Landscape and evolution of cis- and trans-regulatory divergence in chicken genome between two contrasted breeds analyzed in three tissues at one day age

CURRENT STATUS: ACCEPTED

Qiong Wang
Chinese Academy of Fishery Science Yellow Sea Fisheries Research Institute

Yaxiong Jia
Institute of Animal Science, Chinese Academy of Agricultural Sciences

Yuan Wang
Department of Animal Science and Technology, Qingdao Agricultural University, Qingdao

Zhihua Jiang
Department of Animal Sciences, Center for Reproductive Biology, Veterinary and Biomedical Research Building, Washington State University

Xiang Zhou
College of Animal Sciences and Veterinary Medicine, Huazhong Agricultural University, Wuhan

Zebin Zhang
College of Animal Science and Technology, China Agricultural University

Changsheng Nie
College of Animal Science and Technology, China Agricultural University

Junying Li
College of Animal Science and Technology, China Agricultural University

Ning Yang
College of Animal Science and Technology, China Agricultural University

Lujian Qu
China Agricultural University

quluj@163.com Corresponding Author
Abstract

Background: Gene expression variation is an important mechanism underlying phenotypic variation, and can occur via cis- and trans-regulation. In order to understand the role of cis- and trans-regulatory variation on population divergence of chicken, we developed reciprocal crosses of two chicken breeds, White Leghorn and Cornish Game, with major differences in body size and reproductive traits, and used them to identify the degree of cis versus trans variation in brain, liver and muscle of both male and female samples at 1 day age. Results: We provided a landscape about how the transcriptomes are regulated in the hybrid progenies of two contrasted breeds by allele specific expression analysis. Our results showed that compared with the cis-regulatory divergence, trans-acted genes existed more extensively in the chicken genome. Furthermore, a widespread tendency of compensatory regulation exists in chicken genome. Most importantly, we found the evidence of stronger purifying selection on genes regulated by trans variations than the cis elements. Conclusions: We demonstrated a pipeline to explore the allele-specific expression in the hybrid progenies of inbred lines without specific reference genome. Our research performed the first study to describe the regulatory divergence between two contrasted breeds. The results suggested that artificial selection associated with domestication in chicken may have more often acted on trans-regulatory divergence than cis.

Background

Gene expression is regulated by many transcriptional regulatory factors, which can be classified into cis-regulatory elements and trans-regulatory factors [1]. Cis-regulatory elements, such as promoters, enhancers and silencers, are regions of non-coding DNA which regulate the transcription of nearby genes. In contrast, trans-regulatory factors regulate (or modify) the expression of distant genes by combining with their target sequence [1, 2]. In most cases, gene expression is controlled by complex interactions of cis-regulatory sequences and trans-acting factors [3, 4].

Cis- and trans-regulation elements are thought to differ in important genetic and evolutionary properties [5, 6]. In diploid individuals, cis-regulatory elements regulate gene expression in an allele-specific manner. Heterozygotes for cis-regulatory variation express allelic imbalances at the
transcriptional and translational level. By comparison, trans-regulatory factors interact with target sequences to regulate both alleles [1]. Trans-regulatory divergence is enriched for dominant effect, while the effects of cis-regulatory variants are additivity [6, 7]. Advantageous cis-regulatory variants are more likely to enriched to fixation during the process of evolution, because the additive effects expose rare alleles to selection [5].

Both cis- and trans- regulatory variation are important in phenotypic variation [1, 8-10]. Previous work in a range of species, including Drosophila [7], Mouse [11, 12] and Coffea [13], have used allele specific expression (ASE) analysis [14] to differentiate cis- and trans-regulatory divergence (Table 1). However, regulatory divergence of gene expression in birds may be different from mammals, insects or plants, considering that some genetic mechanisms involved in allele specific expression are distinct in birds. For instance, genomic imprinting has been found in mammals and some plants [15-17], but appears to be largely absent in birds thus far assessed [18-20]. Dosage compensation exists in some diploid species to buffer the effect of copy number difference of genes on sex chromosome [21-23], but it was incomplete in birds [24-28]. Given this, we thought it is important to study the regulatory divergence in birds.

Chicken is a model animal for researches on birds, and a remarkable example of rapid phenotypic divergence, with artificial selection resulting in major size, behavioral and reproductive differences among breeds [29]. Previous studies have identified frequent allele specific expression between different chicken breeds [19, 20]. The rapid change under domestication offers a unique example to detect the relative importance of cis- versus trans-regulatory variation underlying phenotypic change. We used reciprocal crosses of White Leghorn (WL), a key layer breed selected for high egg output, and Cornish Game breeds (CG), a cornerstone broiler breed selected for rapid growth and muscle development [30], to assess the role of different forms of regulatory variation in brain, liver and muscle of both males and females at 1 day age.

