Title page

The ChickenGTEx pilot analysis: a reference of regulatory variants across 28 chicken tissues

Dailu Guan[1†], Zhonghao Bai[2†], Xiaoning Zhu[3†], Conghao Zhong[4†], Yali Hou[5,6†], The ChickenGTEx Consortium, Fangren Lan[4], Shuqi Diao[7], Yuelin Yao[8,9], Bingru Zhao[10], Di Zhu[2,3], Xiaochang Li[4], Zhangyuan Pan[11], Yahui Gao[7,12,13], Yuzhe Wang[3], Dong Zou[5], Ruizhen Wang[5,6], Tianyi Xu[5], Congjiao Sun[4], Hongwei Yin[14], Jinyan Teng[7], Zhiting Xu[7], Qing Lin[7], 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[11,25], Xin Zhao[25], Bingjie Li[26], Haihan Zhang[27], Chenglong Luo[28], Dingming Shu[28], Hao Qu[28], Wei Luo[28], Zhenhui Li[7,29], Qinghua Nie[7,29], Xiquan Zhang[7,29], Zhe Zhang[7], Zhang Zhang[5,6], George E. Liu[12], Hans Cheng[30], 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]Beijing Institute of Genomics, Chinese Academy of Sciences and China National Center for Bioinformation, Beijing 100101, China

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

[7] 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

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

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

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

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

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

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

[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]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;

Emails for all authors:

dguan@ucdavis.edu; zhonghao.bai@qgg.au.dk; zhu0506@cau.edu.cn; chzhong@cau.edu.cn; houyl@big.ac.cn; fangrlan@163.com; saradiao@126.com; zoud@big.ac.cn; s1914230@ed.ac.uk; zhaobru@163.com; zhudi@qgg.au.dk; lixc1219@126.com; zhypan01@163.com; gyhalvin@gmail.com; yuzhe891@cau.edu.cn; zoud@big.ac.cn;

wangrz@big.ac.cn; xuty@big.ac.cn; cjsun@cau.edu.cn; yinhongwei@caas.cn;
kingyan312@live.cn; zhitingxu@126.com; qing_lin1996@126.com; ssr236@163.com;
yzshaodan@163.com; fabien.degalez@inrae.fr; sandrine.lagarrigue@agrocampus-ouest.fr;
ucywang@ucdavis.edu; wangmingshan@mail.kiz.ac.cn; pengminsheng@mail.kiz.ac.cn;
dominique.rocha@inrae.fr; mathieu.charles@inrae.fr; jacqueline.smith@roslin.ed.ac.uk;
kellie.watson@ed.ac.uk; bart.buitenhuis@qgg.au.dk; mogens.lund@qgg.au.dk;
goutam.sahana@qgg.au.dk; warrenwc@missouri.edu; laurent.frantz@lmu.de;
laurent.frantz@palaeo.vetmed.uni-muenchen.de; greger.larson@arch.ox.ac.uk;
sjlamont@iastate.edu; siwei01@caas.cn; xin.zhao@mcgill.ca; bingjie.li@sruc.ac.uk;
zhhous@163.com; chenglongluo1981@163.com; shudm@263.net; qhw03@163.com;
luowei198891@163.com; lizhenhui@scau.edu.cn; nqinghua@scau.edu.cn;
xqzhang@scau.edu.cn; zhezhang@scau.edu.cn; zhangzhang@big.ac.cn; george.liu@usda.gov;
hcheng@msu.edu; hans.cheng@usda.gov; nyang@cau.edu.cn; huxx@cau.edu.cn;
hzhou@ucdavis.edu; lingzhao.fang@qgg.au.dk

Abstract:

Chicken is a valuable model for understanding fundamental biology, vertebrate evolution and diseases, as well as a major source of nutrient-dense and lean-protein-enriched food globally. Although it is the first non-mammalian amniote genome to be sequenced, the chicken genome still lacks a systematic characterization of functional impacts of genetic variants. Here, through integrating 7,015 RNA-Seq and 2,869 whole-genome sequence data, the Chicken Genotype-Tissue Expression (ChickenGTEx) project presents the pilot reference of regulatory variants in 28 chicken tissue transcriptomes, including millions of regulatory effects on primary expression (including protein-coding genes, lncRNA and exon) and post-transcriptional modifications (alternative splicing and 3' untranslated region alternative polyadenylation). We explored the tissue-sharing and context-specificity of these regulatory variants, their underlying molecular mechanisms of action, and their utility in interpreting adaptation and genome-wide associations of 108 chicken complex traits. Finally, we illustrated shared and lineage-specific features of gene regulation between chickens and mammals, and demonstrated how the ChickenGTEx resource can further assist with translating genetic findings across species.

One-Sentence Summary:

The ChickenGTEx provides a multi-tissue reference of regulatory variants for chicken genetics and genomics, functional genomics, precision breeding, veterinary medicine, vertebrate evolution and even human biomedicine.

Main Text:

Introduction

The chicken (*Gallus gallus domesticus*) is not only a globally significant source of protein-rich food through both meat and egg production, but also a fundamental model species. In 2021, the farming industry achieved a staggering production of 111 million tons of eggs and 137 million tons of poultry meat worldwide (https://www.fao.org). Due to its distinct phylogenetic placement as well as its genetic and physiological characteristics, the chicken is also served as a well-recognized model organism in both fundamental and applied research (*1*, *2*), studies of domestication, genome editing, system biology, virology, immunology, oncology, and evolution (*3*–*7*). The chicken retains a remarkable range of phenotypic variation in terms of morphology, physiology, and behavior, primarily driven by artificial selection and breed specialization. Such extensive variation for a wide range of features is ideal for investigating the genetic architecture underlying complex traits. One example of such traits is dwarfism, which is characterized by a short stature and is observed in various forms in chickens, including sex-linked dwarfism, autosomal dwarfism, and the bantam phenotype, according to their physiological and genetic properties (*8*, *9*). In addition, long-term bidirectional selection lines have been established in chickens to study how polygenetic selection influences complex traits such as body weight (*10*) and feather pecking (*11*).

The Red Jungle Fowl (*G. gallus*), the ancestor of domestic chicken, was one of the first food-producing animals that had its genome assembled (*1*). Recently, a near complete version of the chicken reference genome assembly was reported, revealling distinct sequence and epigenetic features of microchromosomes (*12*). Several population-scale studies of chicken genome variation have focused on various aspects of its evolution, including speciation and domestication (*13*–*15*), signatures of selection(*14*, *16*, *17*), admixture and introgression(*14*, *18*, *19*), feralization(*20*, *21*), and phenotypic adaptation(*22*–*24*). Meanwhile, linkage mapping and genome-wide association studies (GWAS) have identified tens of thousands of genomic loci associated with numerous complex traits in chickens (*5*, *25*–*27*). As most genetic variants behind such adaptive evolutionary and complex traits are non-coding, a systematic annotation of regulatory variants in the chicken genome becomes indispensable for understanding their underlying genetic regulatory circuitry (*28*–*30*). The expression quantitative trait locus (eQTL) analysis is presently the most powerful approach to measure regulatory effects of sequence variants on individual gene expression in their native genomic and cellular contexts (*31*), as documented in the human Genotype-Tissue Expression (GTEx) project series of studies (*32*–*34*) and eQTL Catalogue in humans (*35*). In contrast, previous eQTL studies in chickens have been limited in sample size, the number of studied genomic features, and tissue/cell types (*36*–*40*). For instance, in an intercross population of 125 chickens, Johnsson et al. (2015) identified 6,318 *cis*-eQTL that influence female femoral gene expression, as measured by microarrays (*36*).

To fully unlock the genetic code of the chicken genome, the Chicken Genotype-Tissue Expression (ChickenGTEx) project, as part of the international Farm animal GTEx (FarmGTEx) initiative (*41*), has been lunched to build a comprehensive reference panel of regulatory variants based on the chicken transcriptome in various biological contexts (e.g., development, sex and environmental exposure). In this pilot study, through analyzing 7,015 bulk RNA-Seq datasets from 52 tissues/cell types (hereafter referred to as “tissues”) and 2,869 whole genome sequences (WGS) from over 100 breeds/lines (hereafter referred to as “breeds”) worldwide, we

systematically associated approximately1.5 million genomic variants with five transcriptomic phenotypes in 28 chicken tissues with sufficient sample size (ranging from 44 to 741). We then explored tissue-sharing and context-dependent patterns of these regulatory variants, their underlying molecular mechanisms of action, and their utility in deciphering GWAS loci of 108 complex traits *via* multiple complementary integrative methods such as transcriptome-wide association studies (TWAS), colocalization, and Mendelian Randomization (MR). Additionally, we compared gene regulation and the phenotypic implications between chicken and three mammalian species (*i.e.* human, cattle and pig). Altogether, our study provides novel and profound insights into the regulatory hierarchy of genetic variation in chicken transcriptomes and complex phenotypes, providing the first large-scale mapping of regulatory variants in the chicken genome and their links to complex phenotypes. Meanwhile, the atlas of regulatory variants identified in this study will facilitate the genetic improvement of chicken populations worldwide in health, production, and resilience and inform a wide range of genetic and genomic research in animal and plant species. Furthermore, we have also well-developed a ChickenGTEx online resource that is freely accessible at http://chicken.farmgtex.org.

