Title page

Genetic regulation of gene expression across multiple tissues in chickens

Dailu Guan[1†], Zhonghao Bai[2†], Xiaoning Zhu[3†], Conghao Zhong[4†], Yali Hou[5†], Di Zhu[2,3†], The ChickenGTEx Consortium, Houcheng Li[2], Fangren Lan[4], Shuqi Diao[6], Yuelin Yao[7,8], Bingru Zhao[9], Xiaochang Li[4], Zhangyuan Pan[5], Yahui Gao[6,10,11], Yuzhe Wang[3], Dong Zou[12], Ruizhen Wang[12,13], Tianyi Xu[12], Congjiao Sun[4], Hongwei Yin[14], Jinyan Teng[6], Zhiting Xu[6], Qing Lin[6], Shourong Shi[15], Dan Shao[15], Fabien Degalez[16], Sandrine Lagarrigue[16], Ying Wang[1], Mingshan Wang[17], Minsheng Peng[17], Dominique Rocha[18], Mathieu Charles[18], Jacqueline Smith[19], Kellie Watson[19], Albert Johannes Buitenhuis[2], Goutam Sahana[2], Mogens Sandø Lund[2], Wesley Warren[20], Laurent Frantz[21,22], Greger Larson[23], Susan J. Lamont[24], Wei Si[5,25], Xin Zhao[25], Bingjie Li[26], Haihan Zhang[27], Chenglong Luo[28], Dingming Shu[28], Hao Qu[28], Wei Luo[28], Zhenhui Li[6,29], Qinghua Nie[6,29], Xiquan Zhang[6,29], Ruidong Xiang[30,31,32], Shuli Liu[33], Zhe Zhang[6], Zhang Zhang[12,13], George E. Liu[10], Hans Cheng[34], Ning Yang[4]*, Xiaoxiang Hu[3]*, Huaijun Zhou[1]*, Lingzhao Fang[2]*

[1]Department of Animal Science, University of California, Davis, CA, 95616, USA

[2]Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, 8000, Denmark

[3]State Key Laboratory of Animal Biotech Breeding, College of Biological Sciences, China Agricultural University, Beijing, 100193, China

[4]College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China

[5]Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, 100193, China

[6]State Key Laboratory of Livestock and Poultry Breeding, Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou 510642, China

[7]MRC Human Genetics Unit at the Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh EH4 2XU, UK

[8]School of Informatics, The University of Edinburgh, Edinburgh EH8 9AB, UK

[9]Jiangsu Livestock Embryo Engineering Laboratory, College of Animal Science and Technology, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China

[10]Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service, USDA, Beltsville, Maryland 20705, USA

[11]Department of Animal and Avian Sciences, University of Maryland, College Park, Maryland 20742, USA

[12]Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China

[13]University of Chinese Academy of Sciences, Beijing 100049, China

[14]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Livestock and Poultry Multi-omics of MARA, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518124, China

[15]Poultry Institute, Chinese Academy of Agricultural Sciences, Yangzhou, Jiangsu, 225125, China

[16]PEGASE, INRAE, Institut Agro, 35590, Saint Gilles, France.

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

[18]Paris-Saclay University, INRAE, AgroParisTech, GABI, Jouy-en-Josas, 78350, France

[19]The Roslin Institute, Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian EH25 9RG, UK

[20]Department of Animal Sciences, Data Science and Informatics Institute, University of Missouri, Columbia, MO 65201

[21]Palaeogenomics Group, Department of Veterinary Sciences, Ludwig Maximilian University, Munich 80539, Germany

[22]School of Biological and Behavioural Sciences, Queen Mary University of London, London E1 4DQ, United Kingdom

[23]The Palaeogenomics & Bio-Archaeology Research Network, School of Archaeology, University of Oxford, Oxford, UK

[24]Department of Animal Science, Iowa State University, Ames, Iowa 50011, USA

[25]Department of Animal Science, McGill University, Quebec, H9X 3V9, Canada

[26]Scotland's Rural College (SRUC), Roslin Institute Building, Midlothian EH25 9RG, UK

[27]College of Animal Science and Technology, Hunan Agricultural University, Changsha 410128, China

[28]State Key Laboratory of Swine and Poultry Breeding Industry, Guangdong Key Laboratory of Animal Breeding and Nutrition, Institute of Animal Science, Guangdong Academy of Agricultural Sciences, Guangzhou, 510640, Guangdong China

[29]Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding, and Key Laboratory of Chicken Genetics, Breeding and Reproduction, Ministry of Agriculture, College of Animal Science, South China Agricultural University, Guangzhou, Guangdong, China

[30]Agriculture Victoria, Agribio, Centre for AgriBiosciences, Bundoora, VIC 3083, Australia

[31]Cambridge-Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, 3004, VIC, Australia

[32]School of Agriculture, Food and Ecosystem Sciences, The University of Melbourne, Parkville, 3052, VIC, Australia

[33]School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China

[34]Avian Disease and Oncology Laboratory, USDA, ARS, USNPRC, East Lansing, MI, USA

Corresponding authors*:

Dr. Lingzhao Fang: Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, Denmark

E-mail: lingzhao.fang@qgg.au.dk;

Prof. Huaijun Zhou: Department of Animal Science, University of California, Davis, CA, 95616, USA

E-mail: hzhou@ucdavis.edu;

Prof. Xiaoxiang Hu: State Key Laboratory of Animal Biotech Breeding, College of Biological Sciences, China Agricultural University, Beijing, 100193, China

E-mail: huxx@cau.edu.cn;

Prof. Ning Yang: College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China

E-mail: nyang@cau.edu.cn;

Abstract:
Chicken is a valuable model for understanding fundamental biology and vertebrate evolution, as well as a major global source of nutrient-dense and lean protein. Despite being first non-mammalian amniote to have its genome sequenced, a systematic characterization of functional variation on the chicken genome remains lacking. Here, we integrated bulk RNA-Seq data from 7,015 samples, single-cell RNA-Seq data from 127,598 cells, and 2,869 whole-genome sequences to present a pilot atlas of regulatory variants across 28 chicken tissues. This atlas reveals millions of regulatory effects on primary expression (protein-coding genes, lncRNA and exon) and post-transcriptional modifications (alternative splicing and 3' untranslated region alternative polyadenylation). We highlighted distinct molecular mechanisms underlying these regulatory variants, their context-dependent behavior, and their utility in interpreting genome-wide associations for 39 chicken complex traits. Finally, our comparative analyses of gene regulation between chickens and mammals demonstrate how this resource can facilitate cross-species gene mapping of complex traits.

Main Text:

Introduction

The chicken is a major source of protein-rich food. Its unique phylogenetic position, combined with its genetic and physiological characteristics, has established the chicken as a well-recognized model organism in both fundamental and applied research[1,2], including studies of domestication, developmental biology, and immunology[3–7]. Furthermore, the extensive phenotypic diversity observed in morphology and physiology, shaped by both natural and artificial selection, makes it an ideal system for dissecting the genetic and molecular architecture underlying complex traits[8–11]. Since the first chicken genome was sequenced in 2004, numerous loci associated with complex phenotypes have been identified. Given that most of these genetic variants are located in non-coding regions, a systematic functional annotation of genetic variants is crucial for understanding the regulatory circuitry underlying these traits[12–14].

Inspired by the success of the human Genotype-Tissue Expression (GTEx) project[15–17], the ChickenGTEx project has been launched as part of the international Farm animal GTEx (FarmGTEx) initiative[18–20]. This project aims to develop a reference panel of regulatory effects across various biological contexts in chickens. In this pilot study, through analyzing 7,015 high-quality bulk RNA-Seq datasets from 52 tissues/cell types (hereafter referred to as "tissues"), single-cell transcriptome of 127,598 cells from six tissues, and 2,869 whole genome sequences (WGS) from over 100 breeds/lines (hereafter referred to as "breeds") worldwide, we systematically associated approximately 1.5 million genetic variants with five transcriptomic phenotypes, representing three aspects of primary expression (protein-coding genes, lncRNAs, and exons) and two types of post-transcriptional modifications (alternative splicing and 3' untranslated region alternative polyadenylation [APA]). These analyses were conducted in 28 chicken tissues with over 40 animals. We further explored molecular mechanisms and context-dependent patterns of these regulatory variants and demonstrated their utility in deciphering the molecular architecture of 39 complex traits in chickens. Additionally, we compared gene regulation and its phenotypic implications between chickens and three mammalian species (humans, cattle, and pigs). The overview of the ChickenGTEx data analysis pipeline is presented in **Fig. 1a**, and the resulting data resource is freely accessible at https://chicken.farmgtex.org/.

Results

Harmonizing large transcriptome and genome datasets in chickens

After filtering out low-quality data, 7,015 out of 8,668 RNA-Seq samples were retained for subsequent analyses, representing 28 tissues (**Fig. 1b, Supplementary Fig. 1, Tables S1**). Across these tissues, an average of 23,056 genes were expressed (Transcripts per Million, TPM > 0.1), with an average of 1,938 genes showing tissue-specific expression, consistent with the known tissue biology (**Extended Data Fig. 1a-b**, **Supplementary Fig. 2a-d**, **Table S2**). Notably, about 54.7% of these tissue-specific genes could be linked to previously annotated tissue-specific promoters or enhancers[14] (**Extended Data Fig. 1c-d, Supplementary Fig. 2e-f**). In addition, an average of 114 genes exhibited sex-biased expression across 18 tissues, with 16 of these genes shared across all tissues and located on sex chromosomes (**Extended Data Fig. 1e-f, Table S3**), consistent with known incomplete sex-chromosome dosage compensation in chickens[21]. Furthermore, 41 out of 45 genes associated with Mendelian traits in chickens[22] showed tissue-specific expression (**Supplementary Fig. 3**), such as *SLCO1B3* for the blue eggshell[23] (**Extended Data Fig. 1g**). Additionally, we annotated 43 cell types by analyzing

single-cell RNA-Seq data of 127,598 cells across six chicken tissues (**Supplementary Note**), and each of them had over 100 bulk RNA-Seq samples (**Supplementary Fig. 4, Tables S4-S5**), allowing us to explore cell-type interaction effects.

To obtain genotypes of bulk RNA-Seq data, we called single nucleotide polymorphisms (SNPs) directly from RNA-Seq data[24], and then imputed them genome wide using the genotype imputation panel built from all WGS data. RNA-Seq samples exhibited a similar population structure to that of the reference panel (**Fig. 2a, Extended Data Fig. 2, Supplementary Figs. 5-8, Table S6**). After filtering out low-quality imputed SNPs, approximately 1.5 million SNPs remained for subsequent analyses. On average, each gene's *cis*-window (i.e., ±1Mb of the gene transcription start site [TSS]) contained around 4,000 high-quality SNPs (**Fig. 2b**), and all linkage disequilibrium (LD) blocks estimated in the ChickenGTEx population were tagged by at least one SNP (**Fig. 2c**). Comprehensive evaluation and validation of the genotype imputation are provided in the **Supplementary Note**, **Extended Data Fig. 2,** and **Supplementary Figs. 5-8**. After removing duplicate samples with an identity by segment (IBS) > 0.9, we retained 28 tissues with more than 40 individuals each for subsequent molecular quantitative trait loci (molQTL) mapping (**Fig. 1c**).

molQTL discovery

To uncover the regulatory effect of genetic variants, we quantified five types of molecular phenotypes, including three primary expression phenotypes: protein-coding genes (PCG, n = 15,046 tested), lncRNA genes (n =15,720), and exons (n =163,812, used to complement LeafCutter-based splicing analysis[25,26]), and two post-transcriptional modifications: alternative splicing (n =15,405) and APA (n =11,880) (**Supplementary Fig 9-10**). On average, the median estimates of *cis*-heritability (*cis*-h^2) for PCG expression, lncRNA expression, exon expression, splicing and APA were 0.20, 0.24, 0.21, 0.26, and 0.13, respectively (**Extended Data Fig. 3, Supplementary Figs. 11)**. MolQTL mapping identified significant regulation by at least one genetic variant in at least one of the 28 tissues for 13,983 (92.9%) PCGs, 11,685 (74.3%) lncRNA, 124,423 (76.0%) exon, 9,669 (61.5%) alternative splicing, and 8,798 (74.1%) APA (**Fig. 2d**). We therefore referred to these molecular phenotypes as eGenes, lncGenes, exGenes, sGenes and 3a'Genes, with their associated QTL termed eQTL, lncQTL, exQTL, sQTL, and 3a'QTL, respectively. All these molQTL were enriched around TSS and transcription end sites (TES) (**Fig. 2e**, **Supplementary Fig. 12**). Furthermore, we conducted independent molQTL mapping and revealed that, on average, 73.6% (10,288) of eGenes, 40.5% (3,914) of sGenes, 60.7% (75,527) of exGenes, 58.9% (6,886) of lncGenes, and 7.3% (640) of 3a'Genes were regulated by more than one independent variant across tissues (**Extended Data Fig. 3d, Supplementary Fig. 13**). Though there were no significant differences in the median gene expression level, eGenes with more independent eQTL tend to have a higher *cis*-h^2 (**Fig. 2f**), lower evolutionary constraints (**Fig. 2g)**, and lead variants located further from the TSS (**Fig. 2h**). Fine-mapping analysis using SuSiE-inf[27] revealed 2,887, 2,366, 2,053, 12,409 and 1,572 potential causal variants for eGenes, sGenes, lncGenes, exGenes and 3a'Genes, respectively.