Results

The profile of the parental genomes and gene expression in different tissues, sexes of progenies
The two inbred chicken strains, Cornish-Game and White-Leghorn, which exhibit major differences in growth rate, egg production and behavior, were used to generate purebred and reciprocal hybrid F1 progenies (Fig. 1). In order to identify breed-specific variants, we sequenced the four parents of the two reciprocal crosses, recovering on average 100.73 million pair-end reads per sample after quality control. We identified on average 4.74 million SNPs per parental genome which were used to generate simulated parental genomes. From these SNPs, we picked SNPs which were homozygous in each parental bird but different from each other in a same cross (heterozygous in the hybrid progenies), resulting in two heterozygous SNP lists with on average 1.4 million of heterozygous SNPs for the two reciprocal crosses, individually, to identify the allele-specific RNA-seq reads of offspring in the following steps.

For each hybrid cross, we collected RNA-Seq data from brain, liver and muscle of three male and three female F1 progenies 1-day post-hatch. On average, we recovered 29.17 million mappable reads per sample. In order to eliminate the effect of the sex chromosomes, we removed all Z and W genes from our analysis and focused entirely on autosomal loci. We recovered significant differences in gene expression between different tissues, sexes and parents of origin (Fig. 2). Tissue was the most powerful factor for gene expression. The sex played a leading role in brain, and the strain effected gene expression of liver the most, while in muscle, the parent of origin seems more powerful because samples were divided into two parts by mother origin. We therefore retained all three variables in our subsequent analysis. It resulted in 12 treatment groups, comprised of 3 tissues, 2 sexes and 2 reciprocal crosses in this study.

**An effective pipeline was applied on the allele-specific expression analysis**

To identify the parental origin of offspring’s mRNA, we tried a new pipeline by using a R package of ‘asSeq’ [31]. Briefly, a set of R scripts are available for genotype phasing with the 1.4 million of heterozygous SNPs identified in the previous step. About 2% of these SNPs located on the exon region. The large number of SNPs provide higher chance that an RNA-seq read could overlap with a heterozygous genetic marker, hence is identified as allele-specific read.

In order to validate the accuracy of our ASE pipeline, we generated two artificial hybrid F1 libraries.
Specifically, we concatenated two male brain RNA-seq fastq files from cross 1 and cross 4, which had roughly equal read depth. We also concatenated two female liver samples the same way. The two simulated hybrid libraries and four original purebred libraries were handled consistently with other hybrid libraries, using the heterozygous SNP lists of both cross 2 and cross 3. We compared the expression ratio of two simulated alleles (CG/WL) to the real expression ratio of two samples (CG/WL) for each gene. A strong correlation between the two measurements were observed (Fig. S1), indicating that our ASE analysis pipeline was robust. Since our pipeline only counted the local reads containing the heterozygous SNPs, we further assessed the correlation of expression fold change (CG/WL) between the local reads method and the method of counting total reads using edgeR [32-34]. The correlation was also strong (Fig. S2). These results proved the feasibility of our pipeline.

**Genes were classified into different categories by the type of regulatory divergency**

A total of 24,881 genes from the Ensembl v87 annotation were analyzed. About one fifth of them contained heterozygous SNPs and expressed in our progeny samples (Table S1). For the genes containing heterozygous SNPs, we observed significant expression difference (p-value <0.05, binomial test corrected for multiple comparisons by q-value method) between the purebred females (cross 1 vs. cross 4) for 14.71 % of them in brain, 36.45 % in liver and 38.38 % in muscle (take the heterozygous SNP list of cross 2 for example). In males, we observed that 17.64 % of genes in brain, 41.87 % in liver and 37.84 % in muscle expressed significantly different (Table S1).

Expressive genes were classified into different categories by the type of regulatory divergency [7, 35, 36] (Fig. 3A, B, Table 1, Fig. S3-S5). Most genes showed conserved or ambiguous expression, as expected given the recent divergence time of these two breeds. There were more than 70% genes in brain, more than 40% genes in liver, and around 50% genes in muscle classified as conserved. Nonetheless, we observed substantial cis- and trans- variation in the hybrid crosses. There were a greater proportion of trans-regulated gene expression variations than cis in most tissues and across both sexes, particularly in muscle (Fig. 3C).

