

^jORCID: 0000-0002-5401-1114.

34

35 In total, there are 8157 words, 4 figures, 2 supplementary figures, 4 supplementary tables, 75
36 references, and 20 references from 2014.

Abstract

Piao chicken, a rare Chinese native poultry breed, lacks primary tail structures, such as pygostyle, caudal vertebra, uropygial gland and tail feathers. So far, the molecular mechanisms underlying tail absence in this breed have remained unclear. We employed comprehensive comparative transcriptomic and genomic analyses to unravel potential genetic underpinnings of rumplessness in the Piao chicken. Our results reveal many biological factors involved in tail development and several genomic regions under strong positive selection in this breed. These regions contain candidate genes associated with rumplessness, including *IRX4*, *IL-18*, *HSPB2*, and *CRYAB*. Retrieval of quantitative trait loci (QTL) and gene functions implied that rumplessness might be consciously or unconsciously selected along with the high-yield traits in Piao chicken. We hypothesize that strong selection pressures on regulatory elements might lead to gene activity changes in mesenchymal stem cells of the tail bud and eventually result in tail truncation by impeding differentiation and proliferation of the stem cells. Our study provides fundamental insights into early initiation and genetic bases of the rumpless phenotype in Piao chicken.

KEYWORDS: Comparative transcriptomics; Population genomics; Rumplessness; Vertebra development

Introduction

Body elongation along the anterior-posterior axis is a distinct phenomenon during vertebrate embryo development. Morphogenesis of caudal structures occurs during posterior axis elongation. The tail bud contributes most of the tail portion [1]. This structure represents remains of the primitive streak and Hensen's node and comprises a dense mass of undifferentiated mesenchymal cells [1]. Improper patterning of the tail bud may give rise to a truncated or even absent tail [2]. Previous investigations implicated many factors in the formation of posterior structures. For example, loss of *T Brachyury Transcription Factor (T)* causes severe defects in mouse caudal structures, including the lack of notochord and allantois, abnormal somites, and a short tail [3]. Genetic mechanisms for rumplessness vary among the different breeds of rumpless chicken. For instance, Wang et al. revealed that rumplessness in Hongshan chicken, a Chinese indigenous breed, is a Z chromosome-linked dominant trait and may be associated with the region containing candidates like *LINGO2* and the pseudogene *LOC431648* [4,5]. In Araucana chicken, a Chilean rumpless breed, the rumpless phenotype is autosomal dominant and probably related to two proneural genes – *IRX1* and *IRX2* [6,7].

Piao chicken, a Chinese autochthonic rumpless breed, is native to Zhenyuan County, Puer City, Yunnan Province, China, and is mainly found in Zhenyuan and adjacent counties [8]. This breed has no pygostyle, caudal vertebra, uropygial gland and tail feathers [8], hence, an ideal model for studying tail development [9]. Through crossbreeding experiments and anatomical observations, Song et al. showed that rumplessness in Piao chicken is autosomal dominant and forms during the embryonic period even though no specific stage was identified [9,10]. However, until now, the genetic mechanisms of rumplessness in this breed have not yet been elucidated.

The advent of next-generation sequencing and microscopy has made it possible to probe embryonic morphogenesis through microscopic examination, to study phenotype evolution using comparative population genomics, and to assess transcriptional profiles associated with specific characteristics via comparative transcriptomics. In this study, we integrated these three methods to uncover the potential genetic bases of the rumpless phenotype in Piao chicken.

Results

Comparative genomic analysis identifies candidate regions for the rumpless trait in Piao chicken

To investigate the genetic mechanisms underlying rumplessness in Piao chicken, we employed comparative population genomics to evaluate population differentiation between the Piao chicken and control chickens with a normal tail. In total, we analyzed the genomes of 20 Piao and 98 control chickens, including 18 red junglefowls (RJFs), 79 other domestic chickens and one green junglefowl (GJF) as an outgroup (Table S1). We identified 27,557,576 single-nucleotide polymorphisms (SNPs), more than half of which (52.5%) were in intergenic regions, 42.1% in introns, and 1.5% in exons. Functional annotation by ANNOVAR [11] revealed that 292,570 SNPs (about 1%) caused synonymous substitutions, while 131,652 SNPs (approximately 0.5%) may alter protein structure and function through non-synonymous mutations or gain and loss of stop codons. A phylogenetic tree and principal component analysis (PCA) showed certain population differences between the Piao chicken and controls (**Figure 1A–C**). High genetic relatedness was found between Piao and domestic chickens from Gongshan, Yunnan (Figure 1C). One Piao chicken sample clearly separated from the other Piao chickens and was close to Chinese domestic chickens from Dali, Yunnan and Baise, Guangxi. Since Zhenyuan, Dali, and Baise are geographically close neighbors, it is possible that Piao chicken is beginning to mix with exotic breeds due to the emerging advancements in transport networks in these regions.

Based on fixation index (F_{ST}), nucleotide diversity (Pi) and genotype frequency (GF), we searched for genomic regions, SNPs, and insertions/deletions (indels) with high population differences between the Piao and control chickens (Materials and methods). We identified 40kb-sliding window genomic regions with strong signals of positive selection in the Piao chicken, out of which 28 were highly recovered in 1000 random samples of 20 controls from the 98 control chickens (Figure 1D–F and S1A, Table S2, and Materials and methods). Three regions with the strongest repeatable selection signals were chr2: 85.52–86.07Mb, chr4: 75.42–75.48Mb and chr24: 6.14–6.25Mb.

Chr2: 85.52–86.07Mb had a high proportion of highly differentiated ($GF > 0.8$) SNPs and indels, approximately 13.5% (66/488) and 10.4% (5/48), respectively (Figure 1F and Table S2). There are many important genes in this region. For instance, *IRX4*, along with *IRX1* and *IRX2*, belongs to the same *Iroquois* genomic cluster, *i.e.*, the *IrxA* cluster [12]. Tena et al. discovered an evolutionarily conserved three-dimensional structure in the vertebrate *IrxA* cluster, which facilitates enhancer sharing and coregulation [12]. *IRX4* was reported to be involved in progenitor cell differentiation [13,14]. *Telomerase Reverse Transcriptase (TERT)* is implicated in spermatogenesis and male infertility [15]. In particular, *ENSGALG00000013155*, a novel gene, showed strong selection signals from both F_{ST} and Pi .