Similar to findings in other species[15,19,20], the statistical power for molQTL mapping depended on tissue sample size (**Fig. 2d**, **Extended Data Fig. 3f-i, Supplementary Fig. 14**). Most eQTL with large effect (i.e., fold change of expression, aFC > 2) became detectable when sample sizes approached 200 (**Extended Data Fig. 3f-i**). In general, the estimated effect size of eQTL was not

correlated with their gene expression levels across tissues (**Extended Data Fig. 3b-c**), similar observations in previous studies[15,20]. By comparing chickens with mammals, we categorized chicken eGenes into orthologues (n = 9,613) and non-orthologues (n = 1,384). Orthologues genes showed higher expression level, lower *cis*-h^2, smaller effect size for lead variants, and lead variants located closer to the TSS compared to non-orthologues (**Fig. 2i-j**). Moreover, the effect sizes of lead eQTL of orthologous genes were significantly, albeit weakly, correlated between chickens and mammals, with correlations lower than those observed within mammals (**Extended Data Fig. 5a**). A similar pattern was noted for *cis*-h^2 of orthologous genes (**Extended Data Fig. 5a**).

To evaluate the potential impact using RNA-Seq-inputed genotypes for *cis*-eQTL mapping, we conducted parallel analyses using both RNA-Seq-imputed genotypes and paired WGS-derived genotypes in a population of 308 chickens. The RNA-Seq imputation approach demonstrated comparable statistical power for eGene discovery (**Extended Data Fig. 4a-c**). Although the two methods often identified different lead variants (**Extended Data Fig. 4d-i**), independent eQTL (**Extended Data Fig. 4j**), and fine-mapped variants (**Extended Data Fig. 4k-l**), the lead variants of both methods were generally in high LD, suggesting that they capture the same causal signals (**Extended Data Fig. 4g-h, m-o**).

To validate molQTL identified above, we conducted the internal validation and observed a high replication rate of eQTL, measured by π_1[28], across tissues, ranging from 0.61 in the hypothalamus to 0.92 in the embryo (**Extended Data Fig. 5b**, **Supplementary Fig. 15**). Moreover, effect sizes derived from the eQTL mapping were significantly and positively correlated with those from allele-specific expression analysis at the same loci (**Extended Data Fig. 5c, Table S7).** By training a deep learning model, DeepSEA[29], based on 310 chicken functional epigenomic profiles (**Table S8**), we also observed that regulatory variants predicted by DeepSEA were significantly enriched among eQTL (**Extended Data Fig. 5d, Supplementary Fig. 16**). Finally, 62%, 60%, and 72% of ChickenGTEx eQTL identified in the hypothalamus, liver, and pituitary, respectively, were replicated in matched tissues from an independent chicken population (**Extended Data Fig 6a**). Although this replication rate is relatively low compared to other eQTL mapping studies[15,19,20], it may reflect differences in sample size, effect size, minor allele frequency (MAF) and LD structure between ChickenGTEx and validation populations (**Extended Data Fig. 6c-j**). When considering only eQTL within shared haplotype blocks between ChickenGTEx and validation populations, almost all eQTL were replicated (**Extended Data Fig. 6j**). Additionally, molQTL mapping using a linear mixed model (LMM) were conducted to account for family relatedness (**Supplementary Note**, **Extended Data Fig. 6b, Supplementary Fig. 17**).

Limited overlap in regulatory mechanisms across five molQTL types

To investigate whether different molecular phenotypes of the same genes share regulatory variants, we conducted colocalization between any two types of molQTL. The LD between lead variants for two molecular phenotypes from the same gene was low, with the median estimates ranging from 0.01 (exQTL *vs.* 3a'QTL) to 0.09 (exQTL *vs.* lncQTL) (**Fig. 3a-b, Supplementary Fig. 18**). Further analyses revealed corresponding colocalization probabilities were also low to moderate, ranging from 0.03 (sQTL *vs.* 3a'QTL) to 0.77 (exQTL *vs.* lncQTL) (**Fig. 3a-b,**

Supplementary Fig. 18). These findings suggest largely distinct regulatory mechanisms among different molQTL types.

Functional enrichment analyses further revealed distinct enrichments of molQTL across various regulatory DNA sequences (**Fig. 3e**). All molQTL exhibited a significant enrichment in missense variants (**Fig. 3c**), indicating that some regulatory variants may alter protein-coding sequence[30]. In addition, molQTL were most highly enriched in promoter-like states (E1-E5, average 3.64-fold), followed by enhancer-like states (E6-E10, average 1.98-fold) and ATAC islands (E11, average 1.87-fold) (**Fig. 3d**). Interestingly, super-enhancers showed lower enrichment compared to standard enhancers (E1) (**Extended Data Fig. 7a**). Moreover, a notable fraction of molQTL was supported by previously predicted regulator–gene pairs: 20% of eQTL, 26% of sQTL, 3.4% of lncQTL, 17.9% of exQTL, and 14.5% of 3a'QTL[31] (**Extended Data Fig. 7b)**. Analysis of 3D chromatin loop[31] revealed that 20-60% of molQTL-gene pairs were located within the same topologically associating domain (TAD) across tissues (**Extended Data Fig. 7c, Supplementary Fig. 19a**), with the highest enrichment observed for pairs approximately 400-600kp away from the TSS of target genes after adjusting for distance (**Extended Data Fig. 7d**). Likewise, 41-73% of eQTL-eGene pairs were located in the same CTCF-loops across tissues[14] (**Supplementary Fig. 19b-d**). An illustrative gene example is shown in **Extended Data Fig. 7e**.

Context-dependence of molQTL

In general, tissues with similar biological functions were clustered based on molQTL effect similarity (**Fig. 4a-b, Supplementary Fig. 20**). Notably, molQTL for post-transcriptional modifications (sQTL and 3a'QTL) showed higher tissue specificity than those for primary expression (eQTL, exQTL and lncQTL) (**Fig. 4a, Supplementary Fig. 20-21)**. In total, 10.6% of eQTL, 25.8% of exQTL, 27.4% of lncQTL, 29.6% of 3a'QTL and 32.1% of sQTL were active exclusively in one tissue (**Fig. 4c**). The eQTL active in more tissues tended to have smaller effect size (**Extended Data Fig. 8a**), higher MAF (**Extended Data Fig. 8b**), and stronger enrichment around the TSS (**Extended Data Fig. 8c-d**). Unlike mammalian GTEx studies[15,19,20], blood in chickens formed a distinct outgroup (**Fig. 4a**, **Extended Data Fig. 8f-h**), possibly due to the presence of nucleated red blood cells in avian blood. Moreover, genetic regulation in embryonic tissues was distinct from that in adult tissues (**Fig. 4a, Supplementary Fig. 20**), similar to observations in pigs[20], indicating unique regulatory landscapes during early development. In addition, 59 eQTL exhibited opposite directional effects on the same genes (n = 51) between tissues (**Table S9**). For instance, the G-allele of *rs313608694* significantly upregulated the expression of *ELAC2* in the embryo but downregulated it in the adult spleen (**Extended Data Fig. 8i**); reduced *ELAC2* expression has been linked to growth arrest via suppression of transforming growth factor-beta[32].

We also examined breed-sharing patterns for eQTL in brain, liver, muscle, and spleen —tissues with more than two breeds and sample size > 40 each. On average, 81% of eQTL were replicated between breeds (**Extended Data Fig. 8j-k, Table S10**). For instance, the T-allele of *rs314795649* consistently upregulated *PRKCDBP* expression in the liver across all four breeds being tested (**Extended Data Fig. 8l**).

To further explore context-dependent gene regulation, we mapped eQTL interacting with sex (sb-eQTL), transcription factor (TF-eQTL), and cell type (ci-eQTL) (**Extended Data Fig. 9**). For

sb-eQTL mapping, we focused on eight tissues with at least 30 individuals per sex, identifying 1,138 SNPs that displayed sex-biased regulation of 962 eGenes (sb-eGene, FDR < 0.01) (**Tables S11-12**). For instance, *rs317663121* was significantly associated with *TCFL5* expression in male liver only (**Extended Data Fig. 9a**). Analysis of 956 chicken transcription factors from AnimalTFDB 3.0[33] detected an average of 1,941 TF-eQTL across 17 tissues, representing 503 TFs. In muscle, for instance, the effect of *rs313600592* on *ATP6V1A* expression was significantly modulated by the abundance of transcription factor *TCF25* **(Extended Data Fig. 9d)**. For ci-eQTL mapping (**Fig. 4d**, **Supplementary Fig. 22-23**), we identified an average of 129 ci-eGenes across six tissues (**Extended Data Fig. 9e**). Compared with standard eQTL, ci-eQTL exhibited lower enrichment around the TSS (**Extended Data Fig. 9f**), higher enrichment in weak promoters and enhancers (**Extended Data Fig. 9g**). The low LD (median = 0.02) between lead variants for standard eGenes and ci-eGenes further indicates distinct regulatory mechanisms (**Extended Data Fig. 9h**). An example is shown in **Extended Data Fig. 9i**, where the G-allele of *rs16084159* was associated with decreased expression in muscle, particularly among individuals with a low proportion of dendritic cells, an eQTL effect that diminished in individuals with high dendritic cell composition. Using MashR[34], we found that ci-eQTL sharing across cell types was significantly lower than eQTL sharing across tissues (Wilcoxon rank sum test, $P < 2.2 \times 10^{-16}$) (**Fig. 4e, Extended Data Fig. 9j**). An illustrative example is the regulation of *GH*: one SNP (*rs315662005*) significantly regulated *GH* in the liver, while another independent SNP (*rs14303039*, LD = 0.01) showed significant interaction with erythrocyte enrichment on its expression (**Fig. 4f**).

Interpreting genetic regulation behind complex traits

To demonstrate the potential of molQTL for elucidating regulatory mechanisms underlying complex traits in chickens, we systematically integrated molQTL with genome-wide association (GWAS) results for 39 chicken complex traits from an advanced intercross line (AIL)[35] (**Table S13**). We applied four methods including fastENLOC-based colocalization[36], summary-data-based MR (SMR)[37], single-tissue transcriptome-wide association study (sTWAS)[38], and multi-tissue TWAS (mTWAS)[39] to prioritize causal variants and genes at each GWAS loci. Among the 579 GWAS loci examined, 75.1%, 58.55%, 19.5%, and 36.44% were linked to molecular phenotypes by S-MultiXcan (mTWAS), S-prediXcan (sTWAS), fastENLOC and SMR, respectively (**Fig. 5a, Supplementary Fig. 24a-b**). The high discovery rate via both TWASs may partially result from LD contamination[40] (**Supplementary Fig. 25**). Considering SMR and fastENLOC results together, 222 (38%) out of 579 GWAS loci could be linked to at least one molQTL (**Fig. 5b**), a relatively low colocalization rate that mirrors observations in human studies[15,41]. This low overlap may reflect differences in LD patterns between ChickenGTEx and AIL GWAS populations[35] (**Supplementary Fig. 25a-b**) as well as systematic difference in discovery of molQTL and GWAS loci[42,43] (**Supplementary Fig. 26**). Notably, among 105 eQTL-GWAS colocalizations, 97 (92.38%) were not colocalized with lead GWAS variants of the nearest genes (**Fig. 5c**). For example, sTWAS analysis for body weight gain from week 6 to 8 (WG6.8) linked GWAS loci to 43 genes across 21 tissues (**Table S14**); among these, *KPNA3* exhibited the strongest association in the retina, followed by the pituitary (**Supplementary Fig. 25c, Table S14**). The highest colocalization for a WG6.8 GWAS loci was observed for a retina eQTL at *KPNA3* (*rs314814283*) (**Fig. 5d**). *KPNA3*, which encodes karyopherin subunit alpha 3, is critical for photoreceptor development [44], highlighting its importance in regulating growth [45,46]. The heterozygous variants in *KPNA3* are linked to infantile-onset hereditary spastic paraplegia in

humans, characterized by delayed motor development[47]. Fine-mapping analysis revealed that 193 of the 579 GWAS loci harbored at least two independent signals, and loci with multiple signals were more likely to colocalize with several molQTL than loci with a single signal (**Supplementary Fig. 24c-d**).

Comparing genetic regulation between chickens and mammals

To explore conservation of gene expression and genetic regulation between chickens and mammals, we conducted comparative analyses using data from FarmGTEx[19,20] and human GTEx[15] (**Supplementary Fig. 27-28**). Our analyses revealed a degree of conservations in the number of expressed genes, overall gene expression pattern, and tissue specificity (**Supplementary Fig. 27c-e**). However, correlations of *cis*-h^2 and eQTL effects for orthologous genes were notably lower between chickens and mammals than among mammalian species (**Extended Data Fig. 5a**). Leveraging these conservation patterns[20,48], we further explored cross-species similarities in complex trait genetics using TWAS results for orthologous genes. This approach identified 73 significant trait-pairs across tissues with correlations between chickens and three mammalian species (FDR < 0.1) (**Table S15**). Many of these correlations recapitulated known biological relationships; for example, body weight (BW) in chickens was significantly correlated with average daily gain (ADG) in pigs (**Fig. 6, Supplementary Figs. 24h-i** and **29**).