Results

Harmonizing large transcriptome and genome datasets in chickens

We analyzed 8,668 bulk RNA-Seq samples using a uniform pipeline, yielding 304.4 billion clean reads. After filtering out low-quality data, 7,015 samples remained for subsequent analyses, representing 28 tissues (**fig. S1, Tables S1** and **2**). Based on gene expressions, samples were clustered well regarding their tissue types (**fig. 1a, fig. S2**). Across all the tissues, an average of 23,056 (94.7% of all annotated genes) genes were expressed (Transcripts per Million, TPM > 0.1) (**Table S3**), showing patterns of ubiquitous or tissue-specific expression (**fig. S3**). An average of 1,938 tissue-specific genes were then detected across tissues (**fig. S3d**), and their functions recapitulated the known tissue biology (**fig. 1b**, **fig. S3e**, **Table S4, URL**). For instance, a total of 1,425 genes were specifically and highly expressed in the bursa of Fabricius, a bird-specific primary lymphoid organ, which were significantly enriched in the immune response to bacteria (**Table S4**). An average of 54.7% of tissue-specific genes could be linked to at least one tissue-specific promoter/enhancer (**fig. S4a-d**). For instance, *MSLNL* with bursa-specific promoters and enhancers showed a specific expression in the bursa (**fig. S4d**). An average of 114 genes exhibited sex-biased expression across 18 tissues (FDR < 0.01), among which 17 genes were shared in all these tissues and located in sex chromosomes (**figs. S4e-f, Table S5**). This was in agreement with the notion of incomplete sex-chromosome dosage compensation in chickens(*42*). In addition, out of 45 genes associated with Mendelian traits in chickens(*43*), 41 showed tissue-specific expression (**fig. S5a**). For instance, *SLCO1B3* is the causal gene of blue eggshell in chickens(*44*), which was specifically and highly expressed in the liver and had liver-specific promoters and enhancers (**fig. S5b**).

To further annotate the function of chicken genes with this extensive transcriptome data, we conducted gene co-expression network and transcript assembly analyses. Based on co-expression analysis, we identified 3,583 co-expression modules containing 25,023 genes, 41.3% (10,332) of which were not functionally annotated in the current Gene Ontology (GO) database (**figs. S6**). In the set of 2,940 unannotated protein-coding genes, 56.3% (1,654) were able to be assigned to co-expression modules. Compared to annotated genes, these unannotated genes exhibited more tissue-specificity, lower gene expression level, and small proportion of chicken-human

orthologous genes (**fig. S6d**). For instance, 8 unannotated genes were co-expressed with 12 annotated genes in the muscle, which were significantly enriched in myeloid cell development and erythrocyte differentiation networks (**fig. S6f)**. Through the transcript assembly analysis, we predicted 247,383 transcripts at 48,800 loci, including 184,374 protein-coding transcripts derived from 17,215 loci, 13,140 lncRNA transcripts from 3,436 loci, and 49,869 other noncoding RNA transcripts from 34,350 loci (**fig. S7, URL**). Of all these predicted transcripts, 90% were not annotated in the previous reference and 4-10% were even transcribed from novel genomic loci (**fig. S7d** and **g**). For instance, we observed a new transcript on chromosome 2 was highly and specifically expressed in the testis (**fig. S7h**).

To obtain genotypes of these RNA-Seq samples, we called ~9 million single nucleotide polymorphisms (SNPs) from bulk RNA-Seq data using the GATK best practice pipeline(*45*) (**fig. S8**). To impute missing genotypes, we built a chicken multi-breed genotype imputation reference panel consisting of 2,869 global WGS data sets, which had a similar population composition as the RNA-Seq data (**fig. 1c, Table S6**). Adopting a missing rate of 0.6, the imputation accuracy of 1.5 million SNPs reached 97% (**figs. S8b-h**). The independent datasets from three different chicken breeds confirmed a high concordance rate (> 90%) between imputed genotypes and those directly called from WGS data of the same individuals (**fig. S8g**). After removing duplicated samples based on their genetic relatedness, 28 tissues (each consisting of over 40 individuals) were retained for subsequent molecular quantitative trait loci (molQTL) mapping (**fig. 1d**).

Discovery of molQTL

To comprehensively explore the genetic regulation of the chicken transcriptome, we conducted *cis*-molQTL mapping for five molecular phenotypes, including protein-coding gene expression (eQTL), lncRNA expression (lncQTL), exon expression (exQTL), splicing variation (sQTL), and 3'UTR alternative polyadenylation (3a'QTL), across 28 chicken tissues (**fig. 2a, fig. S2, figs. S9-10)**. In total, 13,983 (92.9%) of 15,046 tested protein-coding genes, 11,685 (74.3%) of 15,720 lncRNAs, 124,423 (76.0%) of 163,812 exons, 9,669 (61.5%) of 15,405 loci with alternative splicing events, and 8,798 (74.1%) of 11,880 loci with 3'UTR alternative polyadenylation (3'UTR APA) were significantly (FDR < 0.05) regulated by at least one genetic variant in at least one tissue, and are thus referred to as eGenes, lncGenes, exGenes, sGenes and 3a'Genes, respectively. All the molQTL tended to be enriched around transcription start sites (TSS) and transcription end sites (TES), while 3a'QTL and sQTL showed a higher enrichment in TES and gene body, respectively, compared to other molQTL (**fig. 2b**, **figs. S11a-e**). Furthermore, an average of 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 (eVariants) across tissues (**fig. 2c, fig. S12**). The further fine-mapping analysis for molQTL with SuSiE (*46*) revealed 2,887, 2,366, 2,053, 12,409 and 1,572 potential causal variants for eGenes, sGenes, lncGenes, exGenes and 3a'Genes, respectively (**URL**).

The statistical power of molQTL mapping depends on the sample size of the tissue, similar to findings in other species(*32*, *47*, *48*) (**fig. 2d-g, figs. S13** and **14**). The down-sampling analysis in the liver and muscle further confirmed the relationship between sample size and eQTL discovery power (**fig. 2g**). Most eQTL with large effect (i.e., fold change of expression, aFC > 2) were detectable when sample size reached around 200 (**fig. 2f** and **g**), and eQTL with larger effect were more enriched around TSS (**fig. S14l**) and had lower minor allele frequency (MAF) (**fig.**

S14). In general, the estimated effect size of eQTL was not correlated with their gene expression levels across tissues (**figs. S15 a** and **b**). Of note, chromosome size was significantly and positively correlated with eGene heritability (**fig. 2e**), eGene discovery (**fig. S15c**), and MAF of lead eVariants (**fig. S15d**). This was only observed in chickens and not in mammals (i.e., pig, cattle and human) (**figs. S15e-g**). Such influences of chromosome size on eQTL effects might be due to differences in evolutionary constraints between microchromosomes and macrochromosomes in chickens(*49*), which was further supported by the observation that phastCons scores of lead eVariants were also negatively correlated with chromosome size (**fig. 2e**).

To validate molQTL identified above, we first applied linear mixed model (LMM), by which we observed that the effect size and significance level of genetic variants estimated by the LMM were highly correlated with those estimated by the linear regression, implemented in tensorQTL (*50*) (**fig. S16**). We then randomly and evenly divided samples from 15 tissues with a sample size of over 100 into two subgroups, and then carried out eQTL mapping separately. A high replication rate, measured by π_1(*51*), was observed between subgroups across tissues, ranging from 0.61 in the hypothalamus to 0.92 in the embryo (**fig. 2h**, **fig. S17**). The effect size of eQTL also exhibited a high Spearman's correlation (an average of 0.77 across tissues) between subgroups (**fig. 2h**). Moreover, we observed that effect sizes derived from the eQTL mapping were positively and significantly correlated (an average of 0.52 across tissues) with those from the allele-specific analysis at the same loci (**fig. 2i, Table S8**). Finally, we trained a deep learning model of regulatory effects based on 310 functional epigenomic profiles in chicken *via* DeepSEA (*50*) (**fig. S15h**, **Table S7**), and observed that regulatory variants predicted by DeepSEA were more significantly enriched in eVariants than the expected (**fig. 2j**). Altogether, these results demonstrated the reliability of molQTL identified in this study.

Limited sharing of regulatory mechanisms underlying five molQTL types

Out of all 27,203 tested genes, 16,097 (59.2%) had significant QTL for at least two molecular phenotypes (**fig. S18a**). The LD of lead variants of any two molQTL types from the same genes was low, ranging from 0.04 (exQTL *vs.* 3a'QTL) to 0.29 (exQTL *vs.* lncQTL) (**fig. 3a**). The colocalization analysis further confirmed the limited sharing of regulatory control among these molecular phenotypes (**fig. 3a**), indicative of their distinct genetic regulatory mechanisms. **Fig. 3b** takes *NLRC5* as an example, four molecular phenotypes of which were controlled by distinct genomic loci, and LD between the respective lead variants was lower than 0.07 (**fig. S18b**). Among these molQTL, eQTL and exQTL exhibited a relatively high colocalization probability (average H4 = 0.72) (**fig. 3a**).

