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]

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

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

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

[4]*Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Canada*

[5]*Zoological 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*

[a]ORCID**:** 0000-0001-8165-1320.

[b]ORCID**:** 0000-0002-0115-8710.

[c]ORCID**:** 0000-0002-6478-4364.

[d]ORCID**:** 0000-0002-9149-4776.

[e]ORCID**:** 0000-0001-6131-4933.

[f]ORCID**:** 0000-0003-2609-7579.

[g]ORCID**:** 0000-0002-5408-7127.

[h]ORCID**:** 0000-0001-9847-9803.

[i]ORCID**:** 0000-0001-7101-7297.

[j]ORCID**:** 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

4

88　　To investigate the genetic mechanisms underlying rumplessness in Piao chicken, we

89　　employed comparative population genomics to evaluate population differentiation between

90　　the Piao chicken and control chickens with a normal tail. In total, we analyzed the genomes of

91　　20 Piao and 98 control chickens, including 18 red junglefowls (RJFs), 79 other domestic

92　　chickens and one green junglefowl (GJF) as an outgroup (Table S1). We identified 27,557,576

93　　single-nucleotide polymorphisms (SNPs), more than half of which (52.5%) were in intergenic

94　　regions, 42.1% in introns, and 1.5% in exons. Functional annotation by ANNOVAR [11]

95　　revealed that 292,570 SNPs (about 1%) caused synonymous substitutions, while 131,652

96　　SNPs (approximately 0.5%) may alter protein structure and function through

97　　non-synonymous mutations or gain and loss of stop codons. A phylogenetic tree and principal

98　　component analysis (PCA) showed certain population differences between the Piao chicken

99　　and controls (**Figure 1**A−C). High genetic relatedness was found between Piao and domestic

100　　chickens from Gongshan, Yunnan (Figure 1C). One Piao chicken sample clearly separated

101　　from the other Piao chickens and was close to Chinese domestic chickens from Dali, Yunnan

102　　and Baise, Guangxi. Since Zhenyuan, Dali, and Baise are geographically close neighbors, it is

103　　possible that Piao chicken is beginning to mix with exotic breeds due to the emerging

104　　advancements in transport networks in these regions.

105　　　Based on fixation index (F_{ST}), nucleotide diversity (*Pi*) and genotype frequency (GF), we

106　　searched for genomic regions, SNPs, and insertions/deletions (indels) with high population

107　　differences between the Piao and control chickens (Materials and methods). We identified 112

108　　40kb-sliding window genomic regions with strong signals of positive selection in the Piao

109　　chicken, out of which 28 were highly recovered in 1000 random samples of 20 controls from

110　　the 98 control chickens (Figure 1D−F and S1A, Table S2, and Materials and methods). Three

111　　regions with the strongest repeatable selection signals were chr2: 85.52–86.07Mb, chr4:

112　　75.42–75.48Mb and chr24: 6.14–6.25Mb.

113　　　Chr2: 85.52–86.07Mb had a high proportion of highly differentiated (GF > 0.8) SNPs and

114　　indels, approximately 13.5% (66/488) and 10.4% (5/48), respectively (Figure 1F and Table

115　　S2). There are many important genes in this region. For instance, *IRX4*, along with *IRX1* and

116　　*IRX2*, belongs to the same *Iroquois* genomic cluster, *i.e.*, the *IrxA* cluster [12]. Tena et al.

117　　discovered an evolutionarily conserved three-dimensional structure in the vertebrate *IrxA*

118　　cluster, which facilitates enhancer sharing and coregulation [12]. *IRX4* was reported to be

119　　involved in progenitor cell differentiation [13,14]. *Telomerase Reverse Transcriptase* (*TERT*)

120　　is implicated in spermatogenesis and male infertility [15]. In particular,

121　　*ENSGALG00000013155*, a novel gene, showed strong selection signals from both F_{ST} and *Pi*.

5

122   This gene had 39 highly differentiated SNPs (23 located in introns and 16 in the intergenic

123   region with *MOCOS*) and two highly differentiated indels (in introns). We retrieved

124   expression profiles of this gene in different chicken tissues and development stages from three

125   NCBI projects (SRA Accessions: ERP003988, SRP007412 and DRP000595) used in our

126   previous study [16]. The results showed that the expression levels of this gene were high in

127   three tissues, *i.e.*, adipose, adrenal gland and cerebellum, and gradually increased during the

128   early development of the chicken embryo (Figure S2A). The adrenal gland affects lipid

129   metabolism and sexual behavior [17,18], while the cerebellum is involved in emotional

130   processing [19]. These results suggest that *ENSGALG00000013155* plays important roles in

131   fat deposition, sexual behavior and embryogenesis in chicken.

132   Chr4: 75.42–75.48Mb includes one gene – *Ligand Dependent Nuclear Receptor*

133   *Corepressor Like* (*LCORL*). *LCORL* was reported as a candidate gene for chicken internal

134   organ weight [20], horse body size [21], and cattle production performance [22].