Because gene synteny and LD patterns often differ between species, cross-species meta-TWAS analysis offers a promising strategy to identify shared causal genes and underlying molecular mechanisms for complex traits[48,49]. For example, a two-species meta-TWAS analysis in muscle identified several growth-related traits (e.g., pig backfat thickness and human height) that helped detect additiional genes associated with chicken BW **(Supplementary Fig. 24f-g)**. Expanding this approach to three species (chicken, pig, and human) and focusing on 676 one-to-one orthologous genes related to chicken BW, pig backfat thickness, and human height. We identified 61 significant genes, of which 57 were new for chicken BW (**Extended Data Fig. 10a, Table S16**). Among these 57 genes, 14 showed nominal significance ($P < 0.05$) in two independent chicken populations, a number significantly higher than expected (one-tailed Fisher's exact test, $P = 0.038$)[50] (**Extended Data Fig. 10b**). Further phenome-wide association (PheWAS) analyses and examples in humans are provided in **Extended Data Fig. 10c-d**, and **Supplementary Fig. 30.**

Discussions

Here, we present the pilot ChickenGTEx resource: a comprehensive catalogue of genetic regulatory effects on five transcriptional signatures across 28 tissues. This resource provides deep insights into the regulatory hierarchy governing genetic variation in the transcriptome and its impact on complex phenotypes in chickens. The ChickenGTEx web portal offers free access to all data and includes user-friendly visualization tools for exploring all the results. Our analyses reveal that different molecular phenotypes of the same gene are often controlled by distinct genomic loci, and each contributing uniquely to complex trait variation. This findings, consistent with observations in humans[51,52], underscores the importance of characterizing diverse molecular phenotypes to capture a comprehensive landscape of genetic regulatory effects on an organism's transcriptome[53]. Moreover, we demonstrate the utility of molQTL in interpreting GWAS findings in chickens and translating genetic insights across species at the functional level of orthologous genes.

Despite the expansive source provided by the current version of ChickenGTEx, challenges remain in accurately assessing genotypes and molecular phenotypes. Future studies will integrate RNA-Seq with paired DNA sequences to better resolve genomic variants e.g., rare and splice-variants[54–56], and will employ pangenome references to annotate complex structural variants[57], mobile element variation [58], and short tandem repeats [59,60]. Additionally, we plan to incorporate other molecular features (e.g., chromatin accessibility) across diverse biological and environmental conditions, in both *in vivo* and *in vitro* systems at single-cell resolution. Experimental follow-ups, e.g., CRISPR-based screens[61], will be essential for functionally validating these regulatory effects. In summary, the current and future iterations of the ChickenGTEx project promises to establish a robust reference panel for investigating functional impacts of genetic variants in chickens.

Acknowledgments:
We thank all the researchers who have contributed to the publicly available data used in this research. We extend special acknowledgment to the human GTEx consortium[15] for their sharing of computational codes for some data analyses and figure generation.

L.F. was supported by Agriculture and Food Research Initiative Competitive grants nos. 2022-67015-36215 (H.Z. and L.F.) from the USDA National Institute of Food and Agriculture, and seed-funding from CellFood Hub (Aarhus University Foundation, AUFF). D. Guan was supported by Agriculture and Food Research Initiative Competitive grants nos. 2020-67015-31175 (H.Z.), and 2022-67015-36215 (H.Z. and L.F.) from the USDA National Institute of Food and Agriculture. H.Z. acknowledges fundings from Agriculture and Food Research Initiative Competitive grants nos. 2015-67015-22940 (H.Z.) and 2022-67015-36215 (H.Z. and L.F.) from the USDA National Institute of Food and Agriculture, Multistate Research Project NRSP8 and NC1170, and the California Agricultural Experimental Station. N.Y. acknowledges fundings from the National Key Research and Development Program of China (2021YFD1300600 and 2022YFF1000204). X. H. was supported by the National Natural Science Foundation of China, (31961133003). Y.W. was supported by the National Natural Science Foundation of China, , (32272862). S Rong acknowledges funding from Jiangsu Agricultural Industry Technology System (JATS[2022]406). Zhe Zhang acknowledges fundings from the National Natural Science Foundation of China (32022078 to Z.Z.), the Local Innovative and Research Teams Project of Guangdong Province (2019BT02N630 to Q.N. and Z.Z.). Zhang Zhang acknowledges fundings from National Natural Science Foundation of China (32030021), National Key Research & Development Program of China (2021YFF0703702) and Technical Support Talent Program of Chinese Academy of Sciences (awarded to DZ). Y.H. acknowledges funding from the National Key Research and Development Program of China (2024YFF1000100), and the Science and Technology Innovation 2030 - Major Project (2022ZD04017). L.W. H.Q. and C.L., were supported by Science and Technology Planning Project of Guangzhou City (201504010017) and Natural Scientific Foundation of China (31402067). G.E.L. was supported in part by USDA NIFA AFRI grant numbers 2019-67015-29321 and 2021-67015-33409 and the appropriated project 8042-31000-112-00-D, of the USDA Agricultural Research Service (ARS).

Author contribution statement:

L.Fang, H. Zhou, D.G., X.H., and N.Y. conceived and designed the project. D.G., Y.Y., B.Z. and Z.P. performed bioinformatic analyses of RNA-Seq data analysis. D.G., F.L., S.D., Y.G. and H.Y. conducted whole-genome sequence data analysis. D.Z., performed the deep learning analysis. D.G. and Y.H. performed multi-omics and single-cell RNA-Seq data analysis. D.G. conducted molQTL mapping. X.Z., C.Z. D.G. performed GWAS integrative analysis. Z.B. and D.G. led the comparison of GTEx between chickens and mammals. L.F., H.Zhou, D.G., X.Z., Q.L., C.Z., Y.H., Y.W., C.S., J.T., F.D., S.L., Y.W., M.W., M.P., D.R., M.C., J.S., K.W., A.J.B., W.W., L.Frantz, G.L., M.S.L., G.S., S.S., D.S., S.J.L., X.Z., X.R., S.L., B.L., H.Zhang, and H.C. contributed to the critical interpretation of analytical results before and during manuscript preparation. Y.H., D.Z., R.W., T.X., and Zhang Zhang built the ChickenGTEx web portal. H. Zhou, L.Fang, N.Y., X.H., G.E.L., Zhe Zhang, S.S., D.S., X.Z., Q.N., Z.L., W.L., H.Q., W. S. and C.L. contributed to the data and computational resources. D.G., Z.B., X.Z., C.Z., Y.W., Y.H. and L.Fang drafted the manuscript. All authors read, edited, and approved the final manuscript.

Competing Interests Statement:

The authors declare no competing interests.

Figure legends:

Figure 1. Workflow and data summary of the ChickenGTEx. (**a**) The overall workflow of data analyses. (**b**) Hierarchical clustering of 7,015 RNA-Seq samples based on gene expression values (quantified as Transcripts per Million, TPM) of 5,000 genes with the highest expression variance (measured by standard deviation) (top), and splicing variations of 1,000 intron clusters with the highest variance (bottom). (**c**). Illustration of tissue types used in molecular quantitative trait loci (QTL) mapping. Sample sizes (in the bracket) and colors of all the 28 tissue types with sample sizes over 40 are depicted. SRA: Sequence Read Archive, GSA: Genome Sequence Archive, RNA-Seq: RNA sequencing, APA: 3'UTR alternative polyadenylation, LM: linear regression model, LMM: linear mixed model, PCG: protein-coding gene, GTEx: Genotype-Tissue Expression, molQTL: molecular QTL, GWAS: genome-wide association study, TWAS: transcriptome-wide association study, MR: Mendelian randomization. eQTL: Expression QTL, sQTL: Splicing QTL, exQTL: Exon expression QTL, lncQTL: lncRNA gene expression QTL, 3a'QTL: QTL for APA. ci-eQTL: cell type interaction eQTL, sb-eQTL: Sex-biased eQTL, TF-eQTL: transcription factor interaction eQTL. CNS: central nervous system.

Figure 2. Genotype imputation and molecular QTL (molQTL) mapping in 28 chicken tissues. (**a**) Scatterplots depicting principal component analysis (PCA) of 2,869 whole-genome sequence (WGS, left) and 7,015 RNA-Seq samples (right). PCA was carried out using 1.52 million SNP genotypes shared by both WGS and RNA-Seq datasets. PC: principal componenet. (**b**) Number of imputed high-quality SNPs ($DR^2 > 0.8$ & MAF > 0.05) per (i.e., ±1 Mb of the transcriptional start site, TSS). (**c**) The number of imputed high-quality SNPs per linkage disequilibrium (LD) block. The median number of SNPs per LD block is 5. The distribution of LD block size is shown at the right corner. The LD block was estimated using PLINK v1.9 software[62]. (**d**) Proportion of ePhenotypes (i.e., molecular phenotypes regulated by at least one genetic variant) detected as a function of tissue sample size. (**e**) Proportion of lead variants as a function of their distance to respective genes. TSS: transcription start site, TES: transcription end site. The length of a gene was split into 10 bins and the corresponding upstream and downstream

was split into 30 bins. Each lead variant was assigned to the bins according to their relevant distance. (**f**) Comparison of *cis*-heritability (left, *cis*-h^2) and gene expression (right, Transcripts per Million [TPM]) between non-eGenes (1,063) and eGenes (13,983) with different number of independent SNPs detected. (**g**) Comparison of sequence constraints (phastCons score) between non-eGenes (1,063) and eGenes (13,983) with different numbers of independent SNPs detected. The phastCons score was obtained from the UCSC genome browser (https://genome.ucsc.edu/). (**h**) Comparison of distance of lead variants to transcription start site (TSS) between non-eGenes (1,063) and eGenes (13,983) with different number of independent SNPs detected. (**j**) Comparison of phastCons score, gene expression, *cis*-h^2, and distance of lead variant to TSS (from top to bottom) between orthologous (n=9,613) and non-orthologous eGenes (n=1,384). The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range. The statistical *P* values in (f) and (h) were obtained by two-sided Student's *t*-test, and in (g) and (j) were obtained by the two-sided Wilcoxon test. (**i**) Functional enrichment of non-orthologous eGenes (1,384) and non-orthologous non-eGene (440).

Figure 3. Colocalization and functional enrichment of molQTL. (**a**) Colocalization analyses among different types of molQTL of the same genes. The "LD" is the linkage disequilibrium (LD, r^2) of lead SNPs of two molecular phenotypes derived from the same gene. "PP_{H4}", estimated by coloc software[63], represents the probability of whether the association of two molecular phenotypes is due to the shared SNP. ePhenotype: molecular phenotype that is regulated by at least one genetic variant. (**b**) Examples of non-colocalization between different types of molQTL of the same genes. Each panel corresponds to an example from (a), displayed vertically. The y-axis represents the significance (-$\log_{10}$ scaled) of eQTL mapping with the linear regression. The proportion and enrichment of molQTL across sequence ontology (i.e., variant types annotated by SnpEff software[64]) (**c**) and 15 chromatin states[14] (**d**). Fold enrichment is shown as mean (dot) ± standard deviation ($\log_2$ scaled, error bar) across 28 chicken tissues. The error bars in the right bar plots indicate standard deviation of the proportion of variants across 28 chicken tissues. The chromatin states were retrieved from Pan et al. (2023) [14]. TssA: strongly active promoters/transcripts, TssAHet: flanking active TSS without ATAC, TxFlnk: transcribed at gene, TxFlnkWk: weak transcribed at gene, TxFlnkHet: transcribed region without ATAC, EnhA: strong active enhancer, EnhAMe: medium enhancer with ATAC, EnhAWk: weak active enhancer, TxFlnkWk: active enhancer no ATAC (hetero), EnhPois: poised enhancer, ATAC_Is: ATAC island, TssBiv: bivalent/poised TSS, Repr: repressed polycomb, ReprWk: weak repressed polycomb. The design of panels (e) and (f) was adapted from the human GTEx project[15–17].

Figure 4. Tissue-sharing and context-dependent patterns of molQTL. (**a**) The heatmap of Spearman's correlation of eQTL effect size between tissues. We employed MashR to do this analysis[34], similar to the human GTEx[15]. Tissues are clustered on the basis of dissimilarities (i.e. 1-d), where d is Euclidean distance calculated from the eQTL effect, with a complete linkage method[65]. **(b)** Similarity of molQTL effect-based tissue clustering patterns[15]. The pairwise Rand Index[66] across five types of molQTL was used for measuring the similarity, ranging from 0 to 1, where 0 means that two tissue clustering patterns do not match at all, while 1 means that two clustering patterns match exactly. **(c)** Fraction of molQTL active across tissues[15,16]. An eQTL is considered active in a tissue when its local false sign rate (LFSR, equivalent to FDR) < 5% in MashR analysis[34]. **(d)** Stacked bar plots showing the fraction of cell types estimated in bulk

RNA-Seq samples (n=741) based on single cell RNA-Seq reference matrix in the liver. This estimation is done by CIBERSORTx[67]. **(e)** Distribution of Spearman's correlations of QTL effect size between tissues (eQTL) and between cell types (cell-type interaction QTL, ci-eQTL). The Spearman's correlations were estimated by running MashR[34]. **(f)** Manhattan plots depicting the significance ($-\log_{10}P$, linear regression) of eQTL (bottom) and ci-eQTL (top panel) for the *growth hormone* (*GH*) in the liver on chromosome 27. This ci-eQTL was detected in the interaction of the erythrocytes and genotypes in liver. The lead eQTL (*rs315662005*) and ci-eQTL (*rs14303039*) were indicated by diamond and triangle shape, respectively. Each dot represents a SNP and its color represents the LD degree between the lead eQTL/ci-eQTL and the rest. The LD (r^2) between *rs315662005* and *rs14303039* is 0.01 in the ChickenGTEx populations.