To elucidate molecular mechanisms of action behind these molQTL, we examined sequence ontology and multi-omics data in chickens, including 15 chromatin states predicted from 377 epigenetic data sets in 23 tissues(*30*), and 9,898 topologically associating domains (TADs) detected from high-throughput chromosome conformation capture (Hi-C) in three tissues (i.e., muscle, liver and testis)(*52*). As expected, conditionally independent molQTL were significantly enriched with various regulatory DNA sequences, including synonymous variants (1.67-fold in lncQTL to 3.03-fold in exQTL), 5'UTR variants (1.82-fold in eQTL to 3.64-fold in sQTL), 3'UTR variants (2.29-fold in sQTL to 3.77-fold in 3a'QTL), and non-coding (NC) transcripts (1.48-fold in 3a'QTL to 2.36-fold in exQTL). Of note, all five types of molQTL also exhibited a significant enrichment in missense variants (**fig. 3c**), indicating that a fraction of transcriptional

regulatory variants may also alter protein amino acid residues(*53*). Compared to other molQTL, sQTL exhibited a higher enrichment with splicing variants (63.97-fold in splice acceptor, 3.97-fold in splice donor, and 3.89-fold in splice region), while 3a'QTL were more enriched with stop retained (5.06-fold) and 3'UTR variants (3.77-fold) (**fig. 3c**).

All five types of molQTL showed the highest enrichment in promoter-like states (E1-E5, an average of 3.64-fold), followed by enhancer-like states (E6-E10, an average of 1.98-fold) and ATAC islands (E11, an average of 1.87-fold). In contrast, they were significantly depleted from repressed regions (E12-E14) (**fig. 3d**). Compared to active enhancer (E6), super-enhancer (i.e. a cluster of enhancers in close genomic proximity, exhibiting exceptionally high levels of H3K4ac signals(*30*)), had a lower enrichment for all five molQTL, suggesting that they may be under a stronger purifying selection due to their essential roles in gene regulation and cell identity (**fig. S19a**). Among the five types of molQTL, 3a'QTL had the highest enrichment in enhancer-like states and ATAC islands (**fig. 3d**), supporting their high tissue-specificity. A total of 20% of eQTL, 26% of sQTL, 3.4% of lncQTL, 17.9% of exQTL, and 14.5% of 3a'QTL were supported by regulator-gene pairs that were predicted based on the correlation of signal density of regulators and gene expression (**fig. S19b, Table S9)**. By examining 3D looping of chromatin(*52*), we found 20-60% of molQTL-gene pairs located with the same TAD across tissues (**figs. S19 c** and **d**), with the highest enrichment observed at ~400-600kp away from TSS of target genes after accounting for their distance (**fig. S19e**). As expected, 3a'QTL showed the highest enrichment at ~600-1000kb downstream of their target genes (**fig. S19e**). Likewise, 41-73% of eQTL-eGene pairs located in the same CTCF-loops that were identified from 22 chicken tissues (*30*) (**figs. S19f-h**). These results indicate that the long-distance eQTL exert effects possibly through disrupting TFBS in long-distance enhancers that interact with promoters *via* 3D looping of chromatin (**figs. S19c-h**). As shown in **fig. S19i,** eVariant *rs317368746* regulates expression of *TIMM17B* in the brain only, and it resides in a brain-specific enhancer and is located within the same TAD (346kb upstream) as the TSS of *TIMM17B*. Altogether, these results indicate that regulatory variants exert widespread effects on the transcriptome *via* multiple mechanisms such as changing transcript structure, function, stability, transcription/translation rate and chromatin conformation.

Tissue- and breed-sharing of molQTL

All five types of molQTL were either tissue-specific or ubiquitous, among which 3a'QTL and eQTL exhibited the highest and lowest tissue-specificity, respectively (**fig. 4a** and **b, fig. S20a, fig. 21)**. This was also supported by the meta-tissue analysis (**fig. S22)**. In total, 10.6% of eQTL, 32.1% of sQTL, 27.4% of lncQTL, 25.8% of exQTL, and 29.6% of 3a'QTL were active in one tissue only. Of note, eQTL that were active in more tissues showed a higher enrichment around TSS (**fig. 4c**, **fig. S20b**), a smaller effect size (**fig. 4d**) and a higher MAF (**fig. S20c**). Tissue-shared eQTL (i.e., active in at least two tissues, LFSR < 0.05) also tended to be more enriched for promoter-like states, whereas tissue-specific eQTL were more enriched for enhancer-like states (**fig. S20d**). In general, tissues with similar biological functions (e.g., immune tissues) tended to be clustered together based on eQTL effect correlation (**fig. 4a, fig. S21**), which was similar for the remaining four types of molQTL (**fig. 4b, fig. S21)**. Unlike GTEx in mammals(*32*, *47*, *48*), blood formed the primary outgroup in chickens regarding eQTL and lncQTL, while, for the remaining three types of molQTL, brain and testis were first separated from the rest of the tissues. **Fig. S20e** demonstrates an eQTL (9_16035177_G_A) that significantly regulated the expression of *ALG3* only in the blood. The *ALG3* gene encodes alpha-1,3-mannosyltransferase

with the function of inducing glycosylation of TGF-β receptor II(*54*), which might modulate blood pressure homeostasis(*55*) and affect hematopoiesis(*56*). In contrast to eQTL shared in other tissues, blood-specific eQTL had a lower MAF (**fig. S20f**) and a larger effect (**fig. S20g**). Moreover, genetic regulation of all five molecular phenotypes in the embryo was distinct from those in the primary tissues (**fig. 4a, fig. S21**), similar to that in pigs(*47*), indicating a distinct regulation of early development. In addition, we detected 59 eQTL with opposite directional effects on the same genes (n = 51) between tissues (**Table S10**). For instance, the T-allele of *rs315639985* increased the expression of *FBXO5* in the spleen but decreased its expression in the whole blood. The *FBXO5* gene encodes F-box protein 5, which is associated with systolic blood pressure in human(*57*) (**fig. S20h**). Another example was *rs313608694*, whose G-allele significantly upregulated the expression of *ELAC2* in the embryo but downregulated it in the spleen (**fig. S20h**). This gene encodes elaC ribonuclease Z 2, and the reduction of its expression could induce growth arrest by suppressing transforming growth factor-beta(*58*).

We examined breed-sharing of eQTL in the brain, liver, muscle and spleen, as all of them had more than two breeds and each with a sample size > 40. As a result, the majority of eQTL (an average of 81%) could be replicated between breeds and the replication rate was associated with tissue sample size (**fig. S23a, Table S11**). Furthermore, the eQTL effect was substantially shared between breeds (**fig. 4e**). For instance, the T-allele of *rs314795649* significantly upregulated expression of *PRKCDBP* in the liver across all four breeds being tested, including Cobb ($\beta = 0.57$, $P = 2.67 \times 10^{-6}$), Leghorn ($\beta = 0.33$, $P = 3.10 \times 10^{-6}$), Rhode Island Red ($\beta = 0.39$, $P = 3.06 \times 10^{-6}$) and Ross ($\beta = 0.37$, $P = 5.02 \times 10^{-10}$) (**fig. S23b**). In addition, we detected 376 (Red Jungle Fowl vs. Ross) and 185 (Red Jungle Fowl vs. Leghorn) breed-interaction eQTL (bi-eQTL) in the brain, and with genes regulated by them were enriched in functionals related to brain development (**Table S12**).

Context-dependence of molQTL

To explore the context-dependent nature of gene regulation, we systematically detected eQTL interacting with sex (sb-eQTL), transcription factor (TF-eQTL) and cell type (ci-eQTL). For sb-eQTL mapping, we only considered eight tissues, where each sex had data from over 30 individuals available. In total, 1,138 SNPs displayed sex-biased regulation of 962 eGenes (sb-eGene, FDR < 0.01), ranging from 3 in the small intestine to 954 in the liver (**URL**). Taking the liver as an example, we further performed the sb-eQTL mapping in a single breed, Rhode Island Red (n_{male}= 32; n_{female}= 46), resulting in 48 sb-eQTL regulating 30 eGenes (**fig. 4f, figs. S23c-d, Table S13**). For instance, the significant association of *rs317663121* with *TCFL5* expression was only observed in male liver (**fig. 4f**). Moreover, 14% (164) of sb-eGenes overlapped with sex-biased expressed genes in all eight studied tissues. These sb-eGenes detected in the blood, hypothalamus, and liver were significantly enriched in biological processes related to amino acid metabolism, signaling transduction pathway, and fatty acid metabolism (**Table S14**). Through the examination of 956 chicken transcription factors retrieved from the AnimalTFDB 3.0(*59*), we detected an average of 1,941 TF-eQTL in 17 tissues, representing 503 TFs (**fig. S23f, URL**). **Fig. S23e** illustrates that effect of *rs313600592* on *ATP6V1A* expression was significantly associated with the expression of transcription factor *TCF25* in the muscle. For ci-eQTL mapping, we first annotated 13 cell types from single-cell RNA-Seq data in chicken heart and muscle (**Table S15**). Based on the cellular composition of bulk RNA-Seq samples of muscle and heart estimated by the *in silico* cell-type deconvolution (**fig. S24**), we identified an average of 105 ci-eGenes in the muscle, ranging from 11 with interactions in adipocytes to 214 with Schwann cells, and an

average of 19 ci-eGenes in the heart, ranging from 6 interacting with fibroblasts to 36 with cardiomyocytes (**fig. S23g, Table S16**). For instance, *rs733070738* regulated expression of *PLVAP* by interacting with myocyte enrichment in the muscle (**fig. 4g**). These results highlight the dynamics of genetic regulatory effects across distinct biological contexts.