This gene had 39 highly differentiated SNPs (23 located in introns and 16 in the intergenic region with *MOCOS*) and two highly differentiated indels (in introns). We retrieved expression profiles of this gene in different chicken tissues and development stages from three NCBI projects (SRA Accessions: ERP003988, SRP007412 and DRP000595) used in our previous study [16]. The results showed that the expression levels of this gene were high in three tissues, *i.e.*, adipose, adrenal gland and cerebellum, and gradually increased during the early development of the chicken embryo (Figure S2A). The adrenal gland affects lipid metabolism and sexual behavior [17,18], while the cerebellum is involved in emotional processing [19]. These results suggest that *ENSGALG00000013155* plays important roles in fat deposition, sexual behavior and embryogenesis in chicken.

Chr4: 75.42–75.48Mb includes one gene – *Ligand Dependent Nuclear Receptor Corepressor Like (LCORL)*. *LCORL* was reported as a candidate gene for chicken internal organ weight [20], horse body size [21], and cattle production performance [22].

Chr24: 6.14–6.25Mb exhibited the strongest selection signal on chromosome 24 and has a fundamental gene – *Interleukin 18 (IL18)* (Figure 1F). *IL18* is well known as a proinflammatory and proatherogenic cytokine, as well as an IFN- γ inducing factor [23]. The gene modulates acute graft-versus-host disease by enhancing Fas-mediated apoptosis of donor T cells [24]. It can also induce endothelial cell apoptosis in myocardial inflammation and injury [25]. Interestingly, this gene has been reported to play prominent roles in osteoblastic and osteoclastic functions that are crucial for bone remodeling by balancing formation and resorption [26,27]. *IL-18* affects the bone anabolic actions of parathyroid hormone *in vivo* [26], and drives osteoclastogenesis by elevating the production of receptor activator of nuclear factor- κ B (NF- κ B) ligand in rheumatoid arthritis [27]. *IL-18* is also known as a trigger of NF- κ B activation [25]. Sequestration of NF- κ B in zebrafish can disrupt the notochord differentiation and bring about *no tail (ntl)*-like embryos [28]. This no tail phenotype can be rescued by the T-box gene *ntl* (*Brachyury* homologue) [28]. Indeed, in chicken, blocking NF- κ B expression in limb buds gives rise to a dysmorphic apical ectodermal ridge, a decrease of limb size and outgrowth, as well as a failure of distal structures, through the interruption of mesenchymal-epithelial communication [29].

Developmental transcriptome analysis reveals differentially expressed genes (DEGs) associated with tail development

The tail bud begins to form at the Hamburger-Hamilton stage 11 of chicken embryo development [30,31], and undergoes multidimensional morphogenesis in the subsequent three

to ten days [32]. Thus, it should be possible to analyze transcriptional diversity associated with tail generation during this period. Based on this premise, we sought to outline the functional factors involved in caudal patterning using RNA-Sequencing data from 9 Piao and 12 control chicken embryos after seven to nine days of incubation (Table S1). To minimize bias, the control samples were collected from 6 Gushi and 6 Wuding chicken embryos (Figure S2B and Materials and methods).

An evaluation of gene expression identified 437 DEGs between the Piao and control chickens across the three developmental days (**Figure 2A–B**, Table S3, and Materials and methods), including the gene *T* that is implicated in tail development [3]. Gene ontology (GO) enrichment analysis of all DEGs by the database for annotation, visualization and integrated discovery (DAVID v6.8) [33] showed that many biological processes were related to posterior patterning, including muscle development, bone morphogenesis, somitogenesis, as well as cellular differentiation, proliferation and migration (Figure 2C).

There were some interesting DEGs that had an expression fold change (FC) greater than 2 between the Piao and control chickens, including *T-box 3* (*TBX3*), *homeobox B13* (*HOXB13*), *myosin light chain 3* (*MYL3*), *myosin heavy chain 7B* (*MYH7B*), *KERA*, and *paired box 9* (*PAX9*) (Figure 2D). *TBX3*, encoding a transcriptional factor, had an expression in the Piao chickens three times higher than in controls. This gene regulates osteoblast proliferation and differentiation [34]. *HOXB13*, located at the 5' end of the *HOXB* cluster, had an almost undetectable expression in the Piao chickens, but 36.7-fold higher levels in the controls. In mice, loss-of-function mutations in *HOXB13* cause overgrowth of the posterior structures derived from the tail bud, due to disturbances in proliferation inhibition and apoptosis activation [35]. *MYL3* showed 2.2 times lower expression in the Piao chickens compared to controls. This gene encodes a myosin alkali light chain in slow skeletal muscle fibers and modulates contractile velocity [36]. *MYH7B*, with 2.4-fold lower expression in the Piao chickens, encodes a third myosin heavy chain [37]. Mutations in *MYH7B* cause a classical phenotype of left ventricular non-compaction cardiomyopathy [37]. *KERA* encodes an extracellular matrix keratocan, which acts as an osteoblast marker regulating osteogenic differentiation [38]. The expression of *KERA* in the Piao chickens was 2.3-fold higher than in controls. *PAX9* and *PAX1* function redundantly to influence the vertebral column development [39]. Compared to controls, the expression level of *PAX9* in the Piao chickens is nearly 2.6 times lower. Analyzing gene expression patterns revealed that multiple genes are potentially involved in chicken tail development.

Co-expression modules delineate the biological processes relevant to posterior patterning

To further elucidate principal biological pathways regarding caudal development, we constructed correlation networks through weighted gene co-expression network analysis (WGCNA) [40] (Materials and methods). We captured twelve co-expression modules, six of which (M4, M5, M7, M8, M9 and M10) were significantly correlated with the rumpleless phenotype (Pearson correlation, P value < 0.05) (Figure 3A–G).

Functional annotation for these significant modules revealed close linkages with embryonic development. Modules M5 and M10, which were negatively correlated with rumplelessness, showed functional enrichment for skeletal system development and myogenesis, respectively (Figure 3C and G). Modules in positive correlation with rumplelessness were implicated in several different pathways, including axon guidance and osteoblast differentiation for M4, calcium signaling pathway and neural crest cell migration for M7, actin cytoskeleton organization and dorso-ventral axis formation for M8, as well as transcription and tight junction for M9 (Figure 3B and D–F).