Fig. 5. Interpretation of GWAS loci with molQTL. (**a**) UpsetR[68] plot depicting the number of GWAS loci linking to at least one molQTL by four integrative methods, including fastENLOC-based colocalization[36], Summary-based Mendelian randomization (SMR)[37], single tissue-based transcriptome-wide association study (sTWAS)[38], and multi-tissue TWAS (mTWAS)[39]. (**b**) UpsetR[68] plot depicting the number of GWAS loci linking to at least one type of molQTL by SMR[37] and fastENLOC[36] methods. (**c**) The proportion of GWAS loci (n = 579) colocalizing with eQTL regarding the integration results using fastENLOC-based colocalization[36], and Summary-based Mendelian randomization (SMR)[37]. No colocalization: GWAS loci that are not linked to any eGenes in 28 tissues. Not nearest gene: GWAS loci linked to eGenes that are not nearest genes to GWAS lead SNPs. Nearest gene: GWAS loci linked to eGenes that are the nearest ones to GWAS lead SNPs. The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range. (**d**) Linking of GWAS loci of weight gain from week 6 to 8 (WG6.8) with molQTL. The top panel shows the Manhattan plot of GWAS associations of SNPs (linear mixed model) with WG6.8 on chromosome 1. The color indicated linkage disequilibrium (LD, r^2) of SNPs with the colocalized SNP (black diamond, *rs314814283*). The bottom plot represents eQTL mapping (linear regression) results of *KPNA3* in the retina. The color represents LD (r^2) values of the colocalized SNP (*rs314814283*, black diamond) with the rest. The bottom panel shows the GWAS results before (gray) and after (maroon) conditioning on the colocalized variant (*rs314814283*).

Fig. 6. Comparative analyses of gene regulation and transcriptome-wide associations (TWAS) between chickens and mammals. Significance (at $\log_{10}$ transformed, vertical bar, y-axis) for TWAS-based correlations calculated from one-to-one orthologous gene effect between chicken and mammals (i.e. pig, chicken, cattle). Panels **(a)-(c)** respectively present the associations of chicken body weight **(a)**, chicken egg weight at 52 week **(b)**, chicken feed intake from 6 to 8 weeks **(c)** with available TWAS data in mammals. The red dashed line depicts the FDR[69]-based threshold of significance (FDR < 0.1). The red dot and sign ("+" and "-") indicates they are significant and the correlation direction, respectively. The right side is shown the corresponding quantile–quantile plot, draw by the qqman package[70]. **(d)** Phenome-wide association (PheWAS) study of the *BRICD5* gene with human traits. *x*-axis shows the negative base-10 logarithm of the *P*-value (obtained in the GWAS atlas at https://atlas.ctglab.nl), and *y*-axis indicates traits tested. Each dot represents a trait, colored by trait categories, the red dash line indicated the FDR[69]-based threshold (0.05).

Extended data figures

Extended Data Fig. 1. Characterization of gene expressions. **(a)** Gene expression levels according to the number of tissues where genes are expressed (> 0.1 TPM in >80% of RNA-Seq samples). The expressions (TPM) of genes were averaged in a given number of tissues. TPM: Transcripts per Million. **(b)** Functional enrichment of tissue-specific genes based on the Gene Ontology (GO) database. The color scale from light to deep means a negative logarithm of false discovery rate (FDR)[69] at the base of 10, obtained by the clusterProfiler 4.0 package with default settings[71]. **(c)** Number of tissue-specific genes that were linked to at least one epigenetic regulator in the respective tissue. The annotation of epigenetic regulators was retrieved from Pan et al. (2023)[14]. **(d)** Proportion of tissue-specific genes that were linked to at least one regulator (Regulator-specific) across tissues. The other is for the rest of the tissue-specific genes in each of the tissues. **(e)** Liver genes differentially expressed (n = 327, presenting at x-axis) by sex. Genes were identified according to expression differences between males and females. Each bar represents a gene being tested. **(f)** Ideogram[72] depicting the locations of 17 genes (triangles) showing sex-biased expression across 18 tissues. Each triangle represents a gene that exhibited sex-biased expression across 18 tissues. The color scale represents the gene density in a 1Mb window. **(g)** An example showing the expression of *SLCO1B3* across 27 tissues (n=5,273 samples) and the chromatin states around it across 23 tissues[14]. This gene has been associated with the chicken blue eggshell, and the potential causal mutation was identified to be an EAV-HP insertion located in 5' flanking region of the *SLCO1B3* [23]. The expression level in liver (n=903) is significantly higher than the retina (Student two-sided t-test, Bonferroni adjusted $P = 1.12 \times 10^{-239}$). The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range. The middle panel indicates the chromatin interaction map retrieved from the UCSC genome browser (http://genome.ucsc.edu/s/zhypan/galGal6_FAANG_V1)[14]. The color from lighter to darker represent the interaction from low to strong. TssA: strongly active promoters/transcripts, TssAHet: flanking active TSS without ATAC, TxFlnk: transcribed at gene, TxFlnkWk: weak transcribed at gene, TxFlnkHet: transcribed region without ATAC, EnhA: strong active enhancer, EnhAMe: medium enhancer with ATAC, EnhAWk: weak active enhancer, TxFlnkWk: active enhancer no ATAC (hetero), EnhPois: poised enhancer, ATAC_Is: ATAC island, TssBiv: bivalent/poised TSS, Repr: repressed polycomb, ReprWk: weak repressed polycomb. The design of panels (a) and (e) is adapted from the human GTEx project[15–17].

Extended Data Fig. 2. Genotype imputation. **(a)** The correlation between the number of high-quality SNPs (DR2>0.8 & MAF > 0.05, *y*-axis) in *cis*-windows and the corresponding gene expression at $\log_2$ scaled (*x*-axis) (Pearson's $r = 0.13$, and *P*-value = 2.2×10^{-16}). *cis*-window: ±1 Mb of the transcriptional start site (TSS). *P* values were computed *via* the asymptotic *t* approximation. **(b)** Number of high-quality SNPs as a function of tissue specificity of gene expression, measured by Tau value[73].). *P* values were computed *via* the asymptotic *t* approximation. **(c)** Percent of high-quality SNPs resides in chromatin states, histone modifications, chromatin accessibility, and CTCF. TssA: strongly active promoters/transcripts, TssAHet: flanking active TSS without ATAC, TxFlnk: transcribed at gene, TxFlnkWk: weak transcribed at gene, TxFlnkHet: transcribed region without ATAC, EnhA: strong active enhancer, EnhAMe: medium enhancer with ATAC, EnhAWk: weak active enhancer, TxFlnkWk: active enhancer no ATAC (hetero), EnhPois: poised enhancer, ATAC_Is: ATAC island, TssBiv: bivalent/poised transcriptional start site (TSS), Repr: repressed polycomb, ReprWk: weak repressed polycomb. **(d)** Percent of epigenomic regulators (the same as the panel

(c)) with at least one high-quality SNP. **(e)** The proportion of imputed and reference genotypes by variant types. **(f)** Genotype concordance and Spearman's correlation using paired whole genome sequences (WGS) and RNA-Seq samples in 6 external populations. CLB: Chinese local chicken breed (WGS, n = 280; RNA-Seq samples [bursa, n = 277; liver, n = 278; muscle, n = 267; spleen, n = 274]), Cobb (WGS, n = 21, RNA-Seq samples [brain, n = 20; liver, n = 20, muscle, n = 20, spleen, n = 20]), Commercial : Commercial chicken population (WGS, n = 13, RNA-Seq samples [jejunum, n = 10; spleen, n=11; kidney, n = 11]), Indigenous: Chinese indigenous chicken populations (WGS, n=18, RNA-Seq samples [jejunum, n = 16; spleen, n=17; kidney, n = 17]), WPB: white plumage broiler (WGS, n=112, RNA-Seq samples [ovary, n = 112]), AIL: Advanced intercross line (WGS, n = 30; RNA-Seq samples [duodenum, n = 25; jejunum, n = 4]). **(g)** Genotype concordance and Spearman's correlation (mean ± s.d.) using paired WGS and RNA-Seq samples with the same data as the panel (e) as a function of variant types**. (h)** The number of imputed SNPs (x-axis) as a function of the number of SNPs directly called from WGS (y-axis) within *cis*-windows (±1 Mb of TSS) of 16,779 genes. The significance was obtained with two-sided Student's *t*-test. **(i)** Relationship between median expression levels of genes across samples (y-axis) and the ratio of imputed to observed SNPs (x-axis). Imputed SNPs are obtained with the genotype imputation and observed SNPs were called directly from paired WGS. The significance was obtained with two-sided Student's *t*-test. **(j)** Number of imputed and observed SNPs as a function of distance to the nearest TSS, stratified by median expression levels of genes across samples with the same data as the panel (f). The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range. The significance was obtained with two-sided Student's *t*-test. *: $P < 0.05$; n.s.: not significant.

Extended Data Fig. 3. Molecular quantitative trait locus (molQTL) mapping across 28 chicken tissues. (a) *cis*-heritability (*cis*-h^2) of 219,728 eGene-tissue pairs and 46,303 non-eGene-tissue pairs. The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range. **(b)** Gene expressions measured as Transcripts per Million (TPM) for eGenes in at least one tissue and non-eGenes across 28 tissues. **(c)** Spearman correlations of lead eQTL effect size and gene expression across 28 tissues. NA (low expr.): not tested due to low expression level, Uncorrelated: tested but not significant correlation, Uninterpretable: a significant correlation but with unclear correlation direction due to effect sizes crossed zero, Positive corr.: positively correlated, Negative corr.: negatively correlated. This figure design is adopted from the human GTEx[15]. **(d)** Conditionally independent eQTL across all 28 tissues. Proportion of eGenes with different numbers of independent eQTL being detected (blue stacked bars; left *y*-axis), and mean number of independent eQTL per eGene (red dots; right *y*-axis). Tissues are sorted from smallest to largest regarding sample size. Tissue color legend is shown at the right side. The figure is generated by using human GTEx project codes[15]. **(e)** Slope (*y*-axis, obtained based on the linear regression implemented in tensorQTL[74]) as a function of allelic fold change (aFC) at $\log_2$ scale of eQTL in the liver (n=741). The $\log_2$aFC was estimated with the aFC script[75] **(f)** The proportion of eQTL detected (*y*-axis) with different effect sizes (from left to right panels) as a function of tissue sample size (*x*-axis)[15]. Dot color legend is the same as the panel (d), shown at the right side. **(g)-(i)** Down-sampling analyses of eGene and eQTL. We carried out down-sampling analyses (10 replications at each sample size) in the liver and muscle, which have the largest sample size among all 28 tissues in the ChickenGTEx. The panel (g) depicts the number of eGenes (left *y*-axis) and mean eQTL per eGene (right *y*-axis) detected at different sample size. The middle panel (h) shows the proportion of detected eQTL of large (absolute $\log_2$aFC ≥ 1, left

655 *y*-axis) and small effect size (absolute $\log_2$aFC ≤ 0.25, right *y*-axis). The right panel (i) presents
656 the number eGenes detected when the regulatory effect size of lead eQTL is large (absolute
657 $\log_2$aFC ≥ 1, left *y*-axis) and small (absolute $\log_2$aFC ≤ 0.25, right *y*-axis). The error bars indicate
658 the standard errors across 10 replications.
659
660 **Extended Data Fig. 4. Comparison of eQTL mapping between RNA-Seq imputed and whole-**
661 **genome sequence (WGS) called genotypes in 308 chickens**. RNA-Seq imputed genotypes were
662 generated with the ChickenGTEx genotype reference panel comsisting of of 2,869 WGS using
663 Beagle 5.1 software[76]. **(a)** Venn diagram showing the overlap of eGenes detected using imputed
664 and observed genotypes (i.e., those directly called from whole-genome sequences, WGS). **(b)**
665 Comparison of significance (-$\log_{10}$) between eGenes detected from observed (y-axis) and imputed
666 genotypes (x-axis) with the linear regression. **(c)** Effect size (i.e., slope) of lead variants between
667 eGenes detected by both imputed and observed genotypes (both, n=10,047), by imputed genotypes
668 only (Imputed SNPs only, n=380), by observed genotypes only (Observed SNPs only, n=571), and
669 non-eGenes in both approaches (Neither, n=3,120). The central band in the boxplot represents the
670 median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 ×
671 the interquartile range. The statistical significances were obtained with two-sided Student's *t*-test.
672 **(d)** Venn diagram displaying the overlap of lead variants (i.e., SNPs with the highest significance
673 in eQTL mapping) identified by imputed genotypes and those by observed genotypes. **(e)**
674 Comparison of distribution of lead variants detected by imputed and observed genotypes around
675 gene transcription start sites (TSS). **(f)** Correlation of effect sizes for lead variants of eGenes
676 identified by imputed and observed genotypes. Same lead: the lead variants of the shared eGenes
677 in panel (a) are the same SNPs. Diff lead: the lead variants of the shared eGenes in panel (a) are
678 different SNPs. *P* values were computed *via* the asymptotic *t* approximation. **(g)** Linkage
679 disequilibrium (LD, r^2) of different lead variants of same eGenes (2,307) detected by imputed and
680 observed genotypes. The "Distance-matched random set" represents the same number of paired
681 SNPs with similar physical distances to the "lead variants" category. The central band in the
682 boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the
683 whiskers extend 1.5 × the interquartile range. The significance was obtained with two-sided
684 Student's t-test. **(h)** Percentage of eGenes detected as a function of LD (r^2) between two different
685 lead variants detected by observed and imputed genotypes. Exact number of eGenes are depicted
686 on each bar. **(i)** Functional enrichment ($\log_2$Fold change, mean ± s.d.) of eQTL obtained from
687 observed genotypes versus those from imputed genotypes across chromatin states[14]. The error bars
688 indicate the standard errors of enrichment fold in three central nueral system-related tissues
689 (hypothalamus, cortex and cerebellum) [14]. TssA: strongly active promoters/transcripts, TssAHet:
690 flanking active TSS without ATAC, TxFlnk: transcribed at gene, TxFlnkWk: weak transcribed at
691 gene, TxFlnkHet: transcribed region without ATAC, EnhA: strong active enhancer, EnhAMe:
692 medium enhancer with ATAC, EnhAWk: weak active enhancer, TxFlnkWk: active enhancer no
693 ATAC (hetero), EnhPois: poised enhancer, ATAC_Is: ATAC island, TssBiv: bivalent/poised TSS,
694 Repr: repressed polycomb, ReprWk: weak repressed polycomb. **(j)** Venn diagram depicting the
695 overlap of eGene-independent SNP pairs using imputed and observed genotypes. **(k)** Number of
696 eGenes (y-axis) with different number of credible sizes (x-axis) identified using SuSiE-inf[77] fine-
697 mapping approach. **(l)** Percentage of fine-mapped variants using observed and imputed genotypes
698 as a function of different posterior inclusion probability (x-axis). Manhattan plots display the
699 significance (-$\log_{10}$) of eQTL mapping with the linear regression as a function of genomic position
700 for imputed (bottom) and observed SNPs (top) of three eGenes: *EPHB2* **(m),** *UBXN6* **(n),** and
701 *ENSGALG00000006465* (o). The lead variants of *EPHB2* were identical in both approaches, while