Interpreting genetic regulation behind complex traits and adaptive evolution

To show the potential of molQTL in understanding complex traits in chickens, we systematically integrated molQTL with GWAS results of 108 complex traits, including growth and development (n = 43), carcass (n = 41), egg production (n = 20), feed efficiency (n = 3), and blood biochemical index (n = 1) (**Table S17**). Enrichment analysis revealed that GWAS loci of all the traits were significantly enriched in all five types of molQTL (**fig. S25a**). Among them, the highest enrichment was observed for 3a'QTL (1.87±0.33), followed by sQTL (1.83±0.28), eQTL (1.81±0.30), lncQTL (1.59±0.27) and finally exQTL (1.56±0.32) (**fig. S25a**). Furthermore, we applied four complementary methods to prioritize causal variants and genes underlying each GWAS loci, including fastENLOC-based colocalization, summary-data-based MR (SMR), single-tissue transcriptome-wide association study (sTWAS), and multi-tissue TWAS (mTWAS). Out of all 1,176 significant GWAS loci, 1,059 (90%) could be explained by at least one molQTL across 28 tissues (**fig. 5a, figs. S26** and **S27**). Of 896 colocalized GWAS loci, 59.9% were not colocalized with the nearest genes of lead GWAS variants, indicative of the regulatory complexity of complex traits (**fig. 5b, fig. S28a**). The number of colocalization events of a trait was determined by the statistical power of both GWAS and molQTL mapping (**fig. S26b-c**). Of all 1,176 GWAS loci, 0.8%, 0.9%, 5.2% and 1.4% were explained uniquely by eQTL, sQTL, exQTL and lncQTL, respectively. This result indicates that each type of molecular phenotype only had a limited contribution to complex traits at distinct levels of gene regulation (**fig 5a**, **fig. S29**). Taking the body weight gain from week 6 to 8 (WG6.8) as an example, sTWAS linked GWAS loci to 43 unique genes (34 protein-coding and 9 lncRNA genes) across 21 tissues (**fig. 5c, fig. S28c, Table S18**). Of them, the expression of the *KPNA3* (karyopherin subunit alpha 3) exhibited the strongest association with WG6.8 in the retina, followed by pituitary and heart (**fig. 5c**). Consistently, it has been documented that the knockdown of the *KPNA3* would restore photoreceptor formation in *Drosophila*(*60*). The highest colocalization between WG6.8 GWAS loci and molQTL of *KPNA3* was observed for a retina eQTL (*rs314814283*, GRCP=0.78) and a pituitary sQTL (*rs13552958*, GRCP=0.54) (**Table S19**). The further SMR analysis pinpointed 10 potential causal mutations across tissues (**Table S20**), among which *rs739579746* was the most significant one (**fig. 5c, Table S20**). The SNPs *rs314814283* and *rs739579746* detected by eQTL mapping were in high LD (r^2 = 0.88), while both showed low LD with *rs13552958* (r^2 < 0.02) detected by sQTL mapping. These findings likely reflect the importance of photoreception for chicken growth and production performance (*61*, *62*), and the promising candidate gene in this region is the *KPNA3*. In addition, we detected 149 significant lncRNA-protein-coding-trait regulation events with SMR-multi analysis (**Table S21**). For instance, an eQTL of a lncRNA (*ENSGALG00000053557*), located on the opposite strand of the *IL20RA*, exhibited significant colocalizations with an eQTL of *IL20RA* in the muscle and GWAS loci of the total stomach weight on chromosome 3 (**fig. 5d**).

To further explore context-specific genetic regulation of complex traits, we conducted colocalization analysis between GWAS loci and three types of context-interaction eQTL detected above (**fig. S25b**). Out of 1,155 GWAS loci, 22.9% (264), 48.7% (562) and 12.3% (142) were explained by sb-eQTL, TF-eQTL and ci-eQTL, respectively (**fig. S25b**). For

instance, GWAS loci of total stomach weight and body weight at 8 weeks of age were significantly colocalized with sb-eQTL of *MFSD4A* and *TOX3* in the brain and spleen, respectively (**figs. S28d-e**). The *TOX3* gene encodes TOX high mobility group box family member 3, playing roles in sex determination and differentiation (*63*, *64*). Despite the limited discovery power of the context-interaction eQTL due to the small sample size, our analysis demonstrated that context-specific regulatory effects were nonnegligible in dissecting the regulatory mechanism of complex traits. Furthermore, we conducted an exploratory analysis to investigate whether domestication and breeding also tend to target on regulatory variants, though examining selection sweeps previously detected between broilers and layers previously (**fig. S25d, fig. S30**) (*14*, *65*). Within the brain, we separately detected eQTL in three chicken lines/breeds separately, including Red Jungle Fowl (n = 46), Ross (n=157) and Leghorn (n = 78). Genomic windows containing at least one eQTL (i.e., eQTL windows) in Ross and Leghorn were under stronger selection (i.e., larger selection values, LSBL) in broilers than expected, whereas those detected in Red Jungle Fowl were not (**fig. S25d, fig. S30**). Likewise, for selection sweeps in layers, eQTL windows in Leghorn were under stronger selection in layers than expected, but not for eQTL windows in Ross and Red Jungle Fowl (**fig. S25d, fig. S30**). Altogether, the current ChickenGTEx can serve as a valuable resource for exploiting regulatory mechanisms underlying complex traits and adaptation in chickens.

Comparing gene regulation and complex trait genetics between chickens and mammals

Based on gene orthology between chickens and three mammals (i.e., cattle, pigs and humans), we found the expression levels of the 1-1-1-1 orthologous genes were significantly higher than those of non-orthologous genes across tissues (**fig. S31)**. The proportion of orthologous genes expressed in chicken tissues was positively (Pearson's $r > 0.8$, $P < 0.004$) correlated with that in mammalian tissues (**fig. S31c**). Based on gene expression profiles, 14,278 samples in the four species were clustered first according to their tissue types, indicating the global conservation of gene expression between chickens and mammals (**fig. 6a**). This was also supported by a high correlation of TAU values of genes, a measure of tissue-specificity of gene expression, between chickens and mammals (**fig. S31d**). The phylogenetic analysis of gene expression revealed different evolutionary rates of tissues across species, where testis and pituitary evolved fastest, while adipose and liver evolved slowest (**figs. S31e**). The effect sizes of lead eQTL of orthologous genes were significantly but weakly correlated between chickens and mammals, which were lower than those within mammals (**fig. 6b**). This was consistent for *cis*-h^2 of orthologous genes (**fig. S32a**). As in pigs and cattle, the distance of lead eQTL to TSS was larger in chickens than that in humans (**fig. S32b** and **c**), which might be partially due to the larger LD of SNPs in farm animals' genomes and lower SNP density in the pilot phase of FarmGTEx compared to human GTEx(*32*). We further divided chicken eGenes of each tissue into two groups: 1) chicken-specific eGenes, and 2) those shared with at least one mammalian species (conserved eGenes) (**see Methods**). In general, compared to chicken-specific eGenes, conserved eGenes showed a higher gene expression level, lower tissue-specificity, were more likely to be differentially expressed between species, have more promoters, and stronger tolerance to loss-of-function mutations (less evolutionarily constrained) (**fig 6c**).

The FarmGTEx-based TWAS results provide new opportunities to systematically explore between-species similarity of complex trait genetics at the functional level of orthologous genes. We thus compared all the 3,024 sTWAS of 108 traits in chickens with 9,112, 1,032 and 6,480 sTWAS in three mammalian species, representing 268, 43 and 135 complex traits, respectively.

Within the matching tissues, we identified a total of 8,312 trait-pairs with significant correlations between chickens and three mammalian species ($P < 9.11 \times 10^{-3}$, permutation-based) (**fig. 6d**, **fig. S32d-f, Table S22**), despite the big differences in the TWAS power between species. Most of the significantly correlated traits between species recapitulated known biological and physiological knowledge. For instance, chicken body weight (BW) showed a high correlation with pig average daily gain (ADG) in the ileum (Pearson's $r = 0.69$, $P = 3.42 \times 10^{-5}$, **fig. 32g**), cattle somatic cell scores (SCS) in the adipose (Pearson's $r = 0.38$, $P = 5.46 \times 10^{-5}$, **fig. 32h**), and human type 2 diabetes (T2D) in the kidney (Pearson's $r = 0.57$, $P = 2.3 \times 10^{-5}$, **fig. 32i**). This was in line with previous findings that larger BW fluctuation was related to an increased T2D risk in human(*66*), and also was positively associated with SCS in cattle (*67*). The expression of *ABCC13* (encoding ATP binding cassette subfamily C member 13) in the ileum was significantly associated with both chicken BW ($P = 0.03$) and pig ADG ($P = 0.04$), which encodes ATP binding cassette subfamily C member 13, which had potential associations with body weight/body mass index in humans (*68*). The expression of three genes, *PIGX*, *MRPL51* and *ABHD14B*, in the adipose were significantly associated with both chicken BW and cattle SCS. Of these, the ABHD14B protein is a lysine deacetylase with the capacity of catalyzing the deacetylation of lysine residues to yield acetyl-CoA, which could significantly alter glucose metabolism and could thus cause significant BW loss(*69*, *70*). The expression of *GABRB2* and *SOX4* in the kidney was significantly associated with both chicken BW and human T2D. The SOX4 is involved in pancreas development with roles in inhibiting insulin secretion and increasing diabetes risk(*71*, *72*). Moreover, taking chicken BW as an example, we carried out cross-species meta-TWAS analysis in the muscle, and found that homologous traits (e.g., ADG and back fat thickness) rather than non-homologous traits (e.g., number of stillborn and weaned pigs) in pigs could help detect more genes associated with BW in chickens **(fig. 6e)**. Similarly, human height and BMI increased the detection power of BW-associated genes in chickens *via* cross-species meta-TWAS analysis in the muscle (**fig. S32k**). These results highlighted that the FarmGTEx resource could facilitate the translation of genetic findings between species at the functional level of orthologous genes rather than the DNA sequence level.