We searched for hub genes in each significant module and visualized the weighted networks with the top hubs (Figure 3B–G, Table S4, and Materials and methods). Interestingly, *TBX3*, one of the six DEGs mentioned above, was an M4 hub gene. Another five DEGs (*HOXB13*, *MYL3*, *MYH7B*, *KERA* and *PAX9*) are all M10 hubs. The M7 hub *ret proto-oncogene (RET)* encodes a transmembrane tyrosine kinase receptor with an extracellular cadherin domain [41]. *RET* induces enteric neuroblast apoptosis through caspases-mediated self-cleavage [42]. *Thrombospondin-1 (THBS1)*, an M5 hub gene, has effects on epithelial-to-mesenchymal transition and osteoporosis [43,44]. An M8 hub *WASL* is essential for Schwann cell cytoskeletal dynamics and myelination [45]. The M9 hub *WNT11* is crucial for gastrulation and axis formation [46,47]. These findings suggest that functional polygenic inter-linkages influence posterior patterning during chicken development.

The only two DEGs under strong selection co-localized with *IL18*

To check whether our DEGs were strongly selected, we retrieved genes in the 40kb-sliding window regions with strong repeatable selection signals. We simultaneously searched for any highly differentiated SNP and indel in or flanking these DEGs. Finally, we only obtained two DEGs under strong selection, *i.e.*, *Heat Shock Protein Beta-2 (HSPB2)* and *Alpha-Crystallin B Chain (CRYAB)*. Unfortunately, we found no highly differentiated SNP or indel related to these two DEGs. *HSPB2* and *CRYAB* are both located near *IL18* in chr24: 6.14–6.25Mb (Figure 1F). They encode a small heat shock protein and are essential for calcium uptake in

myocyte mitochondria [48]. Nevertheless, they function non-redundantly: *HSPB2* balances energy as a binding partner of dystrophin myotonic protein kinase, while *CRYAB* is implicated in anti-apoptosis and cytoskeletal remodeling [49].

Discussion

Accurate molecular regulation and control are vital for biological development and existence. Interfering these functional networks can lead to embryo death, diseases, deformities, or even the evolution of new characters [3–7]. Selection makes domestic animals achieve numerous phenotypic changes in morphology, physiology or behavior by modulating one or several components of primary biological networks [16].

In this study, we combined comparative transcriptomics and population genomics to explore the genetic mechanisms underlying rumplessness of Piao chicken. Our transcriptomic analyses presented many biological pathways that might be important to the late development of chicken tail. Genome-wide comparative analyses revealed several genomic regions under robust positive selection in the Piao chicken. These regions contain some fundamental genes, including *TERT*, *ENSGALG00000013155*, *IRX4*, *LCORL*, *IL-18*, *HSPB2*, and *CRYAB*, only two of which (*HSPB2* and *CRYAB*) were DEGs between the Piao and control chickens during D7–D9. Some of these genes might be associated with performance traits in the Piao chicken, such as *TERT* for egg fertilization, *ENSGALG00000013155* for fat deposition, and *LCORL* for body weight. Others might be implicated in axis elongation, such as *IRX4*, *IL-18*, *HSPB2* and *CRYAB*, through signaling pathways like NF- κ B, calcium or apoptosis. Meanwhile, by retrieving all available QTL from the Animal QTLdb [50], we found that these regions were associated with production traits, such as growth, body weight, egg number, duration of broodiness and broody frequency (Table S2). In spite of lack of information about the evolutionary history of the Piao chicken, we know that this breed is characterized by high production, including elevated fat deposition rates, meat production, egg fertilization and egg hatchability [8,51]. Moreover, although Piao chicken originated in a relatively closed environment with limited genetic admixture with exotic breeds [51], it has high genetic variability and five maternal lineages [52], implying high hybrid fertility in the breed. Thus, we speculate that rumplessness, which exposes the posterior orifice of Piao chicken, might make intra-population or inter-population mating easier for the breed. This easier mating might increase egg fertilization, egg number, broody frequency and genetic variability in the Piao chicken. We propose that rumplessness might be consciously or unconsciously selected

along with the high-yield traits of Piao chicken (**Figure 4A**).

In addition, we found that all highly differentiated SNPs and indels, related to the above candidate genes, are located in noncoding regions or consist of synonymous mutations. Therefore, we hypothesize that positive selection pressures on the regulatory elements of some of these candidate genes would produce functional changes, which then leads to the rumpless phenotype in Piao chicken (Figure 4A). These activity changes likely take place at a very early development stage. In fact, when examining embryogenesis, we observed extreme tail truncation in most of the Piao embryos during the fourth day of incubation and later, while some Piao and all control embryos presented normal posterior structures (Figure 4B). The caudal morphogenesis in Piao embryos confirmed the descriptions from Song et al. [9] and Zwilling [53]. Song et al. found that the rumpless phenotype in Piao chicken is autosomal dominant [9]. Zwilling stated that dominant rumpless mutants arise at the end of the second day of embryo development and are established by the closure of the fourth day [53]. Previous studies have shown that the tail bud comprises a dense mass of undifferentiated mesenchymal cells and forms most of the tail portion [1]. The structure undergoes multidimensional morphogenesis from the third day of development [32]. Thus, we postulate that ectopic expressions, driven by positive selection pressures, might be initiated in the tail bud and impede normal differentiation and proliferation of the mesenchymal stem cells through signaling pathways like NF- κ B, calcium or proapoptosis (Figure 4A). The result is the derailing of the mesenchymal maintenance in the tail bud and eventual failure of normal tail development. Hindrance to tail formation might have overarching impacts on later developmental processes of distal structures, for instance, *MYL3* and *MYH7B* involved in myogenesis, or *PAX9* and *PAX1* participating in vertebral column formation (Figure 4A). In particular, the strong selection on the proinflammatory cytokine gene *IL-18* could also be an adaptive response for a robust immunity, as rumplessness leaves the posterior orifice exposed to infections. However, as rumplessness in the Piao chicken is autosomal dominant, it is hard to confirm the rumpless phenotype and sample without RNA degradation before the fourth day of development. Therefore, we could barely even validate the expression of the identified genes in the tail bud further.