UBXN6 had different lead variants but with high LD (r^2 = 0.98). The two different lead variants of *ENSGALG00000006465* detected by observed and imputed genotypes were not in LD.

Extended Data Fig. 5. eQTL conservation and validation. **(a)** Correlations of lead variant effects (upper triangle) and *cis*-h^2 (lower triangle) between chicken and mammals, including cattle, pig and human. The chicken eGenes show lower correlations with that of mammals, while within mammals higher correlations were observed. Similar observations for *cis*-h^2. The species tree is retrieved from the Timetree5 (https://timetree.org/) and is visualized with iTOL (https://itol.embl.de)[78]. *P* values were computed *via* the asymptotic *t* approximation. **(b)** Internal validation of eQTL. Bars in light blue indicate the Spearman correlation coefficient of eQTL effect size between validation and discovery groups (left *y*-axis), and red dots represent π_1 statistic estimating the replication rate of eQTL between groups (right *y*-axis). The samples in each of the 15 tissues with over 100 individuals are evenly and randomly divided into two groups, i.e., discovery and validation groups. The tissue color legend is shown at the right side. **(c)** Correlation between effect size of eQTL (*x*-axis, n=2,396) and those of same loci derived from allele-specific expression (ASE, *y*-axis) analysis in the liver. *P* values were computed *via* the asymptotic *t* approximation. **(d)** The proportion of regulatory variants predicted by DeepSEA[29] (prediction score > 0.7) based on 310 functional profiles in chickens (**Table S8**). molQTL_set: conditionally independent molQTL across tissues; Random_set: randomly selected variants with the same MAF as molQTL; Background: all tested 1.5 million variants. *** $P <$ 0.001, obtained with the Chi-squared test.

Extended Data Fig. 6. External validation of eQTL. **(a)** linear model (LM)-based validation. **(b)** linear mixed model (LMM)-based validation. The validation is carried out in three tissues: hypothalamus (upper row), liver (middle row) and pituitary (bottom row). The ChickenGTEx is the discovery population, and the external validation population is the commercial White Plymouth Rock chickens. *P* values were computed *via* the asymptotic *t* approximation. **(c)** The comparison of SNPs in the ChickenGTEx discovery population and validation population. These SNPs are common (MAF > 0.05) in both discovery and validation populations**. (d)** The number of SNPs used in π1 calculation shown in panel (a). **(e)** The number of samples in ChickenGTEx discovery population and validation population. **(f)** The number of eGenes detected in the ChickenGTEx discovery population and validation. **(g)** Effect size of replicated and not replicated eQTLs across tissues. "Replicated" represents the set of significant SNPs in the discovery population that pass the significant threshold in the validation population. The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range. The significance was obtained with two-sided Student's t-test. **(h)** The π1 value (y-axis) plotted as a function of eQTL effect size (x-axis) obtained from the discovery population in the liver. **(i)** Histogram depicting the distribution of eQTL nominal *P* values in the validation population. For each eGene, the nearest variants in the validation population to the corresponding lead variants in the discovery population were selected. **(j)** Histogram showing the distribution of eQTL nominal *P* values in the validation population. For each eGene identified in the discovery population, we selected the top lead variant within the same LD block as the eQTL in the discovery population.

Extended Data Fig. 7. Functional enrichment of molecular quantitative trait loci (molQTL). (a) Fold enrichment (mean ± s.d.) of molQTL in strong enhancers (E6) and super-enhancers. The error bar indicated standard errors of enrichment across 17 tissues common in this study and Pan et al. (2023)[14]. **(b)** Fraction of eGene-eVariant pairs overlapped with

regulatory elements-target gene pairs, which were retrieved from Pan et al. (2023)[14]. The overlap required the target genes of regulatory elements were eGenes and eVariants were located within regulatory elements. **(c)** Percentage of eGene-eVariant pairs located within the same topologically associating domains (TAD) predicted from Hi-C data in the muscle, liver and testis [31]. **(d)** Enrichment (at $\log_2$ transformed odds ratio) of ePhenotype-molQTL pairs within the same Hi-C TAD as a function of different distances of molQTL TSS. The error bar indicated standard errors of enrichment across 28 tissues. ePhenotype: molecular phenotypes that are regulated by at least one genetic variant. The odds ratio was obtained by fitting the linear model: SameTAD = eQTL + |TSS distance| + eQTL* |TSS distance|, where SameTAD represents whether the pair of eGene-eVariant is located within the same TAD (coded as 1) or not (coded as 0). The symbols – and + means upstream and downstream of eVariants to TSS, respectively. The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range. **(e)** Manhattan plot displaying SNP associations with *TIMM17B* gene expression specifically in the brain using the linear regression. The top SNP (i.e., the one with the highest probability, rs317368746) was highlighted with a diamond. The bottom barplot panel depicts the annotation of regulatory elements retrieved from Pan et al. (2023)[14], by which we observed the top SNP was located within an enhancer annotated specifically in brain-related tissues (e.g. cortex, cerebellum, hypothalamus). The color legend of regulatory elements can be accessed in Pan et al. (2023)[14] and found at the right corner. TssA: strongly active promoters/transcripts, TssAHet: flanking active TSS without ATAC, TxFlnk: transcribed at gene, TxFlnkWk: weak transcribed at gene, TxFlnkHet: transcribed region without ATAC, EnhA: strong active enhancer, EnhAMe: medium enhancer with ATAC, EnhAWk: weak active enhancer, TxFlnkWk: active enhancer no ATAC (hetero), EnhPois: poised enhancer, ATAC_Is: ATAC island, TssBiv: bivalent/poised TSS, Repr: repressed polycomb, ReprWk: weak repressed polycomb.

Extended Data Fig. 8. Tissue-sharing properties of eQTL. Absolute effect size (allelic fold change, aFC) of eQTL **(a)**, distance of eQTL to TSS **(b)**, and minor allele frequency **(c)** as a function of number of tissues where the eGene is expressed in. The black line is corresponding median estimates, and the grey shades indicate corresponding interquartile ranges. Correlation tests were carried out using *cor.test* function, and *P* values were computed *via* the asymptotic *t* approximation in R v3.6.3. **(d)** Fraction of eQTL around transcription start site (TSS) according to number of tissues they are active in. **(e)** Fold enrichment ($\log_2$ scaled, mean ± s.d., y-axis) of tissue-specific and -shared eQTL based on MashR analysis (as shown in **Fig. 4a**) across 15 chromatin states. The error bar indicated standard errors of enrichment across 23 tissues in Pan et al. (2023)[14]. E1–E15 represent chromatin states, defined as follows: TssA: strongly active promoters/transcripts, TssAHet: flanking active TSS without ATAC, TxFlnk: transcribed at gene, TxFlnkWk: weak transcribed at gene, TxFlnkHet: transcribed region without ATAC, EnhA: strong active enhancer, EnhAMe: medium enhancer with ATAC, EnhAWk: weak active enhancer, TxFlnkWk: active enhancer no ATAC (hetero), EnhPois: poised enhancer, ATAC_Is: ATAC island, TssBiv: bivalent/poised TSS, Repr: repressed polycomb, ReprWk: weak repressed polycomb. **(f)** The Manhattan plot for eQTL mapping of *ALG3* in the liver, muscle, brain and blood. Color legend represents the degree of LD (r^2) between the top SNP and the rest**. (g)** and **(h)** are for the comparison of blood-specific eQTL (Specific, n = 11,884) and the rest of blood eQTL (Common, n = 9,089) in terms of Minor Allele Frequency (MAF) and effect size, respectively. The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range. The two-tailed Student's t-test was applied to obtain statistical significance. **(i)** The top panel shows that the

effect of lead variant *rs315639985* on *FBXO5* in the blood is opposite to that of *rs312482960* in the spleen. These two SNPs are in high LD (r^2) of 0.92 in the blood and 0.83 in the spleen. The bottom panel shows that the effect of lead variant *rs313608694* on *ELAC2* in the embryo is opposite to that in the spleen. **(j)** Replication (π_1) of eQTL in liver between breeds, including Ross (n = 101), Leghorn (n = 74), Cobb (n = 47) and RIR (Rhode Island Red, n = 78). We carried out eQTL mapping in each breed separately and computed π_1 replication rates based on commonly discovered eQTL between two breeds. **(k)** Heatmap depicting of eQTL effect sharing between breeds. This analysis was done by using MashR[34]. **(l)** Expression (TMM, Trimmed Means of M values) of *PRKCDBP (ENSGALG00000028174)* regulated by *rs314795649* genotypes consistent in all four breeds (n = 286 samples) being studied. The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range.

Extended Data Fig. 9. Context-interaction eQTL. Examples of sex-biased eGene *TCLF5* **(a)**, *CPS1* **(b)** and *SNAI2* **(c),** respectively, in liver (n=137). The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range. **(d)** Dot plots of eGene *ATP6V1A* expression against the transcription factor *TCF25* expression (*y*-axis) across three genotypes of *rs313600592*. The expression of *ATP6V1A* is significantly affected by the interaction between *TCF25* expression and *rs313600592* genotypes. **(e)** Number of cell type interaction QTL (ci-eQTL) detected in each tissue-cell type combination (FDR < 5%). **(f)** Distance of eQTL (red color) and ci-eQTL (blue color) to transcript start site (TSS). **(g)** Fold enrichment (mean ± s.d.) of eQTL and ci-eQTL in 15 chromatin states. Fold enrichment is shown as mean (dot) ± standard deviation ($\log_2$ scaled, error bar), which was obtained with the enrichment tested in 5 tissues (liver, muscle, heart, bursa and spleen)[14]. TssA: strongly active promoters/transcripts, TssAHet: flanking active TSS without ATAC, TxFlnk: transcribed at gene, TxFlnkWk: weak transcribed at gene, TxFlnkHet: transcribed region without ATAC, EnhA: strong active enhancer, EnhAMe: medium enhancer with ATAC, EnhAWk: weak active enhancer, TxFlnkWk: active enhancer no ATAC (hetero), EnhPois: poised enhancer, ATAC_Is: ATAC island, TssBiv: bivalent/poised TSS, Repr: repressed polycomb, ReprWk: weak repressed polycomb. **(h)** Histogram depicting the distribution of linkage disequilibrium (LD) between lead variants of eQTL and ci-eQTL when they have the target regulated genes. The vertical read line indicated the median LD (r^2=0.02). **(i)** A ci-eQTL regulating the *MANSC1* gene in dendritic cells from the muscle. The corresponding bulk tissue eQTL (depicted at the right side) tested in 517 samples. The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range. **(j)** The heatmap of Spearman's correlation of ci-eQTL effect size between cell type-tissue combinations. Tissues are clustered on the basis of dissimilarities (i.e. 1-d), where d is Euclidean distance calculated from the ci-eQTL effect, with a complete linkage method[65]. The color legend of tissues and cell types are shown at the bottom.