Discussions

Summary and general impacts: Through the comprehensive analyses of the so-far largest collection of chicken RNA-Seq and WGS data, we have developed a catalogue of genetic variants with regulatory effects on five transcriptional phenotypes, representing both primary expression (including protein-coding, lncRNA and exon) and post-transcriptional modifications (alternative splicing and 3'UTR APA), across 28 chicken tissues, referred as the ChickenGTEx. We made the findings and resources of ChickenGTEx freely accessible to the entire community through http://chicken.farmgtex.org. This web portal provides an open-access chicken genotype imputation reference panel, which was built-up and maintained as part of this project. The current reference panel consists of approximately 3,000 WGS samples from around the globe, enabling researchers to impute genotypes derived from RNA-Seq, SNP array or low-coverage sequences to the whole-genome sequence level, which can be utilized further to prioritize potential causal variants underlying complex traits of interest through integrating with muti-layer ChickenGTEx resources. Besides, we offer highly-useful visualization tool, Integrative Genomics Viewer (IGV) (*73*) for exploring molecular phenotypes, enhancer-gene interactions, chromatin states, epigenetic modifications, and publicly available GWAS results. The web portal also includes single-cell RNA-Seq data that were collected and analyzed from six chicken

tissues, enabling users to query the expression of their desired genes at both the cellular and bulk tissue level. Additionally, we provide batch data download and advanced search options for data resource generated in this study, and will continue updating the database to ensure its future accuracy and relevance. Overall, this first GTEx resource in avian species serves as a valuable resource for a global atlas of regulatory variants in chickens and informs vertebrate genome evolution at the functional level, benefiting future research in animal, plant, and human genetic and biomedicine research.

MolQTL mapping and the underlying molecular mechanism: We have demonstrated that different molecular phenotypes of the same genes were likely to be controlled by distinct genomic loci through distinct regulatory mechanisms, indicating the importance of integrating omics data corresponding to multi-layer biologically-important molecular phenotypes (e.g., epigenetic mark activity and microRNA expression(*74*)) in future studies. This is consistent with findings in humans that most of the sQTL and 3a'QTL were distinct from eQTL (*75*, *76*). The comparative analysis of regulatory variants reveals several specificities of gene regulation in chickens compared to mammals. For instance, the chicken genome exhibits a chromosome-size dependence in genetic control of gene expression, in contrast to mammals. Avian genomes often have chromosomes of highly variable sizes, with chicken chromosomes ranging from a minimum of 3.4 Mb to a maximum of 200 Mb (*12*). Chicken microchromosomes exhibit a higher gene density, higher GC content and DNA methylation levels(*12*, *77*), and are under stronger evolutionary constraints (*49*), that easily distinguish them from the mammal 'like' macrochromosomes. These distinct genetic and epigenetic features might lead to differences in the genetic regulation of gene expressions across chromosomes in chickens. In addition, we observed a high sharing of eQTL effect across tissues in chickens, while interestingly the blood showed the highest dissimilarity against other tissue types. This observation is in contrast to that of mammals, where the testis showed the highest dissimilarity (*32*, *47*, *48*) that is perhaps a result of nucleated red blood cells in avian blood (*78*, *79*). Moreover, we uncovered a set of genetic variants with regulatory effects interacting with biological contexts, e.g., sex, transcription factor expression, genetic background, and cell type compositions. This context-dependent molQTL explained 10-50% of GWAS loci, revealing the need to consider cell types/states under different developmental stages, nutrition, and physiology status in the future molQTL mapping experiments. By taking account of a wide range of environmental/biological contexts, we can effectively tackle the challenge of "missing regulation" (*80*, *81*). As demonstrated in human studies(*82*, *83*), harmonizing data from diverse chicken breeds/lines increased the detection power of molQTL *via* increasing sample size, facilitating the fine-mapping of causal variants *via* reducing LD of SNPs, as well as allowing breed-specific molQTL mapping (*84*). At the current pilot phase, eQTL with *trans*-regulatory effect (> 1Mb to the TSS of genes) is not considered due to the limited sample size. Discovering *trans*-eQTL, which often has a small effect size, requires hundreds of thousands of samples (*82*, *85*), and will be considered in the future when the sample size of transcriptome data is sufficient.

Potential applications of ChickenGTEx: This multi-tissue gene regulation resource opens the door to decipher the biological mechanism of complex traits, domestication and polygenic adaptation in chickens in-depth. It enables nearly 90% of GWAS loci being tested in this study to be explained by at least of one type of molQTL, a higher proportion than that in humans (78%) (*32*) or in pigs (80%) (*47*). This finding demonstrates the importance of molQTL mapping in functionally dissecting agriculturally important traits in farm animals, with a high potential for accelerating and improving the current animal breeding program and enabling the future precision selection and breeding (*26*, *86*). The focus of cross-species comparison studies in the

past decades was mainly on the DNA sequence level due to the lack of relevant functional data, and the recent Zoonomia project investigated the DNA sequence evolution of regulatory elements while based on *in silico* prediction across species (*87–89*). The ChickenGTEx offers new means to explore the evolutionary impacts of gene regulation on complex traits across species and translate genetic findings between species at the functional level of orthologous genes rather than the DNA sequence level. Our exploratory comparative analysis of large-scale TWAS between chickens and mammals illustrates how to "borrow" information between species for gene mapping (*90*, *91*). We found that cross-species meta-TWAS aided in the identification of more functional genes for homologous traits. We believe that the ChickenGTEx resource will not only contribute significantly to elucidating the molecular architecture underlying phenotypic variation in chickens, but also to developing chicken models for studying human complex traits (e.g., disease and behavior (*3–7*)).

Limitations and outlooks: The current ChickenGTEx provides the most expansive source of regulatory variants in the chicken genome. Some limitations and challenges remain in the genotype and molecular phenotype assessments. New chicken assemblies with more complete representation are becoming available with fewer computational limitations that we experienced using the GRCg6a reference genome (Ensembl version 102) (*12*, *92–94*). Future studies will consider long-read sequences to better resolve splice-variants (*95–97*), and pangenome references to annotate complex structural variants (*98*), mobile element variation (*99*), and short tandem repeats (*92*, *100*). In addition, it would be of great interest to investigate the functional impacts of rare and somatic variants on molecular phenotypes, where multi-tissue samples are collected from the same individuals with deep WGS data available. Beyond the bulk transcriptome, other molecular features could be included, e.g., DNA methylation variation, protein abundance, metabolite profiles, and the composition of the microbiome. For future single-cell genetics in chickens, a comprehensive chicken single-cell atlas will be the first step and is urgently required to explore the cell-type/state-specific gene regulation *via in silico* cell type deconvolution of large bulk tissue samples(*101*). In addition, conducting experimental follow-ups *via* methods, e.g., massively parallel CRISPR-based screens (*102*), is crucial to functionally validate and characterize regulatory effects of genetic variants and to identify functional genes of complex traits on a large-scale. In summary, the current and future versions of the ChickenGTEx project promises to establish a reference panel for studying the functional impacts of genetic variants in their native genomic and cellular contexts in distinct biological contexts, including molQTL mapping, molecular phenotype prediction for individuals with genotypes (including extinct species with ancient DNA available) and the evolution of regulatory variants. The fully developed ChickenGTEx will contribute substantially to research in complex-trait genetics, animal breeding, functional biology, and vertebrate genome evolution at the functional level.

Acknowledgments:

We thank all the researchers who have contributed to the publicly available data used in this research.

Funding:

D. Guan was supported from Agriculture and Food Research Initiative Competitive grants nos. 2020-67015-31175, and 2022-67015-36215 (H.Z.) from the USDA National Institute of Food and Agriculture. H.Z. acknowledges fundings from Agriculture and Food Research Initiative

Competitive grants nos. 2020-67015-31175, 2015-67015-22940, and 2022-67015-36215 (H.Z.) from the USDA National Institute of Food and Agriculture, Multistate Research Project NRSP8 and NC1170 (H.Z.), and the California Agricultural Experimental Station (H.Z.). 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, "Genetic architecture, gene interaction and genomic prediction for chicken growth evaluated using large advanced intercross populations" (31961133003). Y.W. was supported by the National Natural Science Foundation of China, "Deciphering the genetic architecture of polygenic clustered QTL for chicken body weight by integrative omics", (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 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, "Accelerating Genetic Improvement of Ruminants Through Enhanced Genome Assembly, Annotation, and Selection" 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. 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., 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:

The authors declare no competing interests.

Data and materials 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. Details of RNA-Seq, WGS, ChIP-Seq peaks and single-cell RNA-Seq can be found in Table S1, S6, S7 and S15, respectively. All processed data and the full summary statistics of molQTL mapping and genotype imputation reference panel are available at http://chicken.farmgtex.org. 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/FarmOmics/ChickenGTEx_pilot_phase).