Genes regulated by both cis- and trans-regulatory variations were divided into four categories: “cis+trans (same)”, “cis+trans (opposite)”, “cis×trans” and “compensatory”. Genes classified as
“cis+trans (same)” show evidence of cis and trans variations acting in the same direction, while genes classified into the other three categories show evidence of cis and trans variations acting in opposite directions, with different expression preference on the two alleles. We observed that genes more often display the latter pattern, and most genes were classified as “compensatory” (Fig. 3C). The gene proportion of each regulatory category was similar between different tissues and different sexes, except for some differences between the muscle and other two tissues (Fisher’s exact test, Table S2). Unexpectedly, we observed only few loci with consistent cis- or trans-regulatory divergence across different groups (Fig. S6). These stable cis- or trans-regulatory divergence genes appear to play important roles in phenotypic divergence. For example, the gene IGFBP2, TGFBI, PDGFRL and IGF2R all showed significant evidence of expression bias between our two breeds. These genes are related to the chicken growth, which could explain the difference of growth rate between two chicken breeds (Table S3).

**Genes regulated by trans-acting variation show greater sequence conservation**

We counted the number of variants located within 1 kb upstream of transcription start sites of each gene with the genome data of the four parents. The results showed greater variations at upstream of cis-regulatory divergence genes than trans-acted genes in all samples (Fig. 4A). The ratio of the number of non-synonymous SNPs to the number of synonymous SNPs (pN/pS) in coding sequence of each gene was calculated in this research. We found that the value of pN/pS in genes regulated by trans-variant were lower than cis-regulatory divergence genes in all samples (Fig. 4B, Fig. S7-S8).

**Discussion**

For the previous studies on regulatory divergence, the time point they chose were not identical, from embryo to adult [7, 11, 12]. Genes express differentially among different developmental stages [37], therefore different result of the regulatory divergence would be obtained. We chose the 1-day age because it is a critical stage for the chicken development, when they make a transition from embryo to growth stage, and genes responsible for growth and immunity begin to express [38, 39].
On account of the short divergence time, our two chicken inbred strains are not like the mouse inbred lines with high consistency on genome. To enhance the reliability of our results, we have improved our analysis pipeline. Firstly, the SNP list we used to identify the parental origin was strictly filtered from the re-sequencing data of the four parents. The SNPs were statistically homozygous in each parent, hence heterozygous in each hybrid offspring. Secondly, we counted the total number of reads covering at least one SNP marker across the whole transcript, instead of counting the reads number of each SNP. Compared with the method using the existing strain-specific reference genomes, our pipeline could improve the accuracy of parental origin identification for the heterozygous SNPs in hybrid offspring, because we sequenced their parents directly. These SNPs were used to mark the parental origin for the alleles of each gene, and subsequently increased the accuracy of regulation classification. But it also resulted in the limited number of genes could be studied in this research. However, our study provided an example to deal with the similar situation where there was no specific reference genome for different strains.

Although chicken domestication occurred several thousand years ago, commercial populations were established only in the last 200 years [29]. In our study, most genes showed conserved or ambiguous expression, and more trans-regulatory variants than cis, which could attribute to the short differentiation time between White Leghorn and Cornish Game. In theory, the pleiotropic effects of trans-regulatory mutations will result in selection to eliminate most deleterious trans-acting mutations [40]. In contrast, we might expect a large proportion of cis-regulatory mutations to be largely neutral, and therefore to accumulate over time [9, 41]. The large proportion of trans-regulatory mutations observe in our study suggest that artificial selection has primarily acted on trans-regulatory mutations, but the neutral cis-regulatory mutations have not accumulated substantially in the short time since these breeds were established.

Genes regulated by both cis- and trans-variations more often acting in opposite directions, and most genes were classified into “compensatory” in our study. This result consistent with the previous research on house mice [36], indicating that the cis- and trans-variants tend to act convergently to maintain the stability of gene expression [11, 42]. Despite lacking of complete dosage compensation
mechanism on sex chromosome [24-28], an extensive compensatory tendency still exist in chicken genome.

There were few loci with consistent cis- or trans-regulatory variation across different tissues and different sexes. This result consistent with some previous ASE analyses, which suggest rare ASE genes are consistently expressed across tissues [43, 44]. However, the cis- and trans-regulatory divergence classification is more complex than the simple ASE analysis. Gene expression is characterized by spatiotemporal specificity. It always controlled by the interaction of cis-regulatory DNA sequences and trans-regulatory factors, which could complicate the identification of regulatory divergence. The statistical methods could not accurately classify them with the limited expression information. However, even so, the statistical result still referable and valuable for the following analysis.