Previous studies have shown differences in the genetic architectures of taillessness among different chicken breeds [4–7], for example, Hongshan chicken [4,5] versus Araucana chicken [6,7]. Interestingly, Hongshan chicken has normal coccygeal vertebrae, while Araucana and Piao breeds have no caudal vertebrae and their rumplessness is autosomal dominant. Our results reveal potential similarities in the spatiotemporal bases of rumplessness in Piao and

other dominant rumpless chickens [6,7,53]. This implies that autosomal-dominant rumplessness in chicken probably has the same genetic mechanisms and embryogenesis. By integrating comparative transcriptomics, population genomics and microscopic examination of embryogenesis, this study provides a basic understanding of genes and biological pathways that may be related, directly or indirectly, to rumplessness of the Piao chicken. Our work could shed light on the phenomenon of tail degeneration in vertebrates. Future endeavors should address the limitation to discern specific causative mutations that lead to tail absence in chicken.

Conclusion

By combining comparative transcriptomics, population genomics and microscopic examination of embryogenesis, we reveal the potential genetic mechanisms of rumplessness in Piao chicken. This work could facilitate a deeper understanding of tail degeneration in vertebrates.

Materials and methods

Ethics statement

All animals were handled following the animal experimentation guidelines and regulations of the Kunming Institute of Zoology. This research was approved by the Institutional Animal Care and Use Committee of the Kunming Institute of Zoology.

Whole-genome re-sequencing data preparation

DNA was extracted from blood samples from 20 Piao chickens by the conventional phenol-chloroform method. Quality checks and quantification were performed using agarose gel electrophoresis and NanoDrop 2000 spectrophotometer. Paired-end libraries were prepared by the NEBNext[®] Ultra[™] DNA Library Prep Kit for Illumina[®] (NEB, USA) and then sequenced on the Illumina HiSeq2500 platform after quantification. 150bp paired-end reads were generated. Finally, we obtained tenfold average sequencing depth for each individual. Additionally, we used genomes from another 96 chickens with a normal tail from an unpublished project in our laboratory, including one outgroup GJF, 18 RJFs and 77 other domestic chickens (Table S1). Sequence coverage for these genomes ranged from about 5 to 110. Control genomic data from two additional individuals were downloaded from NCBI

(SRA Accession: SRP022583) at <https://www.ncbi.nlm.nih.gov/sra> [54].

As the Piao chicken has been highly purified after a long time of conservation, it is hard to get enough normal-tailed Piao chicken samples with no inbreeding. Considering the short divergence time (about 8000 years) between domestic chickens and their ancestor – RJF [55], we used various normal-tailed chicken breeds as controls. Nevertheless, the number of each control breed was kept small. This design was based on previous studies [56,57]. In our opinion, the complex constitution of controls would reduce the background noises from some specific control breeds, but highlight the signatures of the target common trait, when compared with the Piao chicken where rumplessness is the main feature. Besides, comparing with exotic control breeds should weaken population differences among the Piao chicken, but highlight the common traits in this breed, like rumplessness.

SNPs and indels detection

Variant calling followed a general BWA/GATK pipeline. Low-quality data were first trimmed using the software btrim [58]. Filtered reads were mapped to the galGal4 reference genome based on the BWA-MEM algorithm [59], with default settings and marking shorter split hits as secondary. We utilized the Picard toolkit (picard-tools-1.56, <http://broadinstitute.github.io/picard/>) to sort bam files and mark duplicates with the SortSam and MarkDuplicates tools, respectively. The Genome Analysis Toolkit (GATK, v2.6-4-g3e5ff60) [60] was employed for preprocessing and SNP/indel calling utilizing the tools RealignerTargetCreator, IndelRealigner, BaseRecalibrator, PrintReads, and UnifiedGenotyper. SNP variants were filtered using the VariantFiltration tool with the criteria: $QUAL < 40.0$, $MQ < 25.0$, $MQ0 \geq 4 \ \&\& \ ((MQ0/(1.0*DP)) > 0.1)$. The same criteria were used for indels except setting: $QD < 2.0 \ || \ FS > 200.0 \ || \ InbreedingCoeff < -0.8 \ || \ ReadPosRankSum < -20.0$.

Population differentiation evaluation

To display the population structure of the Piao and control chickens, we built a phylogenetic tree based on the weighted neighbor-joining method [61] for all SNP sites, and visualised it using the MEGA6 software [62]. We pruned SNPs based on linkage disequilibrium utilizing the PLINK tool (v1.90b3w) [63] with the options of ‘--indep-pairwise 50 10 0.2 --maf 0.05’. PCA was performed using the program GCTA (v1.25.2) [64] without including the GJF outgroup. Following the published formulas, we then computed F_{ST} [65] and P_i [66] values to detect signals of positive selection [56,66]. A 40kb sliding window analysis with steps of

20kb was performed for both F_{ST} and Pi . Nucleotide diversities (ΔPi) was calculated based on the formula $\Delta Pi = Pi \text{ (Control)} - Pi \text{ (Piao)}$. The intersecting 40kb-sliding window regions in the top 1% of descending F_{ST} and ΔPi were regarded as potentially selected candidates, an empirical threshold used previously [56]. Manhattan plots were drawn using the ‘gap’ package in R software [67] ignoring small chromosomes. To reduce false positive rates, we rechecked the candidate regions by analyzing 1000 replicates of 20 controls randomly sampled from all 98 control chickens. We identified intersecting 40kb-sliding window regions in the top 1% of descending F_{ST} and ΔPi for each random sample. We then checked how many times the candidate regions were recovered by these intersecting regions during the 1000 random samples. We calculated the recovery ratio (i.e., times of recovery divided by 1000) for each candidate and defined those with a recovery ratio greater than 0.95 as the final strongly selected sliding regions.

GFs between the Piao and control chickens were compared to retrieve highly differentiated SNPs and indels. In general, there are three genotypes, i.e., 00, 01, and 11. 00 represents two alleles that are both the same as the reference genome. 01 represents one of the two alleles being altered, while the other is the same as the reference genome. 11 represents two alleles that are both altered. Considering sequencing errors, we defined a highly differentiated site using empirical thresholds: first, a site must exist in more than 15 (75%) Piao chickens and more than 50 (51%) control chickens; then, for the eligible site, sum of 01 and 11 must be larger than 0.8 in Piao chicken and less than 0.06 in control chickens. In total, we identified 488 and 48 highly differentiated SNPs and indels, respectively.