135   Chr24: 6.14–6.25Mb exhibited the strongest selection signal on chromosome 24 and has a

136   fundamental gene − *Interleukin 18* (*IL18*) (Figure 1F). *IL18* is well known as a

137   proinflammatory and proatherogenic cytokine, as well as an IFN-γ inducing factor [23]. The

138   gene modulates acute graft-versus-host disease by enhancing Fas-mediated apoptosis of donor

139   T cells [24]. It can also induce endothelial cell apoptosis in myocardial inflammation and

140   injury [25]. Interestingly, this gene has been reported to play prominent roles in osteoblastic

141   and osteoclastic functions that are crucial for bone remodeling by balancing formation and

142   resorption [26,27]. *IL-18* affects the bone anabolic actions of parathyroid hormone *in vivo*

143   [26], and drives osteoclastogenesis by elevating the production of receptor activator of

144   nuclear factor-κB (NF-κB) ligand in rheumatoid arthritis [27]. *IL-18* is also known as a trigger

145   of NF-κB activation [25]. Sequestration of NF-κB in zebrafish can disrupt the notochord

146   differentiation and bring about *no tail* (*ntl*)-like embryos [28]. This no tail phenotype can be

147   rescued by the T-box gene *ntl* (*Brachyury* homologue) [28]. Indeed, in chicken, blocking

148   NF-κB expression in limb buds gives rise to a dysmorphic apical ectodermal ridge, a decrease

149   of limb size and outgrowth, as well as a failure of distal structures, through the interruption of

150   mesenchymal-epithelial communication [29].

151

152   **Developmental transcriptome analysis reveals differentially expressed genes (DEGs)**

153   **associated with tail development**

154   The tail bud begins to form at the Hamburger-Hamilton stage 11 of chicken embryo

155   development [30,31], and undergoes multidimensional morphogenesis in the subsequent three

156 to ten days [32]. Thus, it should be possible to analyze transcriptional diversity associated

157 with tail generation during this period. Based on this premise, we sought to outline the

158 functional factors involved in caudal patterning using RNA-Sequencing data from 9 Piao and

159 12 control chicken embryos after seven to nine days of incubation (Table S1). To minimize

160 bias, the control samples were collected from 6 Gushi and 6 Wuding chicken embryos (Figure

161 S2B and Materials and methods).

162   An evaluation of gene expression identified 437 DEGs between the Piao and control

163 chickens across the three developmental days (**Figure 2**A−B, Table S3, and Materials and

164 methods), including the gene *T* that is implicated in tail development [3]. Gene ontology (GO)

165 enrichment analysis of all DEGs by the database for annotation, visualization and integrated

166 discovery (DAVID v6.8) [33] showed that many biological processes were related to posterior

167 patterning, including muscle development, bone morphogenesis, somitogenesis, as well as

168 cellular differentiation, proliferation and migration (Figure 2C).

169   There were some interesting DEGs that had an expression fold change (FC) greater than 2

170 between the Piao and control chickens, including *T-box 3* (*TBX3*), *homeobox B13* (*HOXB13*),

171 *myosin light chain 3* (*MYL3*), *myosin heavy chain 7B* (*MYH7B*), *KERA*, and *paired box 9*

172 (*PAX9*) (Figure 2D). *TBX3*, encoding a transcriptional factor, had an expression in the Piao

173 chickens three times higher than in controls. This gene regulates osteoblast proliferation and

174 differentiation [34]. *HOXB13*, located at the 5' end of the *HOXB* cluster, had an almost

175 undetectable expression in the Piao chickens, but 36.7-fold higher levels in the controls. In

176 mice, loss-of-function mutations in *HOXB13* cause overgrowth of the posterior structures

177 derived from the tail bud, due to disturbances in proliferation inhibition and apoptosis

178 activation [35]. *MYL3* showed 2.2 times lower expression in the Piao chickens compared to

179 controls. This gene encodes a myosin alkali light chain in slow skeletal muscle fibers and

180 modulates contractile velocity [36]. *MYH7B*, with 2.4-fold lower expression in the Piao

181 chickens, encodes a third myosin heavy chain [37]. Mutations in *MYH7B* cause a classical

182 phenotype of left ventricular non-compaction cardiomyopathy [37]. *KERA* encodes an

183 extracellular matrix keratocan, which acts as an osteoblast marker regulating osteogenic

184 differentiation [38]. The expression of *KERA* in the Piao chickens was 2.3-fold higher than in

185 controls. *PAX9* and *PAX1* function redundantly to influence the vertebral column development

186 [39]. Compared to controls, the expression level of *PAX9* in the Piao chickens is nearly 2.6

187 times lower. Analyzing gene expression patterns revealed that multiple genes are potentially

188 involved in chicken tail development.

189

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 rumpless phenotype (Pearson correlation, *P* value < 0.05) (**Figure 3**A−G).