Cis-regulatory elements are primarily located at upstream of coding sequences. Our results, is consistent with resent studies in Drosophila [7], detected greater variants within 1 kb upstream of transcription start sites of cis-regulatory divergence genes than trans-acted genes, suggesting that our classification results were reliable. Meanwhile, genes regulated by trans-variant showed a lower value of pN/pS than cis-acted genes. The pN/pS was usually used to assess the degree of selective constraint. Genes under high constraint were expected to have lower pN/pS [45, 46]. Our results suggest that trans-regulatory divergence genes are subject to increased selective constraint during chicken domestication and may have been under stronger artificial selection. It is consistent with similar studies in mice [11], which pointed that trans-regulated genes display greater sequence conservation by computing the GERP (Genomic Evolutionary Profiling) score for each exon.

Conclusions
In this study, we demonstrated a pipeline to explore the allele-specific expression in the hybrid progenies of inbred lines without specific reference genome. With the genome sequence of parents and RNA-seq data of offspring, we classified the expressed genes on chicken genome into different categories by the type of regulatory divergency. More trans-regulatory divergence than cis was detected due to the short divergence history of the two parental breeds. A widespread compensatory
tendency for the regulatory divergence exists in chicken genome. The sequence conservation analysis suggested that artificial selection associated with domestication may have more often acted on genes regulated by trans-variations in the history of chicken commercial breeds establishment.

**Methods**

**Samples**

The inbred chickens used in our study were derived from National Engineering Laboratory for Animal Breeding of China Agricultural University. All the animals were fed and handled according to the regulations and guidelines established by the Animal Care and Use Committee of China Agricultural University, and all efforts were made to minimize suffering. DNA was extracted from brachial vein blood of four parents of two reciprocal crosses, using phenol-chloroform protocols with the standard protocols. Total RNA was extract from three tissues (brain, liver and breast muscle) of 23 chickens at the age of 1 day using the Trizol reagent (Invitrogen, Carlsbad, CA, USA) with the standard protocols from the manufacturer. The DNA and RNA quality were assessed by NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific Inc.) and agarose gel electrophoresis (AGE).

**DNA & RNA sequencing and data alignment**

Whole-genome sequencing of parent genomes and RNA-seq of offspring were performed using an Illumina HiSeq 2500 sequencer. Library construction and sequencing were performed following manufacturer protocols (TruSeq DNA Sample Prep Kit, TruSeq RNA Sample Prep Kit, TruSeq PE Cluster Kit v3-cBot, and TruSeq SBS Kit v3). Both DNA and RNA were sequenced with paired-end 100bp reads with a 300bp insert. All sequencing data was filtered by NGS QC Toolkit (v2.3) [47] with default parameters.

To ensure the accuracy of RNA-seq data alignment, we simulated four parental genomes. The re-sequencing data of four parents were mapped to the chicken reference genome (Gallus_gallus-5.0, http://hgdownload.soe.ucsc.edu/downloads.html#chicken) with the Burrows-Wheeler Aligner (BWA) [48] (v0.7.15). The BAM files were sorted, and duplicate reads were removed using Picard toolkit (https://github.com/broadinstitute/picard). The Genome Analysis Toolkit (GATK) [49] (v3.6) was used for single-nucleotide polymorphism (SNP) calling. Nucleotide from the reference genome was
substituted if the mutant base was supported by more reads than the original reference base, this step was completed by vcftools [50] (v0.1.13). The four simulated parental genomes were used to replace the reference genome in the RNA-seq data alignment of hybrid crosses. For each hybrid cross, we then identified SNPs between two parents that were homozygous in each parent with >10 supporting reads from re-sequencing data. The SNP list divided each hybrid offspring genome into two parts basing on the parent-of-origin.

The RNA-seq data alignment was completed by STAR [51] (v2.5.3a). With the SNP list between two parents, we counted the allele-specific reads from the two parts of each hybrid offspring on exon set level, using the R package of ‘asSeq’ [31]. To be specific, we counted the total number of reads covering at least one SNP across the whole exon set. For the condition of one read containing more than one SNPs, we set the parameter of prop.cut as 0.9, that is, we assigned a read to one of the two parental alleles if the proportion of those heterozygous SNPs suggesting the read is from that allele is larger than 0.9. In practice, this condition could insure all the SNPs on one read were consistent. If not, they would be discarded. We then collapsed counts at the exon level into gene level according to the Ensembl gene annotation file (ftp://ftp.ensembl.org/pub/release-91/gtf/gallus_gallus). We filtered the expressed genes with the following standard: for each sex and each tissue, the total reads of both the three purebred offspring and three hybrid offspring should be in the limits of 6 to 1000. The reads count of each sample was further normalized based on the sum of reads which could be mapped to the whole genome.