Embryos microscopic observation

Fertilized Piao and normal-tailed control chicken eggs were purchased from the Zhenyuan conservation farm of Piao chicken and the Yunnan Agricultural University farm, respectively. Eggs were incubated at 37.5°C with 65% humidity. Embryo tail development was observed from the fourth to tenth day of incubation using a stereoscopic microscope.

RNA isolation and sequencing

A total of 21 samples (9 Piao and 12 control chickens) were collected from the posterior end of embryos after seven to nine days of incubation and stored in RNA_{later} at −80°C. RNA was extracted using Trizol reagent (Invitrogen) and RNeasy Mini Kits (Qiagen) and purified with magnetic oligo-dT beads for mRNA library construction. Paired-end libraries were prepared

by the NEBNext[®] Ultra[™] RNA Library Prep Kit for Illumina[®] (NEB, USA) and sequenced on the Illumina HiSeq2500 platform after quantification. 150bp paired-end reads were generated. Overall, we obtained approximately 5 Gbases (Gbp) of raw data for each library.

For the Piao chickens, there are three biological replicates for each of the three developmental days. By using samples from different developmental stages, we aimed to exclude inconsistent effects during development. Due to sampling difficulty, we used two Chinese native chicken breeds (6 Gushi and 6 Wuding chickens) as controls, in a way similar to the genomic analysis. Both control breeds have two biological replicates for each developmental stage. The Gushi and Wuding chicken are native to Henan and Yunnan, respectively, and both have a normal tail. In our opinion, using two control breeds would reduce trait noises from either breed, but strengthen signals from the common traits between the two breeds, such as having a normal tail compared with Piao chicken where rumplessness is the main feature.

Transcriptomic data processing

We first trimmed low-quality sequence data using the software btrim [58]. Filtered reads were aligned to the chicken genome (galGal4.79, downloaded from <http://asia.ensembl.org/index.html>) using TopHat2 (v2.0.14) [68], with the parameters ‘--read-mismatches’, ‘--read-edit-dist’ and ‘--read-gap-length’ set to no more than three bases. We evaluated gene expression levels by applying HTSeq (v0.6.0) with the union exon model and the whole gene model, coupled with the Cufflinks program available in the Cufflinks tool suite (v2.2.1) [69] using default parameters.

Correction and normalization

To improve the analyses, genes were filtered for expression in the three datasets: in at least 80 percent of the Piao or control samples, gene counts from both HTSeq models (union exon and whole gene) were no less than ten, while lower bound Fragments Per Kilobase of exon model per Million mapped fragments (FPKM) values from Cufflinks were greater than zero. We then performed normalization for gene length and GC content using the ‘cqn’ R package (v1.16.0) [70], based on the filtered count matrix of the HTSeq union exon model. The output values were defined as $\log_2$ (Normalized FPKM). The normalized matrix with genes kept in all three datasets was adjusted for unwanted biological and technical covariates, like development days, breeds, and sequencing lanes, via a linear mixed-effects model as previously described [71]. In detail, we calculated coefficients for these covariates with a linear model and then removed

the variability contributing to them from the original $\log_2$ (Normalized FPKM) values. For example, when adjusting for development days, the number 1, 2 and 3 were used to replace D7, D8 and D9, respectively. We then calculated a coefficient for each gene using the “lm” function in R, with the number substitutes as a covariate. We removed the product of the coefficient and the number substitute from the $\log_2$ (Normalized FPKM) value to obtain the adjusted value. The adjusted data was then used for co-expression network construction.

Differential expression analysis

We applied three methods to identify DEGs. First, 826 DEGs (FDR < 0.05) were identified by the Cuffdiff program in the Cufflinks tool suite with default parameters, using the bam files from TopHat2. Second, 1451 DEGs (FDR < 0.05) were found by DESeq2 [72] based on the read count data from the HTSeq union exon model. Third, 1244 DEGs (FDR < 0.05) were obtained with a linear model, where we used the $\log_2$ (Normalized FPKM) matrix from the normalization step and treated development stages, chicken breeds and sequencing lanes as covariates. In total, 437 DEGs found by all three methods were used as the final DEGs.

Gene co-expression network analysis

To unravel underlying functional processes and genes associated with tail development, we carried out WGCNA in the R package [40] with a one-step automatic and ‘signed’ network type. The soft thresholding power option was set to 12 based on the scale-free topology model, where topology fit index R^2 was first greater than 0.8. The minimum module size was limited to 30. A height cut of 0.25 was chosen to merge highly co-expressed modules (*i.e.*, correlation greater than 0.75). Finally, we obtained a total of 12 modules. M0 consisted of genes that were not included in any other modules, and thus was excluded from further analyses. We performed Pearson correlation to assess module relationships to the rumpless trait, and defined P value < 0.05 as a significant threshold.

Hub genes and network visualization

In general, genes, which have significant correlations to others and the targeted trait, are the most biologically meaningful and thus defined as ‘hub genes’. Here, we referred to module eigengenes (MEs) as ‘hub genes’ dependent on high intramodular connectivity, and absolute values of gene significance (GS) and module membership (kME) greater than 0.2 and 0.8, respectively. GS values reflect tight connections between genes and the targeted trait, while kME mirrors eigengene-based connectivity between a gene expression profile and ME, and is

also known as module membership [40]. The intramodular connectivity values measure the co-expression degree of a gene to other MEs in the module where it belongs. To visualize a weighted network, we ranked hub genes for each module by intramodular connectivity in descending order, and exported network connections between the top 50 hubs into an edge file with a topological overlap threshold of 0.1. The edge files were input to Cytoscape [73] for network analysis. Network plots for modules significantly related to the rumpless trait were then displayed based on decreasing degree values.

Data access

Raw sequence data for the RNA samples and DNA samples of Piao chicken reported in this paper were deposited in the Genome Sequence Archive [74] in BIG Data Center [75], Beijing Institute of Genomics (BIG), Chinese Academy of Sciences (GSA Accession: CRA001387) at <http://bigd.big.ac.cn/gsa>. Codes and input files for the major analytic processes were stored in GitHub (accession is the title of this paper) at <https://github.com>.