Functional annotation for these significant modules revealed close linkages with embryonic development. Modules M5 and M10, which were negatively correlated with rumplessness, showed functional enrichment for skeletal system development and myogenesis, respectively (Figure 3C and G). Modules in positive correlation with rumplessness 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

224 myocyte mitochondria [48]. Nevertheless, they function non-redundantly: *HSPB2* balances

225 energy as a binding partner of dystrophin myotonic protein kinase, while *CRYAB* is implicated

226 in anti-apoptosis and cytoskeletal remodeling [49].

227

Discussion

229 Accurate molecular regulation and control are vital for biological development and existence.

230 Interfering these functional networks can lead to embryo death, diseases, deformities, or even

231 the evolution of new characters [3–7]. Selection makes domestic animals achieve numerous

232 phenotypic changes in morphology, physiology or behavior by modulating one or several

233 components of primary biological networks [16].

234   In this study, we combined comparative transcriptomics and population genomics to

235 explore the genetic mechanisms underlying rumplessness of Piao chicken. Our transcriptomic

236 analyses presented many biological pathways that might be important to the late development

237 of chicken tail. Genome-wide comparative analyses revealed several genomic regions under

238 robust positive selection in the Piao chicken. These regions contain some fundamental genes,

239 including *TERT*, *ENSGALG00000013155*, *IRX4*, *LCORL*, *IL-18*, *HSPB2*, and *CRYAB*, only

240 two of which (*HSPB2* and *CRYAB*) were DEGs between the Piao and control chickens during

241 D7–D9. Some of these genes might be associated with performance traits in the Piao chicken,

242 such as *TERT* for egg fertilization, *ENSGALG00000013155* for fat deposition, and *LCORL* for

243 body weight. Others might be implicated in axis elongation, such as *IRX4*, *IL-18*, *HSPB2* and

244 *CRYAB*, through signaling pathways like NF-κB, calcium or apoptosis. Meanwhile, by

245 retrieving all available QTL from the Animal QTLdb [50], we found that these regions were

246 associated with production traits, such as growth, body weight, egg number, duration of

247 broodiness and broody frequency (Table S2). In spite of lack of information about the

248 evolutionary history of the Piao chicken, we know that this breed is characterized by high

249 production, including elevated fat deposition rates, meat production, egg fertilization and egg

250 hatchability [8,51]. Moreover, although Piao chicken originated in a relatively closed

251 environment with limited genetic admixture with exotic breeds [51], it has high genetic

252 variability and five maternal lineages [52], implying high hybrid fertility in the breed. Thus,

253 we speculate that rumplessness, which exposes the posterior orifice of Piao chicken, might

254 make intra-population or inter-population mating easier for the breed. This easier mating

255 might increase egg fertilization, egg number, broody frequency and genetic variability in the

256 Piao chicken. We propose that rumplessness might be consciously or unconsciously selected

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

258   In addition, we found that all highly differentiated SNPs and indels, related to the above

259   candidate genes, are located in noncoding regions or consist of synonymous mutations.

260   Therefore, we hypothesize that positive selection pressures on the regulatory elements of

261   some of these candidate genes would produce functional changes, which then leads to the

262   rumpless phenotype in Piao chicken (Figure 4A). These activity changes likely take place at a

263   very early development stage. In fact, when examining embryogenesis, we observed extreme

264   tail truncation in most of the Piao embryos during the fourth day of incubation and later, while

265   some Piao and all control embryos presented normal posterior structures (Figure 4B). The

266   caudal morphogenesis in Piao embryos confirmed the descriptions from Song et al. [9] and

267   Zwilling [53]. Song et al. found that the rumpless phenotype in Piao chicken is autosomal

268   dominant [9]. Zwilling stated that dominant rumpless mutants arise at the end of the second

269   day of embryo development and are established by the closure of the fourth day [53].

270   Previous studies have shown that the tail bud comprises a dense mass of undifferentiated

271   mesenchymal cells and forms most of the tail portion [1]. The structure undergoes

272   multidimensional morphogenesis from the third day of development [32]. Thus, we postulate

273   that ectopic expressions, driven by positive selection pressures, might be initiated in the tail

274   bud and impede normal differentiation and proliferation of the mesenchymal stem cells

275   through signaling pathways like NF-κB, calcium or proapoptosis (Figure 4A). The result is

276   the derailing of the mesenchymal maintenance in the tail bud and eventual failure of normal

277   tail development. Hindrance to tail formation might have overarching impacts on later

278   developmental processes of distal structures, for instance, *MYL3* and *MYH7B* involved in

279   myogenesis, or *PAX9* and *PAX1* participating in vertebral column formation (Figure 4A). In

280   particular, the strong selection on the proinflammatory cytokine gene *IL-18* could also be an

281   adaptive response for a robust immunity, as rumplessness leaves the posterior orifice exposed

282   to infections. However, as rumplessness in the Piao chicken is autosomal dominant, it is hard

283   to confirm the rumpless phenotype and sample without RNA degradation before the fourth

284   day of development. Therefore, we could barely even validate the expression of the identified

285   genes in the tail bud further.