One male muscle sample of cross 3 was removed because its expression pattern was abnormal. We speculated that it might be mixed from other cross by error.

**Classification of different regulatory categories**

In order to categorize regulatory variation, we referenced the methods of regulatory divergence research on *Drosophila* [7] and house mouse [36]. Concretely, the binomial test was used to identify differential expression both between the two purebred progenies (P) and between the two alleles of hybrid progenies (H). Fisher’s exact test was used to judge the breed-specific RNA abundance ratio difference between the P and H data sets to detect the trans effects (T). For false discovery rate
control, a method of q-value estimation [52] was adopted to correct the p-value of both binomial test and Fisher’s exact test. We considered it has significantly difference when q < 0.05. The expressed genes were classified into eight categories according to the following criterion:

(1) Cis: Significant difference in P and H, no significant difference in T.
(2) Trans: Significant difference in P, but not H, significant difference in T.
(3) Cis + trans (same): significant difference in P, H and T, the log2-transformed strain-specific ratios in P and H have the same sign, and the difference of P higher than H.
(4) Cis + trans (opposite): significant difference in P, H and T, the log2-transformed strain-specific ratios in P and H have the same sign, and the difference of H higher than P.
(5) cis × trans: significant difference in P, H and T, and the log2-transformed strain-specific ratios in P and H have the opposite sign.
(6) compensatory: Significant difference in H, but not P, significant difference in T.
(7) Conserved: No significant difference in H, P and T.
(8) Ambiguous: All other patterns.

**Sequence conservation analysis**

Re-sequencing data of four parents were used to study the sequence conservation of cis- and trans-regulatory divergence genes. The ratio of pN/pS of coding sequence and the number of variants in the upstream of 1 kb range from transcription start site were calculated with the result of SNP annotation using SnpEff [53]. Nonsynonymous mutation contains missense variant, start codon lost, start codon gained, stop codon lost and stop codon gained. Synonymous mutation refers to the variant in the coding region causing a codon that produces the same amino acid.

**Abbreviations**

ASE: allele specific expression; CG: Cornish Game; WL: White Leghorn

**Declarations**

**Ethics approval and consent to participate**

Our animal experiments were approved by the Animal Care and Use Committee of China Agricultural University (Approval ID: XXCB-20090209). All the animals were fed and handled according to the
regulations and guidelines established by this committee, and all efforts were made to minimize suffering.

**Consent for publication**

Not applicable.

**Availability of data and materials**

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

**Competing interests**

The author NY is a member of the editorial board (Associate Editor) of this journal.

**Funding**

This work was supported by the Beijing Innovation Team of the Modern Agro-industry Technology Research System (BAIC04-2016, BAIC04-2017). The funding body did not exert influence on the design of the study, and collection, analysis, and interpretation of data or in writing of the manuscript.

**Authors' contributions**

QW analyzed and interpreted the sequencing data and drafted the manuscript. YJ participated in performing reciprocal cross experiment and collecting the samples and improved the manuscript. YW performed DNA & RNA extraction and preparation for sequencing, and improved the manuscript. ZJ and XZ provided some suggestion to improve our study. ZZ and CN participated in sample collection. JL provided experimental animal and help for performing reciprocal cross experiment. NY participated in the design of the study. LQ conceived of the study, and participated in its design and coordination. All authors read and approved the final manuscript.

**Acknowledgements**

We thank our colleagues of China Agricultural University for assistance on sample collection, and Professor Judith Mank for helpful comments on the manuscript.

**References**

1. Wittkopp PJ, Haerum BK, Clark AG: Evolutionary changes in cis and trans gene regulation. Nature 2004, 430(6995):85-88.
2. Cowles CR, Hirschhorn JN, Altshuler D, Lander ES: Detection of regulatory variation in mouse genes. Nat Genet 2002, 32(3):432-437.