Extended Data Fig. 10. Cross-species meta-TWAS analysis. (a) Miami plot[79] showing associations of gene expression and chicken body weight *via* single-species TWAS (upper panel) in chickens and cross-species (chicken, pig, and human) meta-TWAS (bottom panel) in the muscle tissue. TWAS: transcriptome-wide association study. The cross-species meta-TWAS analysis is conducted based on chicken body weight TWAS in the muscle and the corresponding TWAS data of mammals' growth-related traits in the muscle. Physiologically similar traits to chicken body weight were defined arbitrarily as growth-related traits, such as backfat thickness and average daily

gain (ADG) in pigs, and body weight and height in humans. Non-physiologically similar traits included growth-unrelated phenotypes, such as the number of mummified pigs and number of weaned piglets in pigs, and type 2 diabetes and heart failure in humans. Quantile–quantile plots were depicted at their respective right side, drawn by the qqman package[70]. **(b)** Nominal TWAS associations of 14 genes detected by cross-species meta-TWAS of the panel (a) in independent chicken populations[50]. Phenome-wide associations (i.e. PheWAS) of the *GOLGA3* gene with pig traits **(c)** and human traits **(d)**. *y*-axis shows the negative base-10 logarithm of the *P*-value, and *x*-axis indicates traits tested. Each dot represents a trait, colored by trait categories, the red dash line indicated the FDR[69]-based threshold (0.05). The pig PheWAS was done by the PigBiobank (http://pigbiobank.farmgtex.org), and the human PheWAS was done by the GWAS atlas (https://atlas.ctglab.nl/).

References

1. Hillier, L. W. & Miller, W. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
2. Burt, D. W. Emergence of the chicken as a model organism: implications for agriculture and biology. *Poultry Science* **86**, 1460–1471 (2007).
3. Beacon, T. H. & Davie, J. R. The chicken model organism for epigenomic research. *Genome* **64**, 476–489 (2021).
4. Garcia, P., Wang, Y., Viallet, J. & Macek Jilkova, Z. The Chicken Embryo Model: A Novel and Relevant Model for Immune-Based Studies. *Frontiers in Immunology* **12**, (2021).
5. Wright, D. *et al.* The genetic architecture of domestication in the chicken: effects of pleiotropy and linkage. *Molecular Ecology* **19**, 5140–5156 (2010).
6. Flores-Santin, J. & Burggren, W. W. Beyond the Chicken: Alternative Avian Models for Developmental Physiological Research. *Front Physiol* **12**, 712633 (2021).
7. Brown, W. R. A., Hubbard, S. J., Tickle, C. & Wilson, S. A. The chicken as a model for large-scale analysis of vertebrate gene function. *Nat Rev Genet* **4**, 87–98 (2003).
8. Wu, Z. *et al.* Heterogeneity of a dwarf phenotype in Dutch traditional chicken breeds revealed by genomic analyses. *Evolutionary Applications* **14**, 1095–1108 (2021).
9. Wang, M.-S. *et al.* An Evolutionary Genomic Perspective on the Breeding of Dwarf Chickens. *Molecular Biology and Evolution* **34**, 3081–3088 (2017).
10. van der Eijk, J. A. J. *et al.* Chicken lines divergently selected on feather pecking differ in immune characteristics. *Physiology & Behavior* **212**, 112680 (2019).
11. Lillie, M. *et al.* Genomic signatures of 60 years of bidirectional selection for 8-week body weight in chickens. *Poultry Science* **97**, 781–790 (2018).
12. Smith, J. *et al.* Fourth Report on Chicken Genes and Chromosomes 2022. *Cytogenetic and Genome Research* **162**, 405–528 (2023).
13. Kern, C. *et al.* Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. *Nat Commun* **12**, 1821 (2021).
14. Pan, Z. *et al.* An atlas of regulatory elements in chicken: A resource for chicken genetics and genomics. *Science Advances* **9**, eade1204 (2023).
15. The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318–1330 (2020).
16. Aguet, F. *et al.* Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
17. The GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).

18. Liu, S. & Fang, L. The CattleGTEx atlas reveals regulatory mechanisms underlying complex traits. *Nat Genet* **54**, 1273–1274 (2022).
19. Liu, S. *et al.* A multi-tissue atlas of regulatory variants in cattle. *Nat Genet* **54**, 1438–1447 (2022).
20. Teng, J. *et al.* A compendium of genetic regulatory effects across pig tissues. *Nat Genet* **56**, 112–123 (2024).
21. Ellegren, H. *et al.* Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. *BMC Biology* **5**, 40 (2007).
22. Nicholas, F. W. Online Mendelian Inheritance in Animals (OMIA): a comparative knowledgebase of genetic disorders and other familial traits in non-laboratory animals. *Nucleic Acids Research* **31**, 275–277 (2003).
23. Wang, Z. *et al.* An EAV-HP Insertion in 5′ Flanking Region of SLCO1B3 Causes Blue Eggshell in the Chicken. *PLOS Genetics* **9**, e1003183 (2013).
24. McKenna, A. *et al.* The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
25. Guan, L., Yang, Q., Gu, M., Chen, L. & Zhang, X. Exon expression QTL (eeQTL) analysis highlights distant genomic variations associated with splicing regulation. *Quant. Biol.* **2**, 71–79 (2014).
26. Li, Y. I. *et al.* Annotation-free quantification of RNA splicing using LeafCutter. *Nat Genet* **50**, 151–158 (2018).
27. Cui, R. *et al.* Improving fine-mapping by modeling infinitesimal effects. *Nat Genet* **56**, 162–169 (2024).
28. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences* **100**, 9440–9445 (2003).
29. Zhou, J. & Troyanskaya, O. G. Predicting effects of noncoding variants with deep learning–based sequence model. *Nat Methods* **12**, 931–934 (2015).
30. Robins, C. *et al.* Genetic control of the human brain proteome. *The American Journal of Human Genetics* **108**, 400–410 (2021).
31. Guan, D. *et al.* Profiling chromatin contacts at micro-scale in the chicken genome. Plant and Animal Genome 30 (2023) .
32. Noda, D. *et al.* ELAC2, a putative prostate cancer susceptibility gene product, potentiates TGF-β/Smad-induced growth arrest of prostate cells. *Oncogene* **25**, 5591–5600 (2006).
33. Hu, H. *et al.* AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. *Nucleic Acids Research* **47**, D33–D38 (2019).
34. Urbut, S. M., Wang, G., Carbonetto, P. & Stephens, M. Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nat Genet* **51**, 187–195 (2019).
35. Wang, Y. *et al.* Genetic Dissection of Growth Traits in a Unique Chicken Advanced Intercross Line. *Front. Genet.* **11**, (2020).
36. Hukku, A., Sampson, M. G., Luca, F., Pique-Regi, R. & Wen, X. Analyzing and reconciling colocalization and transcriptome-wide association studies from the perspective of inferential reproducibility. *The American Journal of Human Genetics* **109**, 825–837 (2022).
37. Zhu, Z. *et al.* Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet* **48**, 481–487 (2016).
38. Gamazon, E. R. *et al.* A gene-based association method for mapping traits using reference transcriptome data. *Nat Genet* **47**, 1091–1098 (2015).
39. Barbeira, A. N. *et al.* Integrating predicted transcriptome from multiple tissues improves association detection. *PLOS Genetics* **15**, e1007889 (2019).

40. Barbeira, A. N. *et al.* Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun* **9**, 1825 (2018).
41. Barbeira, A. N. *et al.* Exploiting the GTEx resources to decipher the mechanisms at GWAS loci. *Genome Biology* **22**, 49 (2021).
42. Mostafavi, H., Spence, J. P., Naqvi, S. & Pritchard, J. K. Systematic differences in discovery of genetic effects on gene expression and complex traits. *Nat Genet* **55**, 1866–1875 (2023).
43. Connally, N. J. *et al.* The missing link between genetic association and regulatory function. *eLife* **11**, e74970 (2022).
44. Sowa, A. S. *et al.* Karyopherin α-3 is a key protein in the pathogenesis of spinocerebellar ataxia type 3 controlling the nuclear localization of ataxin-3. *Proceedings of the National Academy of Sciences* **115**, E2624–E2633 (2018).
45. Cao, J. *et al.* Effect of combinations of monochromatic lights on growth and productive performance of broilers. *Poultry Science* **91**, 3013–3018 (2012).
46. Pan, J., Yang, Y., Yang, B. & Yu, Y. Artificial Polychromatic Light Affects Growth and Physiology in Chicks. *PLOS ONE* **9**, e113595 (2014).
47. Schob, C. *et al.* Dominant KPNA3 Mutations Cause Infantile-Onset Hereditary Spastic Paraplegia. *Annals of Neurology* **90**, 738–750 (2021).
48. Santhanam, N. *et al.* RatXcan: A framework for cross-species integration of genome-wide association and gene expression data. 2022.06.03.494719 Preprint at https://doi.org/10.1101/2022.06.03.494719 (2024).
49. Naqvi, S. *et al.* Conservation, acquisition, and functional impact of sex-biased gene expression in mammals. *Science* **365**, eaaw7317 (2019).
50. Li, J. *et al.* Genome-wide association studies for egg quality traits in White Leghorn layers using low-pass sequencing and SNP chip data. *J Anim Breed Genet* **139**, 380–397 (2022).
51. Qi, T. *et al.* Genetic control of RNA splicing and its distinct role in complex trait variation. *Nat Genet* **54**, 1355–1363 (2022).
52. Li, L. *et al.* An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. *Nat Genet* **53**, 994–1005 (2021).
53. Munro, D. *et al.* Multimodal analysis of RNA sequencing data powers discovery of complex trait genetics. *Nat Commun* **15**, 10387 (2024).
54. Guan, D. *et al.* Prediction of transcript isoforms in 19 chicken tissues by Oxford Nanopore long-read sequencing. *Frontiers in Genetics* **13**, (2022).
55. Kuo, R. I. *et al.* Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human. *BMC Genomics* **18**, 323 (2017).
56. Thomas, S., Underwood, J. G., Tseng, E., Holloway, A. K. & Subcommittee, on behalf of the B. T. B. C. I. Long-Read Sequencing of Chicken Transcripts and Identification of New Transcript Isoforms. *PLOS ONE* **9**, e94650 (2014).
57. Zhang, J. *et al.* Comprehensive analysis of structural variants in chickens using PacBio sequencing. *Frontiers in Genetics* **13**, (2022).
58. Kojima, S. *et al.* Mobile element variation contributes to population-specific genome diversification, gene regulation and disease risk. *Nat Genet* **55**, 939–951 (2023).
59. Wicker, T. *et al.* The repetitive landscape of the chicken genome. *Genome Res.* **15**, 126–136 (2005).
60. Li, M. *et al.* De Novo Assembly of 20 Chicken Genomes Reveals the Undetectable Phenomenon for Thousands of Core Genes on Microchromosomes and Subtelomeric Regions. *Molecular Biology and Evolution* **39**, msac066 (2022).

61. Morris, J. A. *et al.* Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. *Science* **380**, eadh7699 (2023).
62. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *The American Journal of Human Genetics* **81**, 559–575 (2007).
63. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLOS Genetics* **17**, e1009440 (2021).
64. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* **6**, 80–92 (2012).
65. Dodge. *The Concise Encyclopedia of Statistics*. (Springer, New York, NY, 2008). doi:10.1007/978-0-387-32833-1_71.
66. Vavrek, M. fossil: palaeoecological and palaeogeographical analysis tools. https://palaeo-electronica.org/2011_1/238/index.html.
67. Newman, A. M. *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* **37**, 773–782 (2019).
68. Conway, J. R., Lex, A. & Gehlenborg, N. UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics* **33**, 2938–2940 (2017).
69. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)* **57**, 289–300 (1995).
70. Turner, S. D. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software* **3**, 731 (2018).
71. Wu, T. *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).
72. Hao, Z. *et al.* RIdeogram: drawing SVG graphics to visualize and map genome-wide data on the idiograms. *PeerJ Comput. Sci.* **6**, e251 (2020).
73. Kryuchkova-Mostacci, N. & Robinson-Rechavi, M. A benchmark of gene expression tissue-specificity metrics. *Briefings in Bioinformatics* **18**, 205–214 (2017).
74. Taylor-Weiner, A. *et al.* Scaling computational genomics to millions of individuals with GPUs. *Genome Biology* **20**, 228 (2019).
75. Mohammadi, P., Castel, S. E., Brown, A. A. & Lappalainen, T. Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change. *Genome Res.* **27**, 1872-1884 (2017).
76. Browning, B. L., Tian, X., Zhou, Y. & Browning, S. R. Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics* **108**, 1880–1890 (2021).
77. Cui, R. *et al.* Improving fine-mapping by modeling infinitesimal effects. *Nat Genet* **56**, 162–169 (2024).
78. Letunic, I. & Bork, P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* **52**, W78–W82 (2024).
79. White, J. D. Miamiplot: An R package for creating ggplot2 based miami plots. (2020). https://github.com/juliedwhite/miamiplot

Online Methods

Ethics

It is not applicable because no biological samples were collected and no animal handling was performed for this study.