Figures and legends

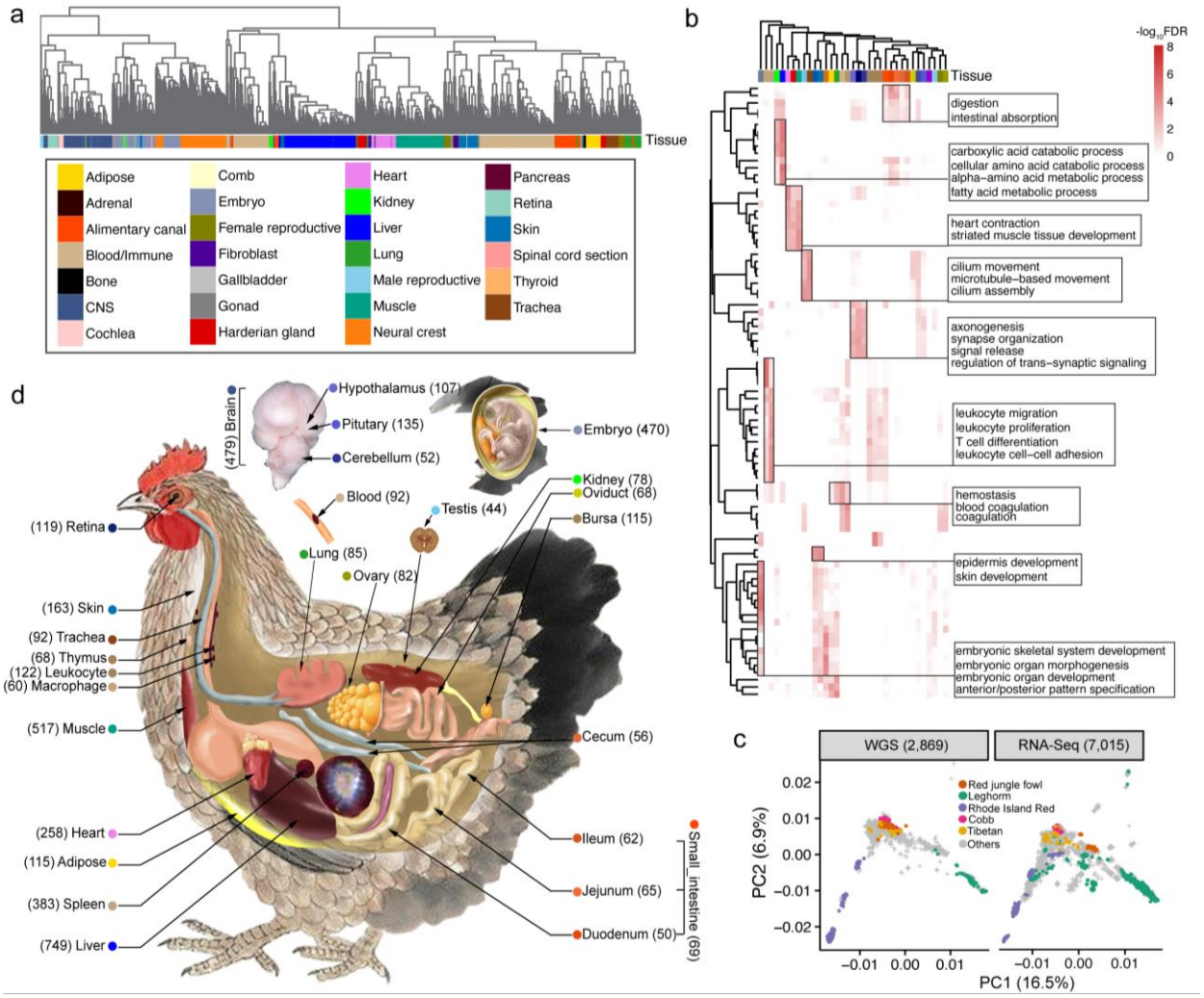


Fig. 1. Data summary in the pilot phase of ChickenGTEx. (**a**) Hierarchical clustering of 7,015 RNA-Seq samples. Distance between samples was calculated using 1-*r*, where *r* is the Pearson correlation coefficient calculated from gene expression values (quantified as Transcripts per Million, TPM) of 5,000 genes with the highest expression variance (measured by standard deviation) across samples. (**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) at the base of 10, obtained by the clusterProfiler 4.0 package with default settings(*103*). (**c**) 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. (**d**) Illustration of tissue types used in molecular quantitative trait loci (molQTL) mapping. Sample sizes (in the bracket) and colors of all the 28 tissues with sample sizes over 40 are depicted.

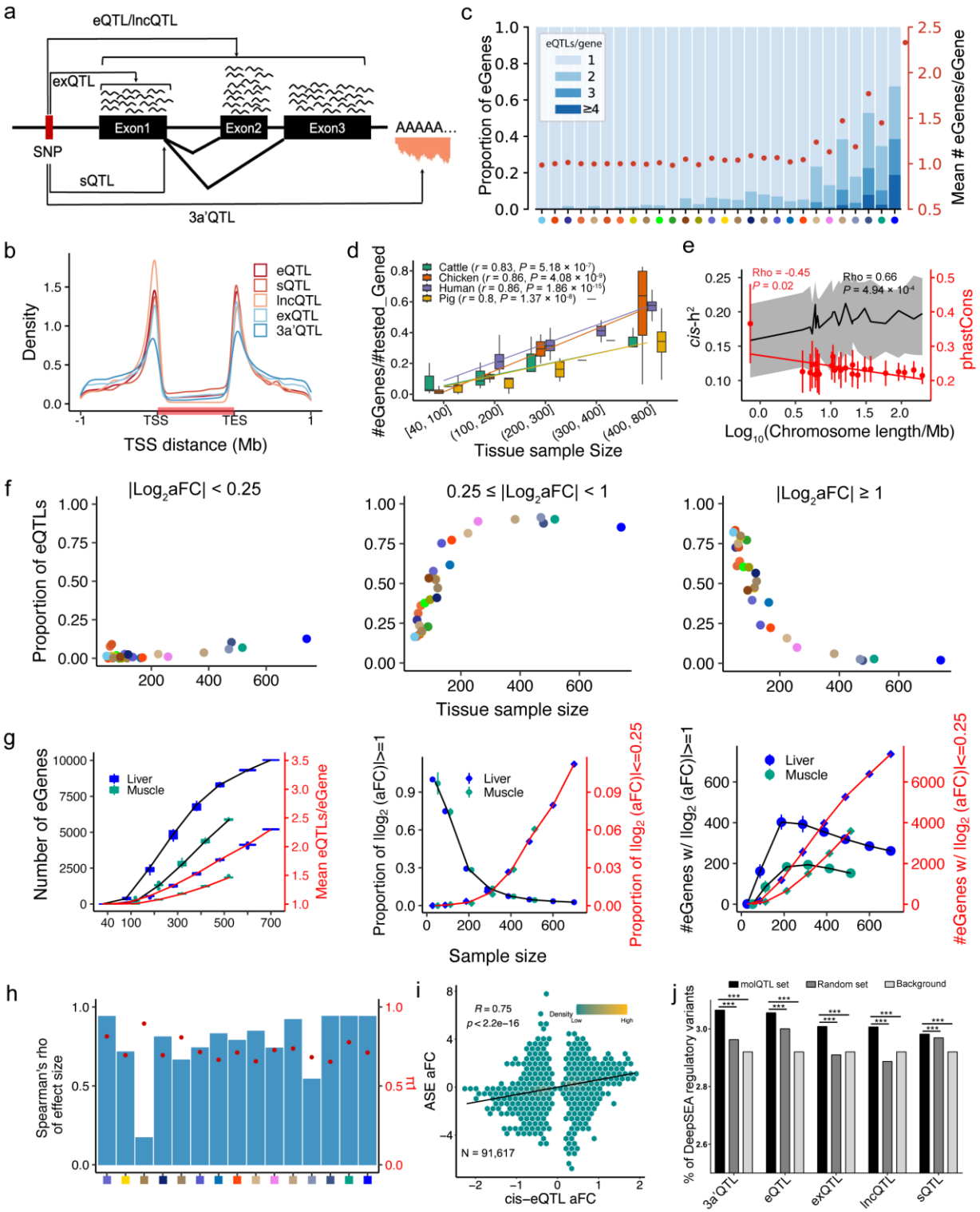
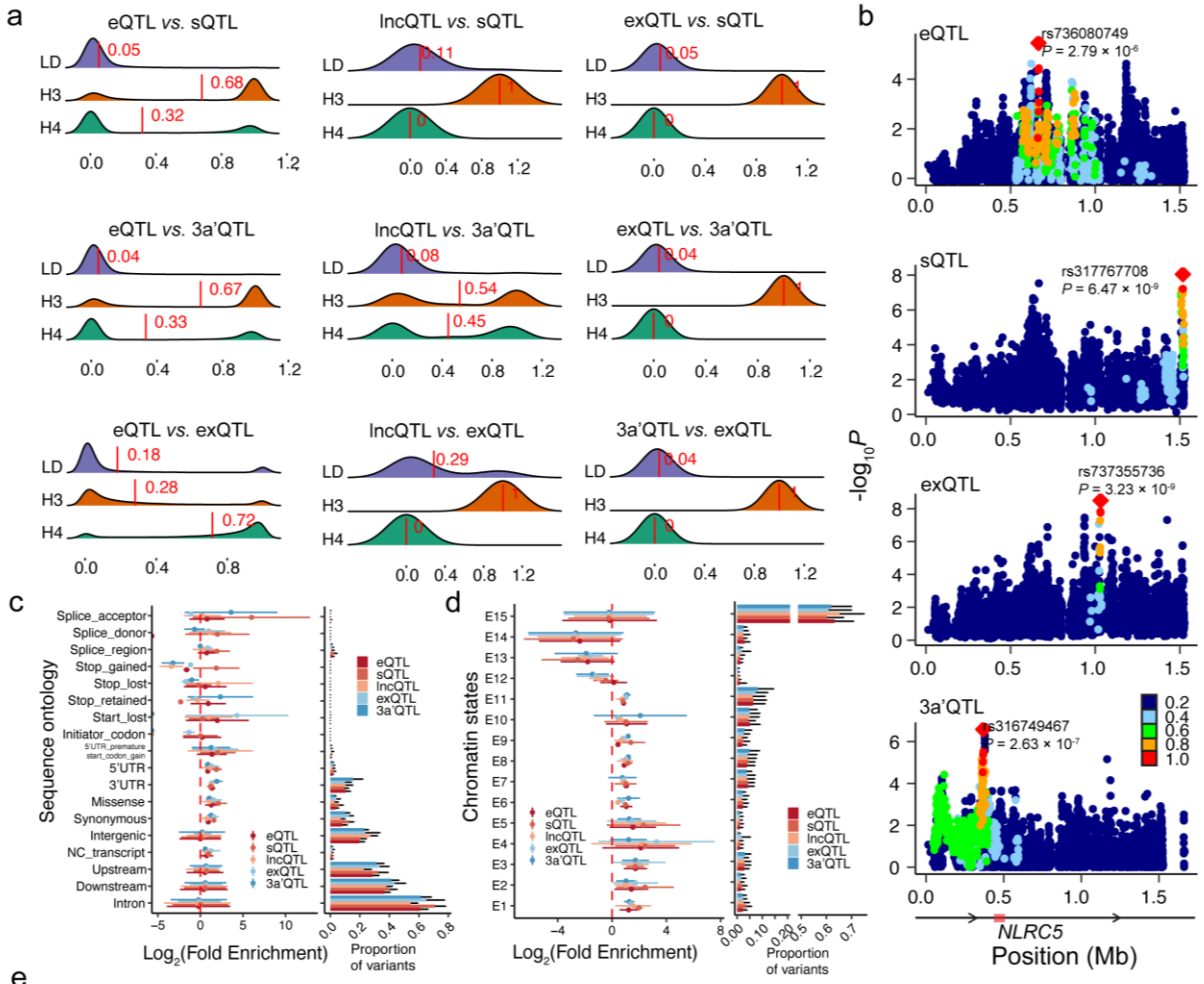


