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

CURRENT STATUS: ACCEPTED

Qiong Wang
Republic of South Africa Department of Environmental Affairs Oceans and Coasts

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

Lujiang Qu quluj@163.com
China Agricultural University

Corresponding Author

DOI: 10.21203/rs.2.10726/v1
SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
cis; trans; regulation; RNA-seq; allele-specific expression
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 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 between two contrasted breeds by allele specific expression analysis. Our results showed that compared with the cis-regulated genes, trans-acted genes existed more extensively in the chicken genome. Furthermore, a widespread compensatory tendency exists in chicken genome. Most importantly, we found the evidence of stronger purifying selection for trans-regulatory variation than the cis-elements.

Conclusions

Our research performed the first study to describe the transcriptome regulation between White Leghorn and Cornish Game breeds and suggests that artificial selection associated with domestication in chicken may have more often acted on trans-regulatory variation.

Background

Gene expression levels are 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 [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 expression in an allele-specific manner, and 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 or recessive alleles, while the effects of cis-regulatory variants are additivity [6,7]. Compared with trans-acting mutations, advantageous cis-regulatory variants more likely to enriched to fixation, because the additive effects expose rare alleles to selection [5].

Both cis- and trans-regulatory variation is important in phenotypic variation [1,8-10], and previous work in a range of species, including Drosophila [7], Mouse [11]...
and Coffea [12], have used allele specific expression (ASE) [13] to differentiate cis- from trans-regulated gene expression variation. Gene expression regulation in birds may be different from mammals, insects or plants. 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 mechanism 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 gene expression regulatory mechanisms in birds.

Chicken is a model animal in 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].

The rapid change under domestication offers a unique example to determine the relative importance of cis- versus trans-regulatory variation underlying phenotypic change. We used reciprocal crosses of White Leghorn (WL), a breed selected for high egg output and a key layer breed, and Cornish Game breeds (CG), a breed selected for rapid growth and muscle development and a cornerstone broiler breed [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

Significant gene expression differences in based in breed, tissue and sex

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 progeny (Fig. 1). In order to determine breed-specific variants, we resequenced the four parents of the two hybrid crosses, recovering on average 100.73 million pair-end reads per sample after quality control. From this, we identified on average 4.74 million SNPs per parental genome which we used to generate simulated parental genomes. From these SNPs, we selected SNPs which were homozygous in each breed (parental birds) but heterozygous in the hybrid birds (cross 2 and cross 3), resulting in 1.4 million of heterozygous SNPs, 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 for three male and three female F1 progeny 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 in based in tissue, sex and parent of origin (Fig. 2). Tissue was the most powerful factor in gene expression. While in different tissues, the sex played a leading role in gene expression of 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. This results in 12 treatment groups, comprised of 3 tissues, 2 sexes and 2 reciprocal crosses in our study. The 24,881 genes from the Ensembl v87 annotation were analyzed. About one fifth of
them contained homozygous SNPs and expressed in our samples (Table S1). We observed significant expression differences (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 genes in brain, 36.45% in liver, and 38.38% in muscle (take the homozygous SNP list of cross 2 for example). In males, we observed significant expression differences for 17.64% of genes in brain, 41.87% in liver, and 37.84% in muscle (Table S1). The differential expression of genes between strains appears to play an important role in phenotypic divergence. For example, the gene \textit{IGFBP2}, \textit{TGFBI}, \textit{PDGFRL} and \textit{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.

An effective pipeline was applied on the allele-specific expression analysis. To identify the origin of offspring’s mRNA, we have 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 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 overlaps with a heterozygous genetic marker, hence is identified as allele-specific read. The tool of ‘asSeq’ has considered the condition of one read containing more than one SNPs [31].

We assigned a read to one of the two parental genomes if the proportion of those heterozygous SNPs suggesting the read is from that genome is larger than 0.9, to ensure the accuracy of parental origin identification.

In order to validate our ASE pipeline, we generated two artificial hybrid F1 libraries. Specifically, we concatenated the RNA-seq fastq files for two male brain samples of 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 were handled consistently with other hybrid libraries, using the homozygous SNP lists of cross 2 and cross 3, respectively. We compared the expression ratio of two alleles (CG WL) of the simulated F1 to the real expression ratio of two strains (CG WL) of each gene, and observed a strong correlation between the two measurements of each group (Fig. S1), indicating that our ASE analysis pipeline was robust. Since our method only counted the local reads containing the homozygous SNPs, we also 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. Our study provided an example to deal with the similar situation where there was no specific reference genome for different strains.

**Genes were classified into different categories by regulatory divergency**

We classified expressed genes into eight categories by the regulatory divergency [7, 35, 36]. Most genes showed conserved or ambiguous expression, as expected given the recent divergence time of these two breeds. There were more than 70% genes classified as conserved in brain, and more than 40% genes in liver, and in muscle, the ratio is around 50%. However, we nonetheless observed substantial cis- and trans- variation in the hybrid F1 crosses. There were a greater proportion of trans-regulated gene expression variations in most tissues and across both sexes, particularly in muscle.

Genes regulated by both cis- and trans-regulatory variations contained 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 in 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”.

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 few loci with consistent significant cis- or trans-regulatory variation across different groups (fig. S6, Table S3).

Genes regulated by trans-acting variation display greater sequence conservation

We counted the number of variants located within 1 kb upstream of transcription start sites of each gene. The results showed greater variations at upstream of cis-regulated genes than trans-regulated genes (Fig. S7).