Authors' contributions

YPZ, DDW and MSW designed the study. YMW and SK performed data analyses. YMW and SRL finished embryo incubation and observation. YMW and DDW wrote the manuscript. SK, NOO, DMI and MT revised the manuscript. MSW helped with comparative genomic analyses. XDR submitted the data. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Acknowledgments

This work was supported by the Bureau of Science and Technology of Yunnan Province (2015FA026), the Youth Innovation Promotion Association, Chinese Academy of Sciences, and the National Natural Science Foundation of China (grant numbers 31771415, 31801054). We are grateful to the team of Prof. Yong-Wang Miao, Yunnan Agricultural University, for blood collection from adult Piao chickens. We also thank the support of the CAS-TWAS President's Fellowship Program for Doctoral Candidates.

References

- [1] Griffith CM, Wiley MJ, Sanders EJ. The vertebrate tail bud: three germ layers from one tissue. *Anat Embryol* 1992;185:101–13.
- [2] Schoenwolf GC. Effects of complete tail bud extirpation on early development of the posterior region of the chick embryo. *Anat Rec* 1978;192:289–95.
- [3] Stott D, Kispert A, Herrmann BG. Rescue of the tail defect of *Brachyury* mice. *Gene Dev* 1993;7:197–203.
- [4] Wang Q, Pi J, Pan A, Shen J, Qu L. A novel sex-linked mutant affecting tail formation in Hongshan chicken. *Sci Rep-uk* 2017;7:10079.
- [5] Wang Q, Pi J, Shen J, Pan A, Qu L. Genome-wide association study confirms that the chromosome Z harbours a region responsible for rumplessness in Hongshan chickens. *Anim Genet* 2018;49:326–8.
- [6] Freese NH, Lam BA, Staton M, Scott A, Chapman SC. A novel gain-of-function mutation of the proneural *Irx1* and *Irx2* genes disrupts axis elongation in the Araucana rumpless chicken. *PLOS ONE* 2014;9:e112364.
- [7] Noorai RE, Freese NH, Wright LM, Chapman SC, Clark LA. Genome-wide association mapping and identification of candidate genes for the rumpless and ear-tufted traits of the Araucana chicken. *PLOS ONE* 2012;7:e40974.
- [8] Huang Y. Introduction and development of variety resources of Zhenyuan Piao chicken (Chinese Language). *Chinese Journal of Animal Husbandry and Veterinary Medicine* 2014:135–6.
- [9] Song C, Chao D, Guo C, Song W, Su J, Han W, et al. Detection and genetic study of rumpless trait in Piao chicken. *China Poultry* 2015;37:12–5.
- [10] Song C, Han W, Hu Y, Xu W, Di C, Su Y, et al. Observation of bone loss in the tail of Piao chicken (Chinese Language). *China Poultry* 2014;36:53.
- [11] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:E164.
- [12] Tena JJ, Alonso ME, de la Calle-Mustienes E, Splinter E, de Laat W, Manzanares M, et al. An evolutionarily conserved three-dimensional structure in the vertebrate *Irx* clusters facilitates enhancer sharing and coregulation. *Nat Commun* 2011;2:310.
- [13] Nelson DO, Lalit PA, Biermann M, Markandeya YS, Capes DL, Addesso L, et al. *Irx4* marks a multipotent, ventricular-specific progenitor cell. *Stem Cells* 2016;34:2875–88.
- [14] Raouf A, Zhao Y, To K, Stingl J, Delaney A, Barbara M, et al. Transcriptome analysis of the normal human mammary cell commitment and differentiation process. *Cell Stem Cell* 2008;3:109–18.
- [15] Gegenschatz-Schmid K, Verkauskas G, Demougin P, Bilius V, Dasevicius D, Stadler MB, et al. *DMRTC2*, *PAX7*, *BRACHYURY/T* and *TERT* are implicated in male germ cell development following curative hormone treatment for cryptorchidism-induced infertility. *Genes* 2017;8:267.
- [16] Wang Y-M, Xu H-B, Wang M-S, Otecko NO, Ye L-Q, Wu D-D, et al. Annotating long intergenic non-coding RNAs under artificial selection during chicken domestication. *BMC Evol Biol* 2017;17:192.
- [17] Kargi AY, Iacobellis G. Adipose tissue and adrenal glands: novel pathophysiological mechanisms and clinical applications. *Int J Endocrinol* 2014;2014.
- [18] Kuhnle U, Bullinger M. Outcome of congenital adrenal hyperplasia. *Pediatr Surg Int* 1997;12:511–5.
- [19] Schutter DJ, Van Honk J. The cerebellum on the rise in human emotion. *The Cerebellum* 2005;4:290–4.
- [20] Moreira GCM, Salvian M, Boschiero C, Cesar ASM, Reecy JM, Godoy TF, et al. Genome-wide association scan for QTL and their positional candidate genes associated with internal organ traits in chickens. *BMC Genomics* 2019;20:669.
- [21] Metzger J, Schrimpf R, Philipp U, Distl O. Expression levels of *LCORL* are associated with body size in horses. *PLOS ONE* 2013;8:e56497.
- [22] Lindholm-Perry AK, Sexten AK, Kuehn LA, Smith TPL, King DA, Shackelford SD, et al. Association,

effects and validation of polymorphisms within the *NCAPG - LCORL* locus located on BTA6 with feed intake, gain, meat and carcass traits in beef cattle. BMC Genet 2011;12:103.

[23] Fantuzzi G, Puren AJ, Harding MW, Livingston DJ, Dinarello CA. Interleukin-18 regulation of interferon γ production and cell proliferation as shown in interleukin-1 β -converting enzyme (caspase-1)-deficient mice. Blood 1998;91:2118–25.

[24] Reddy P, Teshima T, Kukuruga M, Ordemann R, Liu C, Lowler K, et al. Interleukin-18 regulates acute graft-versus-host disease by enhancing *Fas*-mediated donor T cell apoptosis. J Exp Med 2001;194:1433–40.

[25] Chandrasekar B, Vemula K, Surabhi RM, Li-Weber M, Owen-Schaub LB, Jensen LE, et al. Activation of intrinsic and extrinsic proapoptotic signaling pathways in interleukin-18-mediated human cardiac endothelial cell death. J Biol Chem 2004;279:20221–33.

[26] Raggatt LJ, Qin L, Tamasi J, Jefcoat SC, Shimizu E, Selvamurugan N, et al. Interleukin-18 is regulated by parathyroid hormone and is required for its bone anabolic actions. J Biol Chem 2008;283:6790–8.