286   Previous studies have shown differences in the genetic architectures of taillessness among

287   different chicken breeds [4–7], for example, Hongshan chicken [4,5] versus Araucana chicken

288   [6,7]. Interestingly, Hongshan chicken has normal coccygeal vertebrae, while Araucana and

289   Piao breeds have no caudal vertebrae and their rumplessness is autosomal dominant. Our

290   results reveal potential similarities in the spatiotemporal bases of rumplessness in Piao and

10

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^{TM} 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

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

324   As the Piao chicken has been highly purified after a long time of conservation, it is hard

325   to get enough normal-tailed Piao chicken samples with no inbreeding. Considering the short

326   divergence time (about 8000 years) between domestic chickens and their ancestor – RJF [55],

327   we used various normal-tailed chicken breeds as controls. Nevertheless, the number of each

328   control breed was kept small. This design was based on previous studies [56,57]. In our

329   opinion, the complex constitution of controls would reduce the background noises from some

330   specific control breeds, but highlight the signatures of the target common trait, when

331   compared with the Piao chicken where rumplessness is the main feature. Besides, comparing

332   with exotic control breeds should weaken population differences among the Piao chicken, but

333   highlight the common traits in this breed, like rumplessness.

334

335   **SNPs and indels detection**

336   Variant calling followed a general BWA/GATK pipeline. Low-quality data were first trimmed

337   using the software btrim [58]. Filtered reads were mapped to the galGal4 reference genome

338   based on the BWA-MEM algorithm [59], with default settings and marking shorter split hits

339   as   secondary.   We   utilized   the   Picard   toolkit   (picard-tools-1.56,

340   http://broadinstitute.github.io/picard/) to sort bam files and mark duplicates with the SortSam

341   and   MarkDuplicates   tools,   respectively.   The   Genome   Analysis   Toolkit   (GATK,

342   v2.6-4-g3e5ff60) [60] was employed for preprocessing and SNP/indel calling utilizing the

343   tools   RealignerTargetCreator,   IndelRealigner,   BaseRecalibrator,   PrintReads,   and

344   UnifiedGenotyper. SNP variants were filtered using the VariantFiltration tool with the criteria:

345    $QUAL < 40.0$, $MQ < 25.0$, $MQ0 >= 4$ && $((MQ0/(1.0*DP)) > 0.1)$. The same criteria were

346   used for indels except setting: $QD < 2.0$ || $FS > 200.0$ || $InbreedingCoeff < –0.8$ ||

347    $ReadPosRankSum < –20.0$.

348

349   **Population differentiation evaluation**

350   To display the population structure of the Piao and control chickens, we built a phylogenic

351   tree based on the weighted neighbor-joining method [61] for all SNP sites, and visualised it

352   using the MEGA6 software [62]. We pruned SNPs based on linkage disequilibrium utilizing

353   the PLINK tool (v1.90b3w) [63] with the options of '--indep-pairwise 50 10 0.2 --maf 0.05'.

354   PCA was performed using the program GCTA (v1.25.2) [64] without including the GJF

355   outgroup. Following the published formulas, we then computed F_{ST} [65] and Pi [66] values to

356   detect signals of positive selection [56,66]. A 40kb sliding window analysis with steps of

357   20kb was performed for both F_{ST} and Pi. Nucleotide diversities (ΔPi) was calculated based on

358   the formula $\Delta Pi = Pi$ (Control) $- Pi$ (Piao). The intersecting 40kb-sliding window regions in

359   the top 1% of descending F_{ST} and ΔPi were regarded as potentially selected candidates, an

360   empirical threshold used previously [56]. Manhattan plots were drawn using the 'gap'

361   package in R software [67] ignoring small chromosomes. To reduce false positive rates, we

362   rechecked the candidate regions by analyzing 1000 replicates of 20 controls randomly

363   sampled from all 98 control chickens. We identified intersecting 40kb-sliding window regions

364   in the top 1% of descending F_{ST} and ΔPi for each random sample. We then checked how

365   many times the candidate regions were recovered by these intersecting regions during the

366   1000 random samples. We calculated the recovery ratio (i.e., times of recovery divided by

367   1000) for each candidate and defined those with a recovery ratio greater than 0.95 as the final

368   strongly selected sliding regions.

369   GFs between the Piao and control chickens were compared to retrieve highly

370   differentiated SNPs and indels. In general, there are three genotypes, *i.e.*, 00, 01, and 11. 00

371   represents two alleles that are both the same as the reference genome. 01 represents one of the

372   two alleles being altered, while the other is the same as the reference genome. 11 represents

373   two alleles that are both altered. Considering sequencing errors, we defined a highly

374   differentiated site using empirical thresholds: first, a site must exist in more than 15 (75%)

375   Piao chickens and more than 50 (51%) control chickens; then, for the eligible site, sum of 01

376   and 11 must be larger than 0.8 in Piao chicken and less than 0.06 in control chickens. In total,

377   we identified 488 and 48 highly differentiated SNPs and indels, respectively.

378

Embryos microscopic observation

379

380   Fertilized Piao and normal-tailed control chicken eggs were purchased from the Zhenyuan

381   conservation farm of Piao chicken and the Yunnan Agricultural University farm, respectively.