The ratio of the number of non-synonymous SNPs to the number of synonymous SNPs in coding sequence (pN/pS) was used to assess the degree of selective constraint, with genes under high constraint expected to have lower pN/pS [37,38].

In our research, the value of pN/pS in gene regulated by trans variant is significantly lower than cis-regulated genes in most groups (fig. 4, Fig. S8-S9).

Discussion

On account of the short divergence time, the 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 their 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 across the whole exon, instead of counting the reads number of each SNP. Compared with the previous researches possessing strain-specific reference genomes, our pipeline improved the accuracy of parental origin identification for the SNPs in hybrid offspring, since we have sequenced their parents directly. These SNPs were used to mark the parental origin of the alleles of each gene, and subsequently increased the accuracy of regulation classification. However, it also resulted in the limited number of genes could be studied in this research.

We observed a greater proportion of trans-regulated gene expression variations than cis regulatory variations in most samples. 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 may attribute to the short differentiation time between White Leghorn and Cornish Game. Genes regulated by both cis and trans variations more often acting in opposite directions, and most genes were classified as “compensatory”. This result consistent with the previous researches on house mice [36 indicating that the cis and trans variants tend to act convergently to maintain gene expression stability. It may be due to the short divergence history between the Cornish Games and White Leghorn, and the strict selective pressures that formed and maintain each breed. 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 significant cis- or trans-regulatory variation across different groups. This result consistent with some previous ASE analyses, which suggest rare ASE genes are consistently expressed across tissues [39, 40]. However, the cis- and trans-regulation classification is more complex than the simple ASE analysis. Besides the characteristic of spatiotemporal specificity of gene expression, the gene expression is a complex network controlled by the interaction of cis-regulatory DNA sequences and trans-regulatory factors, which could complicate the identification of regulatory divergence. The classification involved in multiple conditions. The statistical methods could not accurately classify them with the limited expression information. 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-regulated genes than trans-regulated genes, suggesting that the classification results were reliable. Meanwhile, genes regulated by trans-variant showed a lower value of pN/pS than cis-regulated genes in most groups. In theory, the pleiotropic effects of trans-regulatory mutations will result in selection to eliminate most deleterious trans-acting mutations [41]. In contrast, however we might expect a large proportion of cis-regulatory mutations to be largely neutral, and therefore to accumulate over time [9].

The large proportion of trans-regulatory mutations that we observe suggest that artificial selection has primarily acted on trans-regulatory mutations, and that neutral
cis-regulatory mutations have not accumulated substantially in the short time since these breeds were established. Trans-regulated genes display lower pN/pS values on average compared to cis-regulated genes in our data, also suggesting that trans-regulated genes are subject to increased selective constraint during chicken domestication and may have been under stronger artificial selection. This is consistent with similar studies in mice [1], which concluded that trans-regulated genes display greater sequence conservation.

Conclusions

Our analyses indicate important differences among tissue, sex and parent of origin in gene expression regulation. The results also suggest that artificial selection associated with domestication may have more often acted on trans-regulatory variation in the history of chicken commercial breeds establishment. We also demonstrated a pipeline to explore the allele-specific expression on the inbred lines without specific reference genome.

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, for whole genome re-sequencing. Total RNA was extract from three tissues (brain, liver and breast muscle) of 23 chickens at the age of 1 day using Trizol.

DNA & RNA sequencing and data alignment
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) with default parameters.

To ensure the accuracy of RNA-seq data alignment, we simulated four parental genomes. The re-sequencing data 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). The BAM files were sorted, and duplicate reads were removed using Picard toolkit (https://github.com/broadinstitute/picard). The Genome Analysis Toolkit (GATK) 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.

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. 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’.

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 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, we used the binomial test 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 for detecting the trans effects (T). For false discovery rate control, a method of q-value estimation was adopted [48].

To correct the p-value of both binomial test and Fisher’s exact test, and we considered it has significantly difference when the q < 0.05. We classified the expressed genes 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-regulated 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 [49].

The nonsynonymous mutation contains missense variant, start codon lost, start codon gained, stop codon lost and stop codon gained; the 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*
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. Bjornerfeldt S, Webster MT, Vila C: Relaxation of selective constraint on dog mitochondrial DNA following domestication. *Genome Res* 2006, 16(8):990-994.

38. 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.

39. 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.

40. 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.

41. 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.

42. 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.

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

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

45. 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, 20(9):1297-1303.

46. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al: The variant call format and VCFtools. Bioinformatics 2011, 27(15):2156-2158.

47. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR: STAR: ultrafast universal RNA-seq aligner. Bioinformatics 2013, 29(1):15-21.
48. Storey JD, Tibshirani R: Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* 2003, 100(16):9440-9445.

49. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM: A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 2012, 6(2):80-92.

Supplementary Legends

Additional files

**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 probability density (y-axis) of variation count (x-axis) in the 1 kb of DNA sequence located upstream of each gene’s transcription start site.

Figure S8 The ratio of the numbers of non-synonymous SNPs to the numbers of synonymous SNPs (pN/pS) in different groups of cross 2.

Figure S9 The ratio of the numbers of non-synonymous SNPs to the numbers of synonymous SNPs (pN/pS) in different groups of cross 3.

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

Figure 1

Cross design. Cornish-Game (CG) and White-Leghorns (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.
Synonymous and non-synonymous SNPs (pN/pS) in cis- and trans- regulatory genes in different groups. The value of y-axis refers to the mean value of all genes in the category. Significance is indicated between the two regulatory categories (* p<0.05, t-test; ** p<0.01, t-test).

Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.

Additional file 1.docx