[27] Dai S-M, Nishioka K, Yudoh K. Interleukin (IL) 18 stimulates osteoclast formation through synovial T cells in rheumatoid arthritis: comparison with IL1 β and tumour necrosis factor α . Ann Rheum Dis 2004;63:1379–86.

[28] Correa RG, Tergaonkar V, Ng JK, Dubova I, Izpisua-Belmonte JC, Verma IM. Characterization of NF- κ B/I κ B proteins in zebra fish and their involvement in notochord development. Mol Cell Biol 2004;24:5257–68.

[29] Bushdid PB, Brantley DM, Yull FE, Blaeuer GL, Hoffman LH, Niswander L, et al. Inhibition of NF- κ B activity results in disruption of the apical ectodermal ridge and aberrant limb morphogenesis. Nature 1998;392:615–8.

[30] Schoenwolf GC. Histological and ultrastructural observations of tail bud formation in the chick embryo. Anat Rec 1979;193:131–47.

[31] Hamburger V, Hamilton HL. A series of normal stages in the development of the chick embryo. J Morphol 1951;88:49–92.

[32] Schoenwolf GC. Morphogenetic processes involved in the remodeling of the tail region of the chick embryo. Anat Embryol 1981;162:183–97.

[33] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 2008;4:44–57.

[34] Govoni KE, Lee SK, Chadwick RB, Yu H, Kasukawa Y, Baylink DJ, et al. Whole genome microarray analysis of growth hormone-induced gene expression in bone: T-box3, a novel transcription factor, regulates osteoblast proliferation. Am J Physiol Endocrinol Metab 2006;291:E128–E36.

[35] Economides KD, Zeltser L, Capecchi MR. *Hoxb13* mutations cause overgrowth of caudal spinal cord and tail vertebrae. Dev Biol 2003;256:317–30.

[36] Cohen-Haguenauer O, Barton PJ, Van Cong N, Cohen A, Masset M, Buckingham M, et al. Chromosomal assignment of two myosin alkali light-chain genes encoding the ventricular/slow skeletal muscle isoform and the atrial/fetal muscle isoform (*MYL3*, *MYL4*). Hum Genet 1989;81:278–82.

[37] Esposito T, Sampaolo S, Limongelli G, Varone A, Formicola D, Diodato D, et al. Digenic mutational inheritance of the integrin α 7 and the myosin heavy chain 7B genes causes congenital myopathy with left ventricular non-compact cardiomyopathy. Orphanet J Rare Dis 2013;8:91.

[38] Igwe JC, Gao Q, Kizivat T, Kao WW, Kalajzic I. Keratocan is expressed by osteoblasts and can modulate osteogenic differentiation. Connect Tissue Res 2011;52:401–7.

[39] Peters H, Wilm B, Sakai N, Imai K, Maas R, Balling R. Pax1 and Pax9 synergistically regulate vertebral column development. Development 1999;126:5399–408.

[40] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC bioinformatics 2008;9:559.

- [41] Takeichi M. Cadherin cell adhesion receptors as a morphogenetic regulator. *Science* 1991;251:1451–5.
- [42] Bordeaux MC, Forcet C, Granger L, Corset V, Bidaud C, Billaud M, et al. The RET proto-oncogene induces apoptosis: a novel mechanism for Hirschsprung disease. *The EMBO Journal* 2000;19:4056–63.
- [43] Jayachandran A, Anaka M, Prithviraj P, Hudson C, McKeown SJ, Lo P-H, et al. Thrombospondin 1 promotes an aggressive phenotype through epithelial-to-mesenchymal transition in human melanoma. *Oncotarget* 2014;5:5782–97.
- [44] Lawler PR, Lawler J. Molecular basis for the regulation of angiogenesis by thrombospondin-1 and -2. *CSH Perspect Med* 2012;2:a006627.
- [45] Jin F, Dong B, Georgiou J, Jiang Q, Zhang J, Bharioke A, et al. N-WASp is required for Schwann cell cytoskeletal dynamics, normal myelin gene expression and peripheral nerve myelination. *Development* 2011;138:1329–37.
- [46] Heisenberg C-P, Tada M, Rauch G-J, Saude L, Concha ML, Geisler R, et al. Silberblick/Wnt11 mediates convergent extension movements during zebrafish gastrulation. *Nature* 2000;405:76–81.
- [47] Tao Q, Yokota C, Puck H, Kofron M, Birsoy B, Yan D, et al. Maternal Wnt11 activates the canonical wnt signaling pathway required for axis formation in *Xenopus* embryos. *Cell* 2005;120:857–71.
- [48] Kadono T, Zhang XQ, Srinivasan S, Ishida H, Barry WH, Benjamin IJ. CRYAB and HSPB2 deficiency increases myocyte mitochondrial permeability transition and mitochondrial calcium uptake. *J Mol Cell Cardiol* 2006;40:783–9.
- [49] Pinz I, Robbins J, Rajasekaran NS, Benjamin IJ, Ingwall JS. Unmasking different mechanical and energetic roles for the small heat shock proteins CryAB and HSPB2 using genetically modified mouse hearts. *The FASEB Journal* 2008;22:84–92.
- [50] Hu Z-L, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res* 2018;47:D701–D10.
- [51] Xie Q, Yang L. A investigation report on Puer Piao chicken resources (Chinese Language). *Yunnan Journal of Animal Science and Veterinary Medicine* 2009:17–8.
- [52] Gongpan P-c, Li-xian L, Da-lin L, Yue-yun Y, Li-min S, Run Y, et al. The investigation of genetic diversity of Piao chicken based on mitochondrial DNA D-loop region sequence. *Journal of Yunnan Agricultural University (Natural Science)* 2011;2:211–23.
- [53] Zwilling E. The development of dominant rumplessness in chick embryos. *Genetics* 1942;27:641–56.
- [54] Fan W-L, Ng CS, Chen C-F, Lu M-YJ, Chen Y-H, Liu C-J, et al. Genome-wide patterns of genetic variation in two domestic chickens. *Genome Biol Evol* 2013;5:1376–92.
- [55] West B, Zhou B-X. Did chickens go North? New evidence for domestication. *J Archaeo Sci* 1988;15:515–33.
- [56] Wang M-S, Huo Y-X, Li Y, Otecko NO, Su L-Y, Xu H-B, et al. Comparative population genomics reveals genetic basis underlying body size of domestic chickens. *J Mol Cell Biol* 2016;8:542–52.
- [57] Guo X, Li Y-Q, Wang M-S, Wang Z-B, Zhang Q, Shao Y, et al. A parallel mechanism underlying frizzle in domestic chickens. *J Mol Cell Biol* 2018;10:589–91.
- [58] Kong Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing technologies. *Genomics* 2011;98:152–3.
- [59] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997* 2013.
- [60] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.