3. Gibson G, Weir B: The quantitative genetics of transcription. Trends Genet 2005, 21(11):616-623.

4. Wittkopp PJ: Genomic sources of regulatory variation in cis and in trans. Cell Mol Life Sci 2005, 62(16):1779-1783.

5. Wray GA: The evolutionary significance of cis-regulatory mutations. Nat Rev Genet 2007, 8(3):206-216.

6. Meiklejohn CD, Coolon JD, Hartl DL, Wittkopp PJ: The roles of cis- and trans-regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene expression. Genome Res 2014, 24(1):84-95.

7. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ: Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res 2010, 20(6):816-825.

8. Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB: Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature 2005, 433(7025):481-487.

9. Wittkopp PJ, Haerum BK, Clark AG: Regulatory changes underlying expression differences within and between Drosophila species. Nat Genet 2008, 40(3):346-350.

10. Tian L, Khan A, Ning Z, Yuan K, Zhang C, Lou H, Yuan Y, Xu S: Genome-wide comparison of allele-specific gene expression between African and European populations. Hum Mol Genet 2018, 27(6):1067-1077.

11. Goncalves A, Leigh-Brown S, Thybert D, Stefflova K, Turro E, Flicek P, Brazma A, Odom DT, Marioni JC: Extensive compensatory cis-trans regulation in the evolution of mouse gene expression. Genome Res 2012, 22(12):2376-2384.
12. Crowley JJ, Zhabotynsky V, Sun W, Huang S, Pakatci IK, Kim Y, Wang JR, Morgan AP, Calaway JD, Aylor DL et al: Analyses of allele-specific gene expression in highly divergent mouse crosses identifies pervasive allelic imbalance. Nat Genet 2015, 47(4):353-360.

13. Combes MC, Hueber Y, Dereeper A, Rialle S, Herrera JC, Lashermes P: Regulatory divergence between parental alleles determines gene expression patterns in hybrids. Genome Biol Evol 2015, 7(4):1110-1121.

14. Pirinen M, Lappalainen T, Zaitlen NA, Consortium GT, Dermitzakis ET, Donnelly P, McCarthy MI, Rivas MA: Assessing allele-specific expression across multiple tissues from RNA-seq read data. Bioinformatics 2015, 31(15):2497-2504.

15. Bartolomei MS, Zemel S, Tilghman SM: Parental imprinting of the mouse H19 gene. Nature 1991, 351(6322):153-155.

16. DeChiara TM, Robertson EJ, Efstratiadis A: Parental imprinting of the mouse insulin-like growth factor II gene. Cell 1991, 64(4):849-859.

17. Zhang M, Zhao H, Xie S, Chen J, Xu Y, Wang K, Zhao H, Guan H, Hu X, Jiao Y et al: Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. Proc Natl Acad Sci U S A 2011, 108(50):20042-20047.

18. Fresard L, Leroux S, Servin B, Gourichon D, Dehais P, Cristobal MS, Marsaud N, Vignoles F, Bed'hom B, Coville JL et al: Transcriptome-wide investigation of genomic imprinting in chicken. Nucleic Acids Res 2014.

19. Wang Q, Li K, Zhang D, Li J, Xu G, Zheng J, Yang N, Qu L: Next-Generation Sequencing Techniques Reveal that Genomic Imprinting Is Absent in Day-Old Gallus gallus domesticus Brains. PLoS One 2015, 10(7):e0132345.

20. Zhuo Z, Lamont SJ, Abasht B: RNA-Seq Analyses Identify Frequent Allele Specific Expression and No Evidence of Genomic Imprinting in Specific Embryonic Tissues of
Chicken. Sci Rep 2017, 7(1):11944.

21. Lyon MF: X-chromosome inactivation. Curr Biol 1999, 9(7):R235-237.

22. Nguyen DK, Disteche CM: Dosage compensation of the active X chromosome in mammals. Nat Genet 2006, 38(1):47-53.

23. Julien P, Brawand D, Soumillon M, Necsulea A, Liechti A, Schutz F, Daish T, Grutzner F, Kaessmann H: Mechanisms and evolutionary patterns of mammalian and avian dosage compensation. PLoS Biol 2012, 10(5):e1001328.

24. Ellegren H, Hultin-Rosenberg L, Brunstrom B, Dencker L, Kultima K, Scholz B: Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. BMC Biol 2007, 5:40.

25. Mank JE, Ellegren H: All dosage compensation is local: gene-by-gene regulation of sex-biased expression on the chicken Z chromosome. Heredity (Edinb) 2009, 102(3):312-320.

26. Zimmer F, Harrison PW, Dessimoz C, Mank JE: Compensation of dosage-sensitive genes on the chicken Z chromosome. Genome Biol Evol 2016.