382   Eggs were incubated at 37.5°C with 65% humidity. Embryo tail development was observed

383   from the fourth to tenth day of incubation using a stereoscopic microscope.

384

RNA isolation and sequencing

385

386   A total of 21 samples (9 Piao and 12 control chickens) were collected from the posterior end

387   of embryos after seven to nine days of incubation and stored in RNA*later* at –80°C. RNA was

388   extracted using Trizol reagent (Invitrogen) and RNeasy Mini Kits (Qiagen) and purified with

389   magnetic oligo-dT beads for mRNA library construction. Paired-end libraries were prepared

390   by the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) and sequenced

391   on the Illumina HiSeq2500 platform after quantification. 150bp paired-end reads were

392   generated. Overall, we obtained approximately 5 Gbases (Gbp) of raw data for each library.

393   For the Piao chickens, there are three biological replicates for each of the three

394   developmental days. By using samples from different developmental stages, we aimed to

395   exclude inconsistent effects during development. Due to sampling difficulty, we used two

396   Chinese native chicken breeds (6 Gushi and 6 Wuding chickens) as controls, in a way similar

397   to the genomic analysis. Both control breeds have two biological replicates for each

398   developmental stage. The Gushi and Wuding chicken are native to Henan and Yunnan,

399   respectively, and both have a normal tail. In our opinion, using two control breeds would

400   reduce trait noises from either breed, but strengthen signals from the common traits between

401   the two breeds, such as having a normal tail compared with Piao chicken where rumplessness

402   is the main feature.

403

404   **Transcriptomic data processing**

405   We first trimmed low-quality sequence data using the software btrim [58]. Filtered reads were

406   aligned to the chicken genome (galGal4.79, downloaded from

407   http://asia.ensembl.org/index.html) using TopHat2 (v2.0.14) [68], with the parameters

408   '--read-mismatches', '--read-edit-dist' and '--read-gap-length' set to no more than three bases.

409   We evaluated gene expression levels by applying HTSeq (v0.6.0) with the union exon model

410   and the whole gene model, coupled with the Cufflinks program available in the Cufflinks tool

411   suite (v2.2.1) [69] using default parameters.

412

413   **Correction and normalization**

414   To improve the analyses, genes were filtered for expression in the three datasets: in at least 80

415   percent of the Piao or control samples, gene counts from both HTSeq models (union exon and

416   whole gene) were no less than ten, while lower bound Fragments Per Kilobase of exon model

417   per Million mapped fragments (FPKM) values from Cufflinks were greater than zero. We then

418   performed normalization for gene length and GC content using the 'cqn' R package (v1.16.0)

419   [70], based on the filtered count matrix of the HTSeq union exon model. The output values

420   were defined as $\log_2$ (Normalized FPKM). The normalized matrix with genes kept in all three

421   datasets was adjusted for unwanted biological and technical covariates, like development days,

422   breeds, and sequencing lanes, via a linear mixed-effects model as previously described [71].

423   In detail, we calculated coefficients for these covariates with a linear model and then removed

424   the variability contributing to them from the original $\log_2$ (Normalized FPKM) values. For

425   example, when adjusting for development days, the number 1, 2 and 3 were used to replace

426   D7, D8 and D9, respectively. We then calculated a coefficient for each gene using the "lm"

427   function in R, with the number substitutes as a covariate. We removed the product of the

428   coefficient and the number substitute from the $\log_2$ (Normalized FPKM) value to obtain the

429   adjusted value. The adjusted data was then used for co-expression network construction.

430

Differential expression analysis

431

432   We applied three methods to identify DEGs. First, 826 DEGs (FDR < 0.05) were identified by

433   the Cuffdiff program in the Cufflinks tool suite with default parameters, using the bam files

434   from TopHat2. Second, 1451 DEGs (FDR < 0.05) were found by DESeq2 [72] based on the

435   read count data from the HTSeq union exon model. Third, 1244 DEGs (FDR < 0.05) were

436   obtained with a linear model, where we used the $\log_2$ (Normalized FPKM) matrix from the

437   normalization step and treated development stages, chicken breeds and sequencing lanes as

438   covariates. In total, 437 DEGs found by all three methods were used as the final DEGs.

439

Gene co-expression network analysis

440

441   To unravel underlying functional processes and genes associated with tail development, we

442   carried out WGCNA in the R package [40] with a one-step automatic and 'signed' network

443   type. The soft thresholding power option was set to 12 based on the scale-free topology model,

444   where topology fit index $R^{\wedge}2$ was first greater than 0.8. The minimum module size was

445   limited to 30. A height cut of 0.25 was chosen to merge highly co-expressed modules (*i.e.*,

446   correlation greater than 0.75). Finally, we obtained a total of 12 modules. M0 consisted of

447   genes that were not included in any other modules, and thus was excluded from further

448   analyses. We performed Pearson correlation to assess module relationships to the rumpless

449   trait, and defined *P* value < 0.05 as a significant threshold.