RNA-Seq data analyses and molecular phenotype definition

We downloaded 8,338 RNA-Seq data sets from the Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra) and 140 public data sets from the Genome Sequence Archive (GSA, https://ngdc.cncb.ac.cn/gsa/). We also included 155 newly-generated RNA-Seq data sets. The metadata relating to all the RNA-Seq samples is summarized in **Table S1**. For quality control, we removed adaptors and trimmed low-quality reads using Trim Galore (v0.6.6, https://github.com/FelixKrueger/TrimGalore) with options of "--gzip --trim-n --length 30 --clip_R1 3 --clip_R2 3 --three_prime_clip_R1 3 --three_prime_clip_R2 3". We aligned the clean reads to the GRCg6a reference genome (Ensembl version 102) using STAR (v2.7.7a)[80] with parameters of "--quantMode GeneCounts --chimSegmentMin 10 --chimOutType Junctions --chimOutJunctionFormat 1 --outFilterMismatchNmax 3". For downstream analyses, only 7,015 samples with uniquely mapping rates ≥ 60 % and a number of clean reads > 500,000 after removing potentially mislabeled samples were kept. For each of these samples, we then obtained raw read counts and normalized expression (i.e., Transcripts Per Million, TPM) of 16,779 PCGs annotated in the Ensembl v102 and 22,792 lncRNA genes annotated by FR-AgENCODE (http://www.fragencode.org/)[81], using featureCounts (v2.0.1)[82] and StringTie (v2.1.5)[83], respectively. Using the same software[82], we counted the total number of reads as a function of annotated exons, which were further transformed into TPM using TBtools v1.09[84]. We performed the tree clustering of all the RNA-Seq samples using the GGTREE package[85]. The distance between samples was measured by 1-*r*, where *r* was Pearson's correlation coefficient based on the $\log_2$(TPM+0.25) of 5,000 genes with the highest variability. We also visualized these samples using the *t*-distributed stochastic neighbor embedding (*t*-SNE) approach implemented in the Rtsne package[86].

We quantified alternative splicing variation from RNA-Seq data using the LeafCutter package[26], which took into account spliced reads so that both novel and known alternative splicing events could be identified and quantified[26]. Briefly, based on the STAR alignments mentioned above, we extracted junctions and defined intron clusters across samples using the script "bam2junc.sh" and "leafcutter_cluster.py", respectively, as provided by the LeafCutter package[26]. For intron clustering, we required at least 30 split reads supporting each cluster and at least 0.1% of reads supporting a junction in a cluster, as well as allowing intron length of up to 500kb. The generated matrix of per-individual counts was normalized and used for clustering samples based on 1-*r*, where *r* is the Pearson's correlation coefficient between samples. To link intron clusters to genes, we mapped their coordinates to the gene model provided by the FR-AgENCODE database [81] using the script "map_clusters_to_genes.R" (https://github.com/broadinstitute/gtex-pipeline)[15]. Afterward, we filtered out introns, by following the human GTEx's pipeline[15], if no reads were detected in >50% of samples or the number of counts was less than max(10, 0.1n) where n is the sample size. In addition, we discarded introns with low variability across samples: $\sum_i(|z_i| < 0.25) \geq n-3$ and $\sum_i(|z_i| > 6) \leq 3$, where z_i is the z-score of the ^{i}th cluster read fraction across individuals. The filtered counts were further normalized between samples using the script "prepare_phenotype_table.py" in the LeafCutter package[26]. The generated normalized splicing counts were stored in BED formatted file for subsequent sQTL mapping.

For the quantification of 3'UTR APA, we utilized the DaPars (v2)[87]. We first extracted distal polyadenylation sites based on the Ensembl annotation (v102) using the script "DaPars_Extract_Anno.py". Then, we computed the genome coverage of STAR alignments mentioned above using the *genomecov* function in the BEDTools (v2.30.0)[88]. The generated wiggle alignment files were then used for quantifying APA usage, resulting in the percentage of distal poly(A) site usage index (PDUI) value for each gene in each sample. We rescaled the PDUI values across samples to the mean of zero and variance of one in each tissue for 3a'QTL mapping.

SNP calling from RNA-Seq samples

To call SNPs from RNA-Seq samples, we marked PCR duplicates in STAR alignment files and split reads that contained Ns in their cigar string using *MarkDuplicates* and *SplitNCigarReads* modules of the Genome Analysis Toolkit (GATK, v 4.1.9.0)[24], respectively. Using the Ensembl dbSNP database (v102), we recalibrated base quality scores using GATK *BaseRecalibrator* and *ApplyBQSR* modules. By following the best practice of germline variant calling from RNA-Seq data, we detected small variants from the recalibrated alignments files, which generated individual Genomic Variant Call Format (GVCF) files using the *HaplotypeCaller* function of the GATK tool[24]. Then, we carried out joint-calling of all GVCF samples using the *GenotypeGVCFs* module from the GATK tool[24]. For selecting high-quality SNPs, we carried out a hard-filtering with criteria of "FS > 30.0 & QD < 2.0", resulting in a total set of 12,191,306 SNPs.

Construction of the multi-breed genotype imputation panel and genotype imputation

We retrieved 1,693 public WGS data sets from SRA (n=1,213) and GSA (n = 480) databases along with 1,176 additional newly generated WGS samples, resulting in a total set of 2,869 WGS samples (**Table S5**). All raw sequence reads passed a uniform computational pipeline, including adaptor removal, read alignment, and SNP calling. Briefly, we trimmed read adaptors and low-quality reads using the Trimmomatic v0.39 software[89]. The obtained clean reads were further aligned against the Ensembl GRCg6a chicken reference genome (v102) using the MEM algorithm of the Burrows-Wheeler Aligner (BWA, v0.7.17)[90]. The alignment files in Binary Alignment Map (BAM) format were sorted using SAMtools (v1.9)[91], and were further passed for the removal of PCR duplicates using GATK (v4.1.9.0)[24]. The obtained BAM files were then used for variant discovery to generate individual GVCF files using the *HaplotypeCaller* function of the GATK tool[24]. The joint-calling of all 2,869 GVCF samples was further done using the *GenotypeGVCFs* module from the GATK tool[24]. For selecting high-quality SNPs, we carried out a hard-filtering with criteria of "QD < 2.0, MQ < 40.0, FS > 60.0, SOR > 3.0, MQRankSum < -12.5, and ReadPosRankSum < -8.0", resulting in a total set of 117,900,812 clean SNPs. To create the genotype imputation reference panel, we first filtered out multi-allelic and sex chromosomal SNPs, as well as those with MAF < 0.01 and missing rate > 0.9 using BCFtools v1.10.2[92], and then imputed missing genotypes using the Beagle 5.1 program[93]. This yielded the final reference panel consisting of 2,869 samples and 10,520,420 SNP genotypes. To better impute SNPs called from RNA-Seq samples, we discarded SNPs called from RNA-Seq samples with MAF < 0.05 using BCFtools v1.10.2[92] and further evaluated the effect of missing rates decreasing from 0.9 gradually to 0.6 on imputation accuracy. This evaluation revealed that the missing rate of 0.6 could reach >95% of imputation accuracy, yielding a set of 1.5 million SNPs for subsequent analysis. The genotype imputation was performed using the Beagle 5.1 program[93].

Molecular QTL mapping

After removing duplicated samples within a tissue, we estimated covariates and conducted molQTL by following human GTEx standard pipeline[15]. Briefly, we estimated the Probabilistic Estimation of Expression Residuals (PEER) in each of the tissues using the PEER software package[94], and computed principal components using LD-pruned imputed genotypes with the option of "--indep-pairwise 200 100 0.1" using PLINK v1.9[62] and EIGENSOFT v8.0.0 package[95]. C*is*-molQTL of each feature, i.e., SNPs distributed around 1 Mb upstream and downstream of the TSS of the gene, were then mapped using tensorQTL v1.0.4[74]. Initializing with the option of "--mode cis_nominal" of the tensorQTL v1.0.4[74], we calculated all nominal associations of all variant-molecular phenotype pairs. The permutation mode was further used for computing empirical *P*-values for a molecular phenotype using the option of "--mode cis" of the tensorQTL v1.0.4. After carrying out a multiple testing correction based on empirical beta-approximated *P*-values[96] using the FDR approach[69], we defined eGenes , i.e., genes that were significantly regulated by at least one variant (FDR < 0.05). For an eGene, the empirical *P*-value that was closest to an FDR of 0.05 was defined as the genome-wide empirical *P*-value threshold (pt), which was used for defining the gene-level significance threshold using qbeta(pt, beta_shape1, beta_shape2) in R (v3.6.3)[97], where beta_shape1 and beta_shape2 were computed by tensorQTL v1.0.4[74]. The significant molQTL were tested SNPs whose nominal *P*-values were lower than the gene-level significance threshold. Moreover, we carried out molQTL mapping using a standard linear mixed model (LMM) implemented in the rMVP package (v1.0.8)[98], with the genomic relationship matrix (GRM) generated using the VanRaden method[99]. Similar to molQTL mapping via tensorQTL, we included 5 genotype PCs and the top 10 PEER factors to account for technical confounders.

Fine-mapping analysis of molQTL

We employed two strategies for fine-mapping independent variants underlying each molQTL. Firstly, we utilized the stepwise regression procedure for mapping conditionally independent molQTL, as used in other GTEx studies [15,19,20]. This analysis was done by using the tensorQTL v1.0.4 with "--mode cis_independent" option[74]. The conditionally independent molQTL mapping was based on the nominal associations mentioned above and ranked variants. Secondly, we fine-mapped putative causal variants for each molecular phenotype by using the "Sum of Single Effects" (SuSiE) model (v 1.0) [100]. We calculated LD correlations between all tested SNPs of a molecular phenotype from the genotype reference panel and then fine-mapped variants using the SuSiE infinitesimal effect model. The posterior probability of 0.1 was used for identifying putative causal variants and credible sets.

Colocalization analysis between molecular phenotypes

To demonstrate whether two types of molecular phenotypes shared genetic regulatory mechanisms, we determined a set of paired molecular phenotypes that were derived from the same gene. We then ran the *coloc.abf* function in the coloc package[101], which is an Approximate Bayes Factor colocalization analysis for detecting significant genetic variants shared by two molecular phenotypes. The package computed posterior probabilities for: 1) no association with either molecular phenotype (H0); 2) association only with the first molecular phenotype (H1); 3) association only with the second molecular phenotype (H2); 3) association with both molecular phenotype but two independent signals (H3); 4) association with both molecular phenotype and shared signals (H4). Moreover, we calculated the linkage disequilibrium (LD) of two lead SNPs for a pair of shared molecular phenotypes using PLINK v1.9[62].

Tissue- and breed-sharing of molQTL

Tissue-sharing of molQTL. To assess the cross-tissue sharing pattern of molQTL, we used Multivariate Adaptive Shrinkage in R (MashR, v0.2.57)[34] and METASOFT v2.0.0[102], as applied in human GTEx[15,16]. For MashR, we used the z-score (slope/slope_se) of top molQTL for a gene as input. To run the *mash* model, we randomly selected 1 million molQTL-gene pairs from nominal associations being tested across all tissues by tensorQTL and obtained their z-score values. If there were missing z-score values, zero was filled and the corresponding standard error was set to 1e[6]. Local false sign rate (LFSR) was then computed by MashR and an LFSR of 0.05 was considered as the significance threshold to define whether a molQTL was active in a tissue. Pairwise Spearman's correlation of effect size of active molQTL was calculated to evaluate tissue similarity. For METASOFT, we combined all significant molQTL across tissues and computed the z-score as described above. We estimated the m-value, which represented the posterior probability indicating whether a molQTL effect exists in a tissue, using the Markov Chain Monte Carlo (MCMC) method[103]. The m-value threshold was set as 0.7.

Breed-sharing eQTL analysis. We considered the brain (Leghorn, n = 78; Red Jungle Fowl, n = 46; Ross, n = 157), spleen (Leghorn, n = 74; Cobb, n = 43) and liver (Leghorn, n = 60; Cobb, n = 47; Ross, n = 101; Rhode Island Red, n = 78), tissues as they had more than two breeds with sample size > 40. For each breed, we ran eQTL mapping independently using tensorQTL software (v1.0.4). The eQTL sharing was assessed using METASOFT v2.0.0[102], and MashR (v0.2.57)[34], as well as π1 statistic in the qvalue package[28,104]. The METASOFT and MashR were run as described above, and the π1 statistic (i.e. replication rate)[104] was used to assess if an eQTL detected in one breed can be replicated in another breed.

Integrating molQTL with GWAS results

GWAS summary statistics. To investigate the regulatory mechanisms underpinning complex traits in chickens, we systematically integrated the identified molQTL with GWAS from 39 complex traits of economic importance, representing five trait domains (i.e., growth and development, carcass, egg production, feed efficiency and blood biochemical index). Detailed information for each GWAS is shown in **Table S12**. To perform the integrative analysis of GWAS and molQTL, we overlapped significant GWAS loci with the 1,522,091 SNPs were tested in the molQTL mapping analysis, resulting in 579 GWAS loci.

Transcriptome-wide association study (TWAS). We conducted single- and multi-tissue TWAS with S-PrediXcan[40] and S-MultiXcan[39] included in the MetaXcan (v0.6.11) family, respectively. Briefly, we trained the Nested Cross validated Elastic Net models with molecular phenotypes (i.e., PCG, lncRNA, splicing, exon, and 3a'Genes) and corresponding SNPs within the 1Mb *cis*-window of molecular phenotypes in all 28 tissues. The predictive models with cross-validated correlation $\rho > 0.1$ and prediction performance $P < 0.05$ were selected for subsequent analyses. Using the S-PrediXcan tool and trained models, we predicted gene-trait associations at the single-tissue level, i.e., single-tissue TWAS results. Further, using the S-MultiXcan tool, we integrated single-tissue predictions, generating the multiple-tissue TWAS results. After carrying out a multiple testing correction with the FDR approach[69], gene-trait associations with corrected-$P < 0.05$ were considered as significant.