27. Barlow DP, Stoger R, Herrmann BG, Saito K, Schweifer N: The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature 1991, 349(6304):84-87.

28. Wang Q, Mank JE, Li J, Yang N, Qu L: Allele-Specific Expression Analysis Does Not Support Sex Chromosome Inactivation on the Chicken Z Chromosome. Genome Biol Evol 2017, 9(3):619-626.

29. Rubin CJ, Zody MC, Eriksson J, Meadows JR, Sherwood E, Webster MT, Jiang L, Ingman M, Sharpe T, Ka S et al: Whole-genome resequencing reveals loci under selection during chicken domestication. Nature 2010, 464(7288):587-591.

30. Association AP: The American standard of perfection, illustrated. A complete
description of all recognized breeds and varieties of domestic poultry. American Poultry Association 2001.

31. Sun W: A statistical framework for eQTL mapping using RNA-seq data. Biometrics 2012, 68(1):1-11.

32. Robinson MD, Smyth GK: Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics 2008, 9(2):321-332.

33. Robinson MD, Smyth GK: Moderated statistical tests for assessing differences in tag abundance. Bioinformatics 2007, 23(21):2881-2887.

34. Lun AT, Chen Y, Smyth GK: It's DE-licious: A Recipe for Differential Expression Analyses of RNA-seq Experiments Using Quasi-Likelihood Methods in edgeR. Methods Mol Biol 2016, 1418:391-416.

35. Landry CR, Wittkopp PJ, Taubes CH, Ranz JM, Clark AG, Hartl DL: Compensatory cis-trans evolution and the dysregulation of gene expression in interspecific hybrids of Drosophila. Genetics 2005, 171(4):1813-1822.

36. Mack KL, Campbell P, Nachman MW: Gene regulation and speciation in house mice. Genome Res 2016, 26(4):451-461.

37. Xu Z, Che T, Li F, Tian K, Zhu Q, Mishra SK, Dai Y, Li M, Li D: The temporal expression patterns of brain transcriptome during chicken development and ageing. BMC Genomics 2018, 19(1):917.

38. Dushyanth K, Bhattacharya TK, Shukla R, Chatterjee RN, Sitaramamma T, Paswan C, Guru Vishnu P: Gene Expression and Polymorphism of Myostatin Gene and its Association with Growth Traits in Chicken. Anim Biotechnol 2016, 27(4):269-277.

39. Schokker D, Hoekman AJ, Smits MA, Rebel JM: Gene expression patterns associated with chicken jejunal development. Dev Comp Immunol 2009, 33(11):1156-1164.

40. Denver DR, Morris K, Streelman JT, Kim SK, Lynch M, Thomas WK: The transcriptional
consequences of mutation and natural selection in Caenorhabditis elegans. Nat Genet 2005, 37(5):544-548.

41. Lemos B, Araripe LO, Fontanillas P, Hartl DL: Dominance and the evolutionary accumulation of cis- and trans-effects on gene expression. Proc Natl Acad Sci U S A 2008, 105(38):14471-14476.

42. Coolon JD, McManus CJ, Stevenson KR, Graveley BR, Wittkopp PJ: Tempo and mode of regulatory evolution in Drosophila. Genome Res 2014, 24(5):797-808.

43. Chamberlain AJ, Vander Jagt CJ, Hayes BJ, Khansefid M, Marett LC, Millen CA, Nguyen TT, Goddard ME: Extensive variation between tissues in allele specific expression in an outbred mammal. BMC Genomics 2015, 16:993.

44. Kukurba KR, Zhang R, Li X, Smith KS, Knowles DA, How Tan M, Piskol R, Lek M, Snyder M, Macarthur DG et al: Allelic expression of deleterious protein-coding variants across human tissues. PLoS Genet 2014, 10(5):e1004304.

45. Bjornerfeldt S, Webster MT, Vila C: Relaxation of selective constraint on dog mitochondrial DNA following domestication. Genome Res 2006, 16(8):990-994.

46. Wang Z, Yonezawa T, Liu B, Ma T, Shen X, Su J, Guo S, Hasegawa M, Liu J: Domestication relaxed selective constraints on the yak mitochondrial genome. Mol Biol Evol 2011, 28(5):1553-1556.

47. Patel RK, Jain M: NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 2012, 7(2):e30619.

48. Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010, 26(5):589-595.

49. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M et al: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010,
Table

Table 1 Studies that have classified the gene regulatory divergence on genome
| Species  | Tissue | Sex    | Cis     | Trans   | Cis and trans | Conserved and ambiguous | Method                      |
|----------|--------|--------|---------|---------|--------------|-------------------------|-----------------------------|
| Drosophila | Whole fly | Female | 12.4%   | 30%     | 35%          | 22.6%                   | Hierarchical statistical analyses |
| Mouse | Liver | Male | 14% | 0.6% | 17.4% | 68% | Maximum likelihood based approach |
| Mouse | Testis | Male | 24% | 9% | 44% | 23% | Hierarchical statistical analyses |
| Coffea¹ | Leaf | Male | 15.5% | 18.5% | 17.5% | 48.0% | Hierarchical statistical analyses |
| Coffee¹ | Leaf | Male | 14.5% | 18.3% | 16.6% | 50.6% | Hierarchical statistical analyses |
| Chicken² | Brain | Female | 3.45% | 3.70% | 4.88% | 87.99% | Hierarchical statistical analyses |
| Chicken² | Brain | Male | 3.75% | 4.86% | 4.37% | 87.01% | Hierarchical statistical analyses |
| Chicken² | Liver | Female | 7.41% | 12.92% | 16.15% | 63.53% | Hierarchical statistical analyses |
| Chicken² | Liver | Male | 8.31% | 13.93% | 17.07% | 60.70% | Hierarchical statistical analyses |
| Chicken² | Muscle | Female | 5.60% | 15.80% | 10.79% | 67.82% | Hierarchical statistical analyses |
| Chicken² | Muscle | Male | 4.72% | 16.73% | 11.58% | 66.99% | Hierarchical statistical analyses |

¹ This article contains two crosses (cross C×E and cross E×C).

² This study contains two crosses (cross 2 and cross 3), and we took the cross 2 as an example.

Supplementary Legends

Additional file 1.docx

Supplemental figures:

Figure S1 Assessment of our analysis pipeline for estimating allele specific expression.

Figure S2 Assessment of our local reads method for estimating allele specific expression.

Figure S3 Classification of genes in brain.

Figure S4 Classification of genes in liver.

Figure S5 Classification of genes in muscle.

Figure S6 Intersection of different groups of cis- and trans- regulatory genes.

Figure S7 The ratio of the numbers of non-synonymous SNPs to the numbers of synonymous SNPs (pN/pS) in different groups of cross 2.

Figure S8 The ratio of the numbers of non-synonymous SNPs to the numbers of synonymous SNPs
Supplemental tables:

Table S1 The summary of differential expression genes in hybrid and purebred progenies
Table S2 The difference of gene proportion of each categories between different groups
Table S3 The gene list of intersection of each group

Figures

Cross design. Cornish-Game (CG) and White-Lehorns (WL) were used to generate purebred and hybrid progenies. There were four crosses, Cross 1: CG × CG, cross 2: CG × WL, cross 3: WL × CG, cross 4: WL × WL (female parent was listed first).
Figure 2

Principal Component Analysis of RNA-seq data. Each point represents one sample, with shape indicating sex, color indicating tissue (All) or cross (Brain, Liver, Muscle). In this step, information of genes on Z chromosome has been excluded.
Figure 3

Classification of genes according to the expression pattern of purebred and hybrid data sets. We take the group of brain male (A) and brain female (B) of cross 2 for example (the other groups see Supplemental Material), each point represents a single gene and is color-coded according to its regulatory category. The coordinate position shows the average log2 expression fold change between the alleles in the hybrids (y-axis) and between the two purebreds (x-axis). The proportion of each category was summarized in the bar graph (C), where we removed the conserved and ambiguous genes, and further subdivided the genes of cis+trans category into two categories, basing on the same direction or opposition direction did the cis and trans variants act in. The number above the bar represents the proportion of genes in this regulatory category, and number on the bar represents the gene count of this category.
Figure 4

Sequence conservation analysis of the cis- and trans-regulatory divergence genes. (A) The probability density (y-axis) of variation count (x-axis) located in the 1 kb of DNA sequence upstream of each gene’s transcription start site. The number followed the regulatory category name in the legend refers to the mean value of variation count of all genes in this category. The p-value above the legend obtained by Mann-Whitney U test. (B) The value of pN/pS in cis- and trans-regulatory divergence genes. The y-axis refers to the mean value of all genes in the category. Significance of the difference between the two regulatory categories is labeled above the bar (* p<0.05, t-test; ** p<0.01, t-test).

Supplementary Files
This is a list of supplementary files associated with this preprint. Click to download.
Additional file 1.pdf