450

Hub genes and network visualization

451

452   In general, genes, which have significant correlations to others and the targeted trait, are the

453   most biologically meaningful and thus defined as 'hub genes'. Here, we referred to module

454   eigengenes (MEs) as 'hub genes' dependent on high intramodular connectivity, and absolute

455   values of gene significance (GS) and module membership (kME) greater than 0.2 and 0.8,

456   respectively. GS values reflect tight connections between genes and the targeted trait, while

457   kME mirrors eigengene-based connectivity between a gene expression profile and ME, and is

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

465

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

472

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

478

479 **Competing interests**
480 The authors declare that they have no competing interests.

481

482 **Acknowledgments**

489

490 **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,

535   effects and validation of polymorphisms within the *NCAPG - LCORL* locus located on BTA6 with feed intake,

536   gain, meat and carcass traits in beef cattle. BMC Genet 2011;12:103.

537   [23] Fantuzzi G, Puren AJ, Harding MW, Livingston DJ, Dinarello CA. Interleukin-18 regulation of interferon γ

538   production and cell proliferation as shown in interleukin-1β–converting enzyme (caspase-1)-deficient mice.

539   Blood 1998;91:2118−25.

540   [24] Reddy P, Teshima T, Kukuruga M, Ordemann R, Liu C, Lowler K, et al. Interleukin-18 regulates acute

541   graft-versus-host disease by enhancing *Fas*-mediated donor T cell apoptosis. J Exp Med 2001;194:1433−40.

542   [25] Chandrasekar B, Vemula K, Surabhi RM, Li-Weber M, Owen-Schaub LB, Jensen LE, et al. Activation of

543   intrinsic and extrinsic proapoptotic signaling pathways in interleukin-18-mediated human cardiac endothelial cell

544   death. J Biol Chem 2004;279:20221−33.

545   [26] Raggatt LJ, Qin L, Tamasi J, Jefcoat SC, Shimizu E, Selvamurugan N, et al. Interleukin-18 is regulated by

546   parathyroid hormone and is required for its bone anabolic actions. J Biol Chem 2008;283:6790−8.

547   [27] Dai S-M, Nishioka K, Yudoh K. Interleukin (IL) 18 stimulates osteoclast formation through synovial T cells

548   in rheumatoid arthritis: comparison with IL1β and tumour necrosis factor α. Ann Rheum Dis 2004;63:1379−86.

549   [28] Correa RG, Tergaonkar V, Ng JK, Dubova I, Izpisua-Belmonte JC, Verma IM. Characterization of NF-κB/Iκ

550   B proteins in zebra fish and their involvement in notochord development. Mol Cell Biol 2004;24:5257−68.

551   [29] Bushdid PB, Brantley DM, Yull FE, Blaeuer GL, Hoffman LH, Niswander L, et al. Inhibition of NF-κB

552   activity results in disruption of the apical ectodermal ridge and aberrant limb morphogenesis. Nature

553   1998;392:615−8.

554   [30] Schoenwolf GC. Histological and ultrastructural observations of tail bud formation in the chick embryo.

555   Anat Rec 1979;193:131−47.

556   [31] Hamburger V, Hamilton HL. A series of normal stages in the development of the chick embryo. J Morphol

557   1951;88:49−92.

558   [32] Schoenwolf GC. Morphogenetic processes involved in the remodeling of the tail region of the chick embryo.

559   Anat Embryol 1981;162:183−97.

560   [33] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using

561   DAVID bioinformatics resources. Nat Protoc 2008;4:44−57.

562   [34] Govoni KE, Lee SK, Chadwick RB, Yu H, Kasukawa Y, Baylink DJ, et al. Whole genome microarray

563   analysis of growth hormone-induced gene expression in bone: T-box3, a novel transcription factor, regulates

564   osteoblast proliferation. Am J Physiol Endocrinol Metab 2006;291:E128−E36.

565   [35] Economides KD, Zeltser L, Capecchi MR. *Hoxb13* mutations cause overgrowth of caudal spinal cordand

566   tail vertebrae. Dev Biol 2003;256:317−30.

567   [36] Cohen-Haguenauer O, Barton PJ, Van Cong N, Cohen A, Masset M, Buckingham M, et al. Chromosomal

568   assignment of two myosin alkali light-chain genes encoding the ventricular/slow skeletal muscle isoform and the

569   atrial/fetal muscle isoform (*MYL3*, *MYL4*). Hum Genet 1989;81:278−82.

570   [37] Esposito T, Sampaolo S, Limongelli G, Varone A, Formicola D, Diodato D, et al. Digenic mutational

571   inheritance of the integrin alpha 7 and the myosin heavy chain 7B genes causes congenital myopathy with left

572   ventricular non-compact cardiomyopathy. Orphanet J Rare Dis 2013;8:91.

573   [38] Igwe JC, Gao Q, Kizivat T, Kao WW, Kalajzic I. Keratocan is expressed by osteoblasts and can modulate

574   osteogenic differentiation. Connect Tissue Res 2011;52:401−7.