Summary-based Mendelian Randomization (SMR). To explore the pleiotropic association between molecular phenotypes and a complex trait, we conducted a Mendelian Randomization analysis. This was done by using the SMR software (v1.3.1) [37], which can utilize summary-level data from GWAS and molQTL. To correctly fit the SMR software, the molQTL data generated

by tensorQTL in this study was initially converted into BESD format with options of "--fastqtl-nominal-format --make-besd". We then ran the SMR test and carried out a multiple testing correction with the FDR approach[69]. The gene-trait pairs with corrected *P*-value < 0.05 and HEIDI > 0.05were selected and deemed as significant.

Colocalization analysis. To identify shared genetic variants between GWAS and molQTL, we conducted a colocalization analysis with fastENLOC (v2.0) [105]. We first fine-mapped putative causal variants for each eGene by using a Bayesian multi-SNP genetic association analysis algorithm, deterministic approximation of posteriors (DAP, the current version is DAP-G, v1.0.0)[106,107]. Leveraging the DAP-G (v1.0.0) [106,107] outcome, we generated a probabilistic annotation of molQTL using the "summarize_dap2enloc.pl" script. We then calculated approximate LD blocks using PLINK v1.9 [62] with options: --blocks no-pheno-req --blocks-max-kb 1000 --make-founders. The posterior inclusion probability (PIP) of GWAS loci was calculated for each LD block using TORUS [108] with the options: --load_zval -dump_pip. By integrating GWAS PIP values, we ran the final colocalization analysis with the fastENLOC (v2.0) tool [36] and obtained the gene variant-level colocalization probability (GRCP). The GRCP > 0.5 was defined as the threshold of significance.

Cross-species TWAS comparison

To explore whether conserved nature of gene regulation can help us translate findings between species, we conducted comparative analyses based on single-tissue TWAS results. We first filtered out correlated TWAS traits (Pearson's $r > 0.8$, based on their TWAS z-scores) within each species, resulting in 232 pig traits[20], 37 cattle traits[19], and 37 human traits[15]. Within a shared tissue, we then correlated TWAS results of a chicken complex trait with all these traits on the basis of z score (beta / standard error) estimated from one-to-one orthologous genes between two corresponding species. To define the threshold of significance, we employed the FDR method[69], considering correlations with an FDR < 0.1 as significant[69].

To investigate whether combing TWAS results of physiologically similar traits across species can help to detect more effect genes, we conducted cross-species meta-TWAS analyses by combining TWAS results from different species based on orthologous genes. For meta-TWAS analysis, we applied a sample-size weighting (SSW) strategy[109] by calculating Z_{TWAS} as follows:

$$Z_{TWAS} = \frac{\sum_{i=1}^{B} N_i Z_{TWASj}}{(\sum_{i=1}^{B} N_i^2)^{1/2}}$$

where Z_{TWASj} is the z-score for *j*th gene in TWAS analysis, *i* is the species, *i.e.*,chicken, humans, pigs, and cattle, N_i is the number of individuals for *i*th species in TWAS, *B* is the number of species in meta-TWAS. The effective sample size is $N_i = 4/(\frac{1}{N_{cases}} + \frac{1}{N_{controls}})$. To obtain the significance level, we calculated *P* values for each gene based on a Chi-squared distribution of z-scores (df=1) calculated before. After a multiple testing correction with the FDR method[69] by replacing original *P* value (TWAS) with *P* value (meta-TWAS) of orthologous genes, the threshold of significance was defined as FDR < 0.05.

Statistics and reproducibility

No statistical method was used to predetermine the sample size. The details of data exclusions for each specific analysis are available in the Methods section. The central band in the boxplot represents the median, the box boundaries represent the 25% to 75% percentiles, and the whiskers extend 1.5 × the interquartile range. The experiments were not randomized, as all the datasets are publicly available from observational studies. The investigators were not blinded to

allocation during experiments and outcome assessment, as the data were not from controlled randomized studies. Data normalization and transformation have been carefully performed to ensure that all the data being analyzed met the assumptions of statistical tests used.

Data availability

All raw data analyzed in this study are publicly available for download without restrictions from SRA (https://www.ncbi.nlm.nih.gov/sra/) and NGDC BioProject (https://bigd.big.ac.cn/bioproject/) databases. The GRCg6a chicken reference genome (v102) is available at Ensembl (https://www.ensembl.org). Details of RNA-Seq, WGS, ChIP-Seq peaks and single-cell RNA-Seq can be found in Supplementary Tables, respectively. All processed data, including metadata, the genotype imputation reference panel, molecular phenotypes, imputed genotypes, and the summary statistics of molQTL and GWAS, are available at http://chicken.farmgtex.org.

Code availability

All the computational scripts and codes for RNA-Seq, WGS, single-cell RNA-Seq and Hi-C datasets analyses, as well as the respective quality control, molecular phenotype normalization, genotype imputation, molQTL mapping, functional enrichment, colocalization, SMR and TWAS are available at the FarmGTEx GitHub website (https://github.com/guandailu/ChickenGTEx_pilot_phase).

Methods-only references:

80. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
81. Jehl, F. *et al.* An integrative atlas of chicken long non-coding genes and their annotations across 25 tissues. *Sci Rep* **10**, 20457 (2020).
82. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
83. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology* **20**, 278 (2019).
84. Chen, C. *et al.* TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data. *Molecular Plant* **13**, 1194–1202 (2020).
85. Yu, G., Smith, D. K., Zhu, H., Guan, Y. & Lam, T. T.-Y. ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution* **8**, 28–36 (2017).
86. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).
87. Xia, Z. *et al.* Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. *Nat Commun* **5**, 5274 (2014).
88. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
89. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
90. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at https://doi.org/10.48550/arXiv.1303.3997 (2013).
91. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

92. Danecek, P. *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).
93. Browning, B. L., Zhou, Y. & Browning, S. R. A One-Penny Imputed Genome from Next-Generation Reference Panels. *The American Journal of Human Genetics* **103**, 338–348 (2018).
94. Stegle, O., Parts, L., Piipari, M., Winn, J. & Durbin, R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. *Nat Protoc* **7**, 500–507 (2012).
95. Price, A. L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909 (2006).
96. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
97. R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.
98. Yin, L. *et al.* rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated Tool for Genome-wide Association Study. *Genomics, Proteomics & Bioinformatics* **19**, 619–628 (2021).
99. VanRaden, P. M. Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science* **91**, 4414–4423 (2008).
100. Zou, Y., Carbonetto, P., Wang, G. & Stephens, M. Fine-mapping from summary data with the "Sum of Single Effects" model. *PLOS Genetics* **18**, e1010299 (2022).
101. Giambartolomei, C. *et al.* Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. *PLOS Genetics* **10**, e1004383 (2014).
102. Duong, D. *et al.* Applying meta-analysis to genotype-tissue expression data from multiple tissues to identify eQTLs and increase the number of eGenes. *Bioinformatics* **33**, i67–i74 (2017).
103. Speagle, J. S. A Conceptual Introduction to Markov Chain Monte Carlo Methods. Preprint at https://doi.org/10.48550/arXiv.1909.12313 (2020).
104. Storey JD, Bass AJ, Dabney A & Robinson D. qvalue: Q-value estimation for false discovery rate control. Storey Lab (2022).
105. Wen, X., Pique-Regi, R. & Luca, F. Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. *PLOS Genetics* **13**, e1006646 (2017).
106. Lee, Y., Luca, F., Pique-Regi, R. & Wen, X. Bayesian Multi-SNP Genetic Association Analysis: Control of FDR and Use of Summary Statistics. 316471 Preprint at https://doi.org/10.1101/316471 (2018).
107. Wen, X., Lee, Y., Luca, F. & Pique-Regi, R. Efficient Integrative Multi-SNP Association Analysis via Deterministic Approximation of Posteriors. *The American Journal of Human Genetics* **98**, 1114–1129 (2016).
108. Wen, X. Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. *The Annals of Applied Statistics* **10**, 1619–1638 (2016).
109. Bhattacharya, A. *et al.* Best practices for multi-ancestry, meta-analytic transcriptome-wide association studies: Lessons from the Global Biobank Meta-analysis Initiative. *Cell Genomics* **2**, (2022).

ChickenGTEx Consortium:

Dailu Guan[1], Zhonghao Bai[2], Xiaoning Zhu[3], Conghao Zhong[4], Yali Hou[5], Di Zhu[2,3], Houcheng Li[2], Fangren Lan[4], Shuqi Diao[6], Yuelin Yao[7,8], Bingru Zhao[9], Xiaochang Li[4], Zhangyuan Pan[5], Yahui Gao[6,10,11], Yuzhe Wang[3], Dong Zou[12], Ruizhen Wang[12,13], Tianyi Xu[12], Congjiao Sun[4], Hongwei Yin[14], Jinyan Teng[6], Zhiting Xu[6], Qing Lin[6], Shourong Shi[15], Dan Shao[15], Fabien Degalez[16], Sandrine Lagarrigue[16], Ying Wang[1], Mingshan Wang[17], Minsheng Peng[17], Dominique Rocha[18], Mathieu Charles[18], Jacqueline Smith[19], Kellie Watson[19], Albert Johannes Buitenhuis[2], Goutam Sahana[2], Mogens Sandø Lund[2], Wesley Warren[20], Laurent Frantz[21,22], Greger Larson[23], Susan J. Lamont[24], Wei Si[5,25], Xin Zhao[25], Bingjie Li[26], Haihan Zhang[27], Chenglong Luo[28], Dingming Shu[28], Hao Qu[28], Wei Luo[28], Zhenhui Li[6,29], Qinghua Nie[6,29], Xiquan Zhang[6,29], Ruidong Xiang[30,31,32], Shuli Liu[33], Zhe Zhang[6], Zhang Zhang[12,13], George E. Liu[10], Hans Cheng[34], Ning Yang[4], Xiaoxiang Hu[3], Huaijun Zhou[1], Lingzhao Fang[2]

[1]Department of Animal Science, University of California, Davis, CA, 95616, USA

[2]Center for Quantitative Genetics and Genomics (QGG), Aarhus University, Aarhus, 8000, Denmark

[3]State Key Laboratory of Animal Biotech Breeding, College of Biological Sciences, China Agricultural University, Beijing, 100193, China

[4]College of Animal Science and Technology, China Agricultural University, Beijing, 100193, China

[5]Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing, 100193, China

[6]State Key Laboratory of Livestock and Poultry Breeding, Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding, College of Animal Science, South China Agricultural University, Guangzhou 510642, China

[7]MRC Human Genetics Unit at the Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh EH4 2XU, UK

[8]School of Informatics, The University of Edinburgh, Edinburgh EH8 9AB, UK

[9]Jiangsu Livestock Embryo Engineering Laboratory, College of Animal Science and Technology, Nanjing Agricultural University, Nanjing, Jiangsu 210095, China

[10]Animal Genomics and Improvement Laboratory, Henry A. Wallace Beltsville Agricultural Research Center, Agricultural Research Service, USDA, Beltsville, Maryland 20705, USA

[11]Department of Animal and Avian Sciences, University of Maryland, College Park, Maryland 20742, USA

[12]Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China

[13]University of Chinese Academy of Sciences, Beijing 100049, China

[14]Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Livestock and Poultry Multi-omics of MARA, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, 518124, China

[15]Poultry Institute, Chinese Academy of Agricultural Sciences, Yangzhou, Jiangsu, 225125, China

[16]PEGASE, INRAE, Institut Agro, 35590, Saint Gilles, France.

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

[18]Paris-Saclay University, INRAE, AgroParisTech, GABI, Jouy-en-Josas, 78350, France

[19]The Roslin Institute, Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Midlothian EH25 9RG, UK

[20]Department of Animal Sciences, Data Science and Informatics Institute, University of Missouri, Columbia, MO 65201

[21]Palaeogenomics Group, Department of Veterinary Sciences, Ludwig Maximilian University, Munich 80539, Germany

[22]School of Biological and Behavioural Sciences, Queen Mary University of London, London E1 4DQ, United Kingdom

[23]The Palaeogenomics & Bio-Archaeology Research Network, School of Archaeology, University of Oxford, Oxford, UK

[24]Department of Animal Science, Iowa State University, Ames, Iowa 50011, USA

[25]Department of Animal Science, McGill University, Quebec, H9X 3V9, Canada

[26]Scotland's Rural College (SRUC), Roslin Institute Building, Midlothian EH25 9RG, UK

[27]College of Animal Science and Technology, Hunan Agricultural University, Changsha 410128, China

[28]State Key Laboratory of Swine and Poultry Breeding Industry, Guangdong Key Laboratory of Animal Breeding and Nutrition, Institute of Animal Science, Guangdong Academy of Agricultural Sciences, Guangzhou, 510640, Guangdong China

[29]Guangdong Provincial Key Lab of Agro-Animal Genomics and Molecular Breeding, and Key Laboratory of Chicken Genetics, Breeding and Reproduction, Ministry of Agriculture, College of Animal Science, South China Agricultural University, Guangzhou, Guangdong, China

[30]Agriculture Victoria, Agribio, Centre for AgriBiosciences, Bundoora, VIC 3083, Australia

[31]Cambridge-Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne, 3004, VIC, Australia

[32]School of Agriculture, Food and Ecosystem Sciences, The University of Melbourne, Parkville, 3052, VIC, Australia

[33]School of Life Sciences, Westlake University, Hangzhou, Zhejiang 310024, China

[34]Avian Disease and Oncology Laboratory, USDA, ARS, USNPRC, East Lansing, MI, USA