Fig. 2. Molecular QTL (molQTL) mapping in 28 chicken tissues. (**a**) Illustration of the definition of five molecular phenotypes and the respective molQTL, including protein-coding gene expression (eQTL), lncRNA expression (lncQTL), exon expression (exQTL), splicing variation (sQTL), and 3' untranslated region alternative polyadenylation (3'UTR APA, 3a'QTL). (**b**) Distribution of molQTL around gene body of eGenes, denoted by the horizontal red bar. TSS: Transcription Start Site, and TES: Transcription End Site. (**c**) 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 can be found in Fig. 1c and Table S2. (**d**) Proportion of eGenes detected as a function of tissue sample size across species, including 28, 34, 24 and 49 tissues in chickens, pigs, cattle and human, respectively. The lines are fitted with a linear model implemented in the geom_smooth function of the ggplot2 package(*104*). Correlations and *P* values were computed with the Spearman method using the cor.test function in R v3.6.5(*105*). (**e**) *cis*-h^2 (cis-heritability, left *y*-axis) and phastCons scores (right *y*-axis) of lead eQTL as a function of chromosome size (log_{10}scaled). The top and bottom boundaries of the grey shade indicate the 25% and 75% of *cis*-h^2 range, respectively, and the black line is the median of *cis*-h^2 values. Red dots are average phastCons of lead eQTL, and red bars are their standard deviations. The correlations were computed with the Spearman method, and *P* values were computed *via* the asymptotic *t* approximation. (**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). (**g**) 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 the 28 tissues. The left panel depicts the number of eGenes (left *y*-axis) and mean eQTL per eGene (right *y*-axis) detected at different sample size. The middle panel shows the proportion of detected eQTL of large (absolute $log_2aFC \geq 1$, left *y*-axis) and small effect size (absolute $log_2aFC \leq 0.25$, right *y*-axis). The right panel presents the number eGenes detected when the regulatory effect size of lead eQTL is large (absolute $log_2aFC \geq 1$, left *y*-axis) and small (absolute $log_2aFC \leq 0.25$, right *y*-axis). (**h**) 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 (*x*-axis) can be found in **Fig. 1c** and **Table S2**. (**i**) Correlation between effect size of eQTL (*x*-axis, n=91,617) and those of same loci derived from allele-specific expression (ASE, *y*-axis) analysis in liver. (**j**) The proportion of regulatory variants predicted by DeepSEA (prediction score > 0.7) based on 310 functional profiles in chickens. 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.



	ATAC	CTCF	DNaseSeq	H3K27ac	H3K27me3	H3K4me1	H3K4me3
3a'QTL	0.86	1.23	0.96	0.70	1.08	1.03	0.80
eQTL	0.98	0.88	0.92	1.35	1.18	0.66	1.31
exQTL	0.99	0.99	0.95	1.12	1.13	0.88	1.08
lncQTL	1.08	1.00	1.08	1.29	1.54	1.27	0.97
sQTL	1.06	0.92	0.98	1.12	1.41	0.59	1.17

Fig. 3. Colocalization and functional enrichment of molQTL. (**a**) Colocalization analyses of different types of molQTL of the same genes. The "LD" is the linkage disequilibrium (LD) of lead SNPs of two molecular phenotypes. "H3" and "H4" represent the probability of whether the association of two molecular phenotypes is due to two independent SNPs or one shared SNP, respectively. The vertical red lines indicate corresponding mean values. (**b**) Associations (i.e., -$\log_{10}$ transformed *P*) of genetic variants with four molecular phenotypes of *NLRC5*. The panels from top to bottom represent gene expression, alternative splicing, exon expression and 3'UTR APA, respectively. Color legend represents the degree of LD between the lead SNP and the others. The proportion and enrichment of five types of molQTL across sequence ontology (i.e., variant types annotated by SnpEff software(*106*)) (**c**) and 15 chromatin states (**d**). Fold enrichment is shown as mean (dot) ± standard deviation ($\log_2$ scaled, error bar) across 28 chicken tissues. The chromatin states were retrieved from Pan et al. (2023) (*30*). (**e**) The enrichment fold (odds ratio, OR) of molQTL in regulatory variants of seven epigenomic marks predicted by DeepSEA(*107*) (prediction score > 0.7). OR = (A/B)/(C/D), where C is the length of molQTL overlapped with annotated features (A), and B is the length of molQTL overlapped with the total genome length (D).