575   [39] Peters H, Wilm B, Sakai N, Imai K, Maas R, Balling R. Pax1 and Pax9 synergistically regulate vertebral

576   column development. Development 1999;126:5399−408.

577   [40] Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC

578   bioinformatics 2008;9:559.

579   [41] Takeichi M. Cadherin cell adhesion receptors as a morphogenetic regulator. Science 1991;251:1451−5.

580   [42] Bordeaux MC, Forcet C, Granger L, Corset V, Bidaud C, Billaud M, et al. The RET proto-oncogene induces

581   apoptosis: a novel mechanism for Hirschsprung disease. The EMBO Journal 2000;19:4056−63.

582   [43] Jayachandran A, Anaka M, Prithviraj P, Hudson C, McKeown SJ, Lo P-H, et al. Thrombospondin 1

583   promotes an aggressive phenotype through epithelial-to-mesenchymal transition in human melanoma.

584   Oncotarget 2014;5:5782−97.

585   [44] Lawler PR, Lawler J. Molecular basis for the regulation of angiogenesis by thrombospondin-1 and -2. CSH

586   Perspect Med 2012;2:a006627.

587   [45] Jin F, Dong B, Georgiou J, Jiang Q, Zhang J, Bharioke A, et al. N-WASp is required for Schwann cell

588   cytoskeletal dynamics, normal myelin gene expression and peripheral nerve myelination. Development

589   2011;138:1329−37.

590   [46] Heisenberg C-P, Tada M, Rauch G-J, Saude L, Concha ML, Geisler R, et al. Silberblick/Wnt11 mediates

591   convergent extension movements during zebrafish gastrulation. Nature 2000;405:76−81.

592   [47] Tao Q, Yokota C, Puck H, Kofron M, Birsoy B, Yan D, et al. Maternal Wnt11 activates the canonical wnt

593   signaling pathway required for axis formation in *Xenopus* embryos. Cell 2005;120:857−71.

594   [48] Kadono T, Zhang XQ, Srinivasan S, Ishida H, Barry WH, Benjamin IJ. CRYAB and HSPB2 deficiency

595   increases myocyte mitochondrial permeability transition and mitochondrial calcium uptake. J Mol Cell Cardiol

596   2006;40:783−9.

597   [49] Pinz I, Robbins J, Rajasekaran NS, Benjamin IJ, Ingwall JS. Unmasking different mechanical and energetic

598   roles for the small heat shock proteins CryAB and HSPB2 using genetically modified mouse hearts. The FASEB

599   Journal 2008;22:84−92.

600   [50] Hu Z-L, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through

601   integrative developments of Animal QTLdb and CorrDB. Nucleic Acids Res 2018;47:D701−D10.

602   [51] Xie Q, Yang L. A investigation report on Puer Piao chicken resources (Chinese Language). Yunnan Journal

603   of Animal Science and Veterinary Medicine 2009:17−8.

604   [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

605   of Piao chicken based on mitochondrial DNA D-loop region sequence. Journal of Yunnan Agricultural

606   University (Natural Science) 2011;2:211−23.

607   [53] Zwilling E. The development of dominant rumplessness in chick embryos. Genetics 1942;27:641−56.

608   [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

609   in two domestic chickens. Genome Biol Evol 2013;5:1376−92.

610   [55] West B, Zhou B-X. Did chickens go North? New evidence for domestication. J Archaeo Sci

611   1988;15:515−33.

612   [56] Wang M-S, Huo Y-X, Li Y, Otecko NO, Su L-Y, Xu H-B, et al. Comparative population genomics reveals

613   genetic basis underlying body size of domestic chickens. J Mol Cell Biol 2016;8:542−52.

614   [57] Guo X, Li Y-Q, Wang M-S, Wang Z-B, Zhang Q, Shao Y, et al. A parallel mechanism underlying frizzle in

615   domestic chickens. J Mol Cell Biol 2018;10:589−91.

616   [58] Kong Y. Btrim: a fast, lightweight adapter and quality trimming program for next-generation sequencing

617   technologies. Genomics 2011;98:152−3.

618   [59] Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv preprint

619   arXiv:1303.3997 2013.

620   [60] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis

621   Toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res

622   2010;20:1297−303.