[61] Bruno WJ, Socci ND, Halpern AL. Weighted neighbor joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol Biol Evol* 2000;17:189–97.

[62] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* 2013;30:2725–9.

[63] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum GeneT* 2007;81:559–75.

[64] Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* 2011;88:76–82.

[65] Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of natural selection. *Genome Res* 2002;12:1805–14.

[66] Wang M-S, Zhang R-w, Su L-Y, Li Y, Peng M-S, Liu H-Q, et al. Positive selection rather than relaxation of functional constraint drives the evolution of vision during chicken domestication. *Cell Res* 2016;26:556–73.

[67] Zhao JH. gap: Genetic Analysis Package. *J Stat Softw* 2007;023.

[68] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol* 2013;14:R36.

[69] Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 2012;7:562–78.

[70] Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile normalization. *Biostatistics* 2012;13:204–16.

[71] Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature* 2016;540:423–7.

[72] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.

[73] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13:2498–504.

[74] Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, et al. GSA: Genome Sequence Archive. *Genom Proteom Bioinf* 2017;15:14–8.

[75] Big Data Center Members. Database resources of the BIG Data Center in 2018. *Nucleic Acids Res* 2018;46:D14–D20.

Figure legends:

Figure 1 Population genomics analysis

A. Weighted phylogenetic tree of Piao chicken (purple) and controls grouped into RJF (red), GJF outgroup (black) and others (green). **B.** PCA plot of Piao and control chickens. Color groups are same as **A**. **C.** PCA plot similar to **B**, but marking control chickens from different places with different colors and shapes (cross is RJF, and the solid point represents other domestic chickens). GX_BS_DC, domestic chicken from Baise, Guangxi; YN_DL_DC, domestic chicken from DL, Yunnan; YN_GS_DC, domestic chicken from Gongshan, Yunnan; for the specification of other abbreviations please see Table S1. **D.** Manhattan plots of F_{ST} and ΔPi based on the median sites of 40kb-sliding window regions. Black arrows point out strongly selected regions mentioned in the main text. **E.** Scatter plot for F_{ST} and ΔPi of 40kb-sliding window regions. Larger dots are 112 selected sliding regions and colored if its recovery ratio of random sampling is greater than 0.95 in both F_{ST} and ΔPi , while those in strongly selected regions mentioned in the main text are marked with different colors. **F.** Lines of F_{ST} and $\Delta Pi \times 100$ values plotted by the median sites of 40kb-sliding windows around the two strongly selected regions, i.e., chr2: 85.52–86.07Mb and chr24: 6.14–6.25Mb. Dark magenta and midnight blue dots represent highly differentiated SNPs and indels, respectively. The colored fillers are genes located in the two regions. Dashed lines indicate the top 1% threshold of descending F_{ST} and $\Delta Pi \times 100$.

Figure 2 Differential expression analysis

A. Relationships of the number of DEGs from Cuffdiff, DESeq2, and the linear model (LM). **B.** Heatmap and hierarchical clustering dendrogram from FPKM of DEGs. Rows represent DEGs while columns show samples. **C.** Significant categories among DEGs by DAVID annotation. The number on the right of a bar presents gene number in the category. X-axis represents $-\log_{10} P$ value. **D.** FPKM values of six representative DEGs in Piao and control chickens.

Figure 3 Gene co-expression module analysis

A. Module relationships with rumpleness. Values in and outside a parenthesis indicate P value and Pearson correlation coefficient, respectively. Modules were gradiently colored by Pearson correlation coefficients. Modules significantly correlated to rumpleness were marked with a red asterisk, while a grey asterisk labeled M0 to indicate that the module was

excluded from the further analyses. **B–G.** Network plots from top 50 hub genes of six significant modules. Significant annotation categories of these modules from DAVID were colored according to network dots. Dot sizes indicate gene connectivity to others in the module.

Figure 4 Proposed mechanisms underlying rumplessness in Piao chicken and microscopic examination of chicken tail embryogenesis

A. Rumplessness might be consciously or unconsciously selected along with the high-yield traits of Piao chicken. Strong positive selection pressures on regulatory elements of some candidate genes might lead to gene activity changes in the tail bud. These ectopic expressions would destroy mesenchymal maintenance by impeding differentiation and proliferation of mesenchymal stem cells in the tail bud, through multiple cell survival and differentiation pathways. This could then disrupt tail formation and prevent later developmental processes of the distal structures. **B.** The morphology of the posterior region during the fourth to fifth day of embryo development in rumpless Piao chicken, Piao chicken with a normal tail, and control chicken.

Supplementary materials

Figure S1 Random sampling recovery ratio of the 112 strongly selected 40kb-sliding window regions

The top two bar plots show the recovery ratio for F_{ST} and ΔPi in 1000 random samples of 20 controls from the 98 control chickens. The blue dashed lines indicate a recovery ratio of 0.95. Red asterisks show sliding window regions with a recovery ratio greater than 0.95 in both F_{ST} and ΔPi , while the short black lines above the asterisks indicate candidate genes in the sliding regions. The bottom two line charts display F_{ST} and ΔPi values based on the 98 control samples. Dashed lines indicate the top 1% threshold of descending F_{ST} and ΔPi . The colored fillers present sliding window regions in the selected regions mentioned in the main text. The 40kb-sliding window regions are presented based on their median sites in order of chromosomal position.

Figure S2 Expression levels of *ENSGALG00000013155* and chicken pictures

A. FPKM values of *ENSGALG00000013155* in different chicken tissues and development stages from the three NCBI projects (SRA Accessions: ERP003988, SRP007412 and

DRP000595) that were used in our previous work [16]. **B.** The pictures of male (M) and female (F) individuals of Piao chicken, Gushi chicken, and Wuding chicken.

Table S1 RNA and DNA sample information

Table S2 Strongly selected regions, highly differentiated SNPs and indels, as well as QTL annotations

Table S3 Expression values of DEGs and their DAVID annotation

Table S4 Hub genes of modules significantly correlated to rumplessness