623 [61] Bruno WJ, Socci ND, Halpern AL. Weighted neighbor joining: a likelihood-based approach to
624 distance-based phylogeny reconstruction. Mol Biol Evol 2000;17:189−97.
625 [62] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: molecular evolutionary genetics analysis
626 version 6.0. Mol Biol Evol 2013;30:2725−9.
627 [63] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for
628 whole-genome association and population-based linkage analyses. Am J Hum GeneT 2007;81:559−75.
629 [64] Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J
630 Hum Genet 2011;88:76−82.
631 [65] Akey JM, Zhang G, Zhang K, Jin L, Shriver MD. Interrogating a high-density SNP map for signatures of
632 natural selection. Genome Res 2002;12:1805−14.
633 [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
634 functional constraint drives the evolution of vision during chicken domestication. Cell Res 2016;26:556−73.
635 [67] Zhao JH. gap: Genetic Analysis Package. J Stat Softw 2007;023.
636 [68] Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. TopHat2: accurate alignment of
637 transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 2013;14:R36.
638 [69] Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, et al. Differential gene and transcript
639 expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 2012;7:562−78.
640 [70] Hansen KD, Irizarry RA, Wu Z. Removing technical variability in RNA-seq data using conditional quantile
641 normalization. Biostatistics 2012;13:204−16.
642 [71] Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, et al. Genome-wide changes in
643 lncRNA, splicing, and regional gene expression patterns in autism. Nature 2016;540:423−7.
644 [72] Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with
645 DESeq2. Genome Biol 2014;15:550.
646 [73] Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment
647 for integrated models of biomolecular interaction networks. Genome Res 2003;13:2498−504.
648 [74] Wang Y, Song F, Zhu J, Zhang S, Yang Y, Chen T, et al. GSA: Genome Sequence Archive. Genom Proteom
649 Bioinf 2017;15:14−8.
650 [75] Big Data Center Members. Database resources of the BIG Data Center in 2018. Nucleic Acids Res
651 2018;46:D14−D20.
652
653

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 $\Delta Pi*100$ 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 $\Delta Pi*100$.

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 rumplessness. 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 rumplessness were marked with a red asterisk, while a grey asterisk labeled M0 to indicate that the module was

687 excluded from the further analyses. **B–G**. Network plots from top 50 hub genes of six

688 significant modules. Significant annotation categories of these modules from DAVID were

689 colored according to network dots. Dot sizes indicate gene connectivity to others in the

690 module.

691

692 **Figure 4   Proposed mechanisms underlying rumplessness in Piao chicken and**

693 **microscopic examination of chicken tail embryogenesis**

694 **A.** Rumplessness might be consciously or unconsciously selected along with the high-yield

695 traits of Piao chicken. Strong positive selection pressures on regulatory elements of some

696 candidate genes might lead to gene activity changes in the tail bud. These ectopic expressions

697 would destroy mesenchymal maintenance by impeding differentiation and proliferation of

698 mesenchymal stem cells in the tail bud, through multiple cell survival and differentiation

699 pathways. This could then disrupt tail formation and prevent later developmental processes of

700 the distal structures. **B.** The morphology of the posterior region during the fourth to fifth day

701 of embryo development in rumpless Piao chicken, Piao chicken with a normal tail, and

702 control chicken.

703

704 **Supplementary materials**

705 **Figure S1   Random sampling recovery ratio of the 112 strongly selected 40kb-sliding**

706 **window regions**

707 The top two bar plots show the recovery ratio for F_{ST} and ΔPi in 1000 random samples of 20

708 controls from the 98 control chickens. The blue dashed lines indicate a recovery ratio of 0.95.

709 Red asterisks show sliding window regions with a recovery ratio greater than 0.95 in both F_{ST}

710 and ΔPi, while the short black lines above the asterisks indicate candidate genes in the sliding

711 regions. The bottom two line charts display F_{ST} and ΔPi values based on the 98 control

712 samples. Dashed lines indicate the top 1% threshold of descending F_{ST} and ΔPi. The colored

713 fillers present sliding window regions in the selected regions mentioned in the main text. The

714 40kb-sliding window regions are presented based on their median sites in order of

715 chromosomal position.

716

717 **Figure S2   Expression levels of *ENSGALG00000013155* and chicken pictures**

718 **A.** FPKM values of *ENSGALG00000013155* in different chicken tissues and development

719 stages from the three NCBI projects (SRA Accessions: ERP003988, SRP007412 and

720 DRP000595) that were used in our previous work [16]. **B.** The pictures of male (M) and

721 female (F) individuals of Piao chicken, Gushi chicken, and Wuding chicken.

722

723 **Table S1 RNA and DNA sample information**

724

725 **Table S2 Strongly selected regions, highly differentiated SNPs and indels, as well as**

726 **QTL annotations**

727

728 **Table S3 Expression values of DEGs and their DAVID annotation**

729

730 **Table S4 Hub genes of modules significantly correlated to rumplessness**

23

A

IRX4
QTL, *TERT*, *ENSGALG00000013155*: growth, body weight, sexual behavior, egg fertilization, and fat deposition

IL-18, HSPB2, CRYAB
QTL: egg number, duration of broodiness, and broody frequency

Positive selection on rumplessness along with the high-yield traits

Gene activity changes in mesenchymal stem cells of the tail bud

IRX4 — mesenchymal stem cell differentiation

IL-18 — Fas etc mediated stem cell apoptosis / osteoblast/osteoclast formation balance / NF-κB-related epithelial-mesenchymal transition

HSPB2 — energy balance and ossification

CRYAB — stem cell apoptosis and ossification

Destroy mesenchymal maintenance by impeding differentiation and proliferation of mesenchymal stem cells in the tail bud

MYL3, MYH7B etc. → Myogenesis

PAX9, PAX1 etc. → Vertebral column formation

KERA etc. → Osteogenesis

B

	D4	D5
Rumpless Piao		
Normal Piao		
Control		