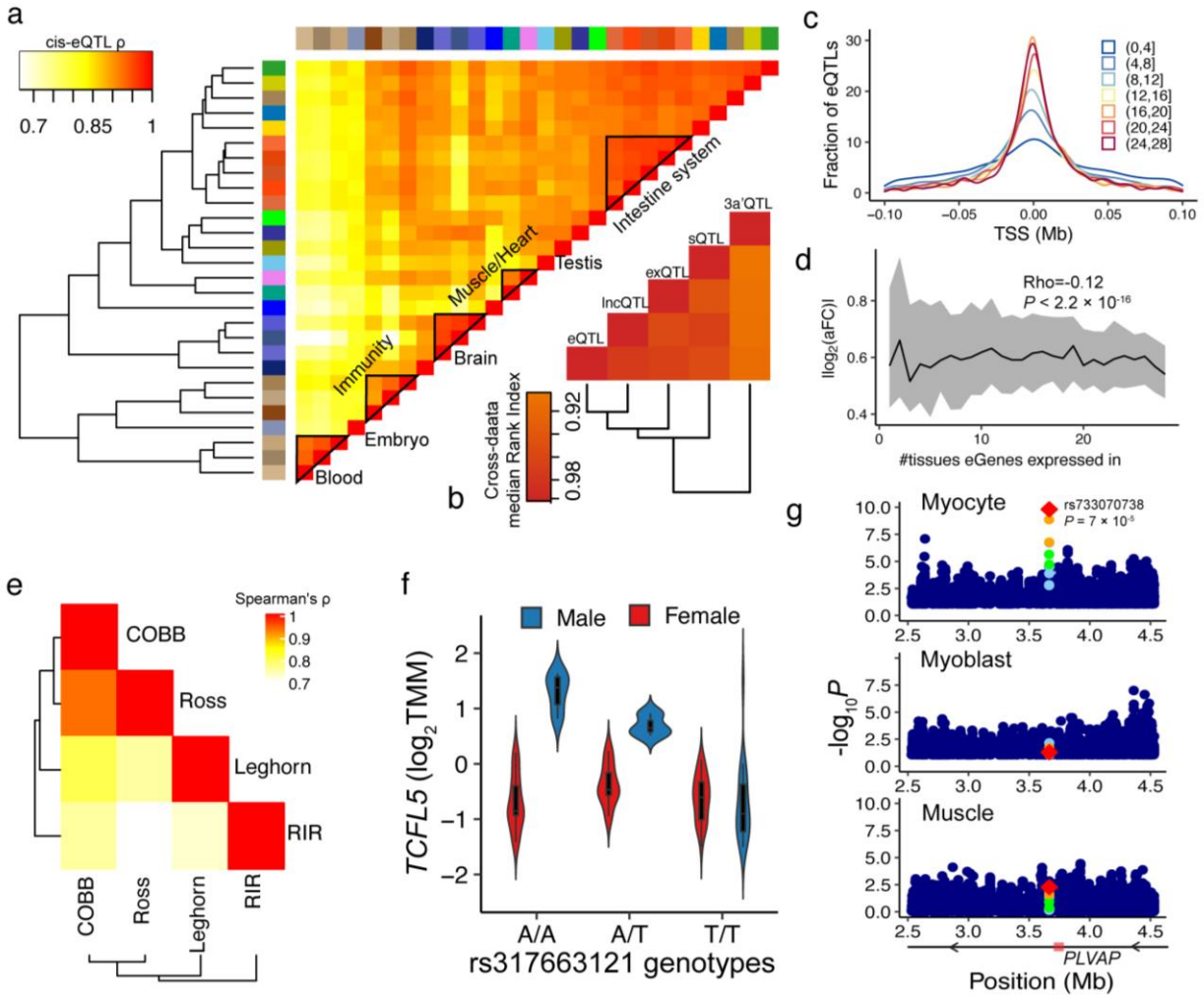


Fig. 4. Tissue-sharing and context-dependent patterns of molQTL. (**a**) The heatmap of Spearman's correlation of eQTL effect size between tissues. Tissues are clustered on the basis of dissimilarities (i.e. 1-d), where d is Euclidean distance calculated from the eQTL effect, with a complete linkage method(*108*). The color legend of tissues is the same as in **Fig. 1c** and **Table S2.** (**b**) Similarity of molQTL effect-based tissue clustering patterns. The pairwise Rand Index across five types of molQTL is 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 eQTL around transcription start site (TSS) according to number of tissues they are active in. (**d**) Absolute effect size (allelic fold change, aFC) of eQTL 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 in R v3.6.3. (**e**) Heatmap depicting of eQTL effect sharing between breeds. This analysis was done by using Multivariate Adaptive Shrinkage (*109*), same as in panel **a**. (**f**) The expression of the *TCFL5* gene across three genotypes of *rs317663121* in males (AA, n=107; AT, n=106; TT, n=92) and female (AA, n=66; AT, n=56; TT, n=73). TMM is the Trimmed Means of M values, representing the normalized gene expression level. (**g**) The significance (-log$_{10}$*P*) of interaction between *rs733070738* genotypes and myocyte enrichment (top panel) and myoblast enrichment (middle panel) on

PLVAP expression in muscle. The bottom panel is the Manhattan plot for eQTL mapping of *PLVAP* in bulk muscle samples. Dot color means linkage disequilibrium (LD) between *rs733070738* and the rest.

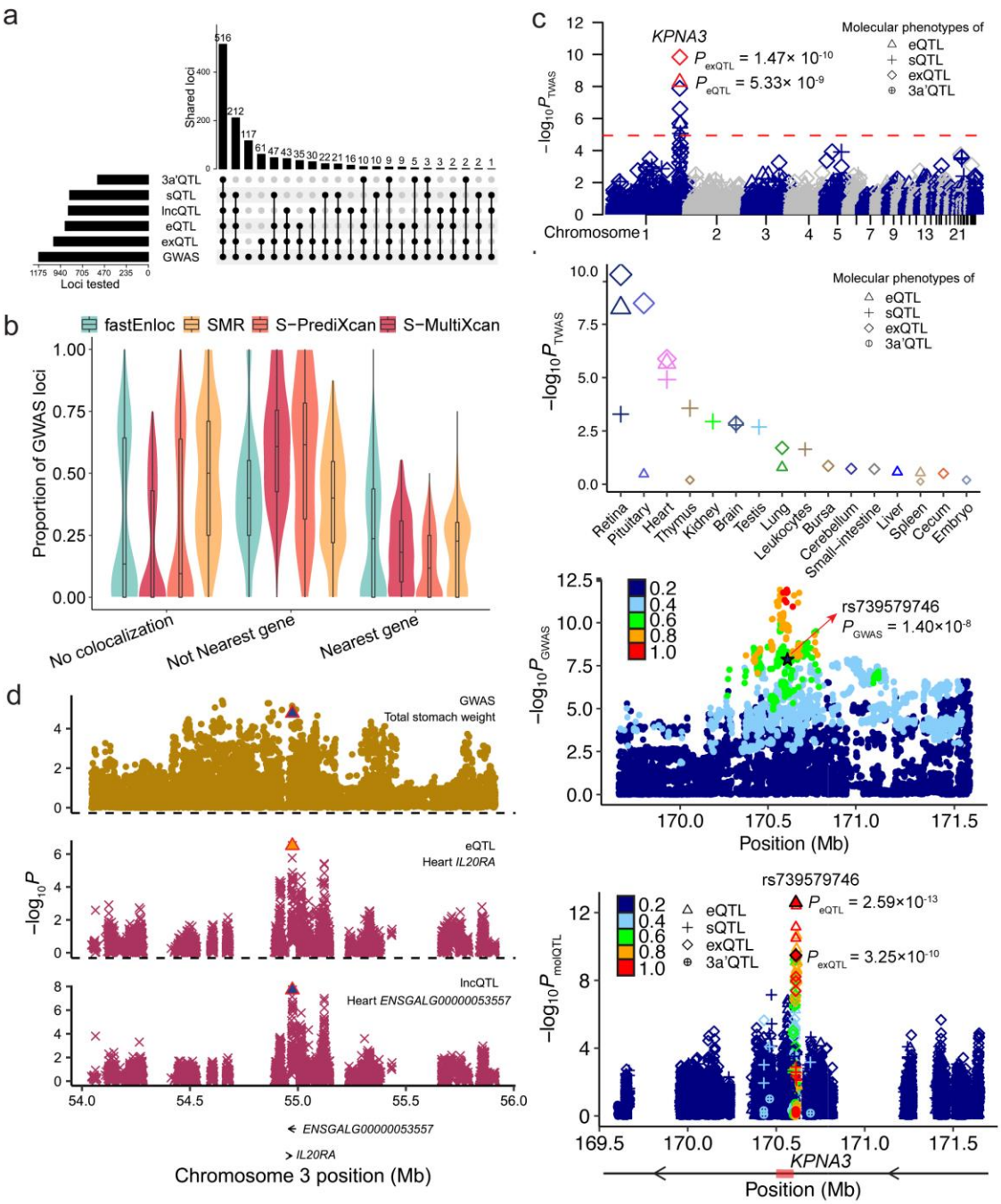


Fig. 5. Interpretation of GWAS loci with molQTL. (**a**) UpsetR plot depicting the number of GWAS loci explained by five types of molQTL, which were detected by at least one of four complementary integrative methods, including fastENLOC-based colocalization, Summary-based Mendelian randomization (SMR), single tissue-based transcriptome-wide association study (sTWAS), and multi-tissue TWAS (mTWAS). (**b**) The proportion of three types of GWAS loci (n = 1,155) colocalizing with eQTL regarding the integration results using 4 methods, including fastENLOC-based colocalization, Summary-based Mendelian randomization (SMR), single tissue-based transcriptome-wide association study (sTWAS), and multi-tissue TWAS (mTWAS). No colocalization: GWAS loci that are not interpreted by any eGenes in 28 tissues. Not nearest gene: GWAS loci are interpreted by eGenes that are not nearest genes to GWAS lead SNPs. Nearest gene: GWAS loci are interpreted by eGenes that are the nearest ones to GWAS lead SNPs. Each dot represents one of 108 complex traits. (**c**) Interpretation of GWAS loci of weight gain from week 6 to 8 (WG6.8) with molQTL. The top panel depicts associations of genes with WG6.8 *via* sTWAS in retina. The second lower panel displays associations ($-\log_{10}P$) of different molecular phenotypes (gene expression, exon expression, alternative splicing and 3'UTR APA) of *KPNA3* with WG6.8 obtained by sTWAS across tissues. The following Manhattan plot exhibits GWAS associations of SNPs with WG6.8 on chromosome 1. The color indicated linkage disequilibrium (LD) of SNPs with the lead one (*rs15497848*, $P = 1.22 \times 10^{-12}$). The colocalized SNP (*rs739579746*, $P = 1.4 \times 10^{-8}$) is denoted as a black star. The bottom plot represents molQTL mapping results of *KPNA3* in retina. The color represents LD values of the colocalized SNP (*rs739579746*, black color) with the rest. (**d**) SMR-multi results of GWAS loci of total stomach weight and eQTL and lncQTL. The top panel depicts GWAS associations of SNPs (represented by dots) with total stomach weight. The middle panel exhibits SMR associations of GWAS loci with eQTL of *IL20RA* in heart, while the bottom panel exhibits SMR associations of GWAS loci with lncQTL of *ENSGALG00000053557*. The triangle shape shows the potential causal SNP (*rs314997637*) across the three biological layers, i.e., expression of *ENSGALG00000053557*, expression of *IL20RA* and total stomach weight.

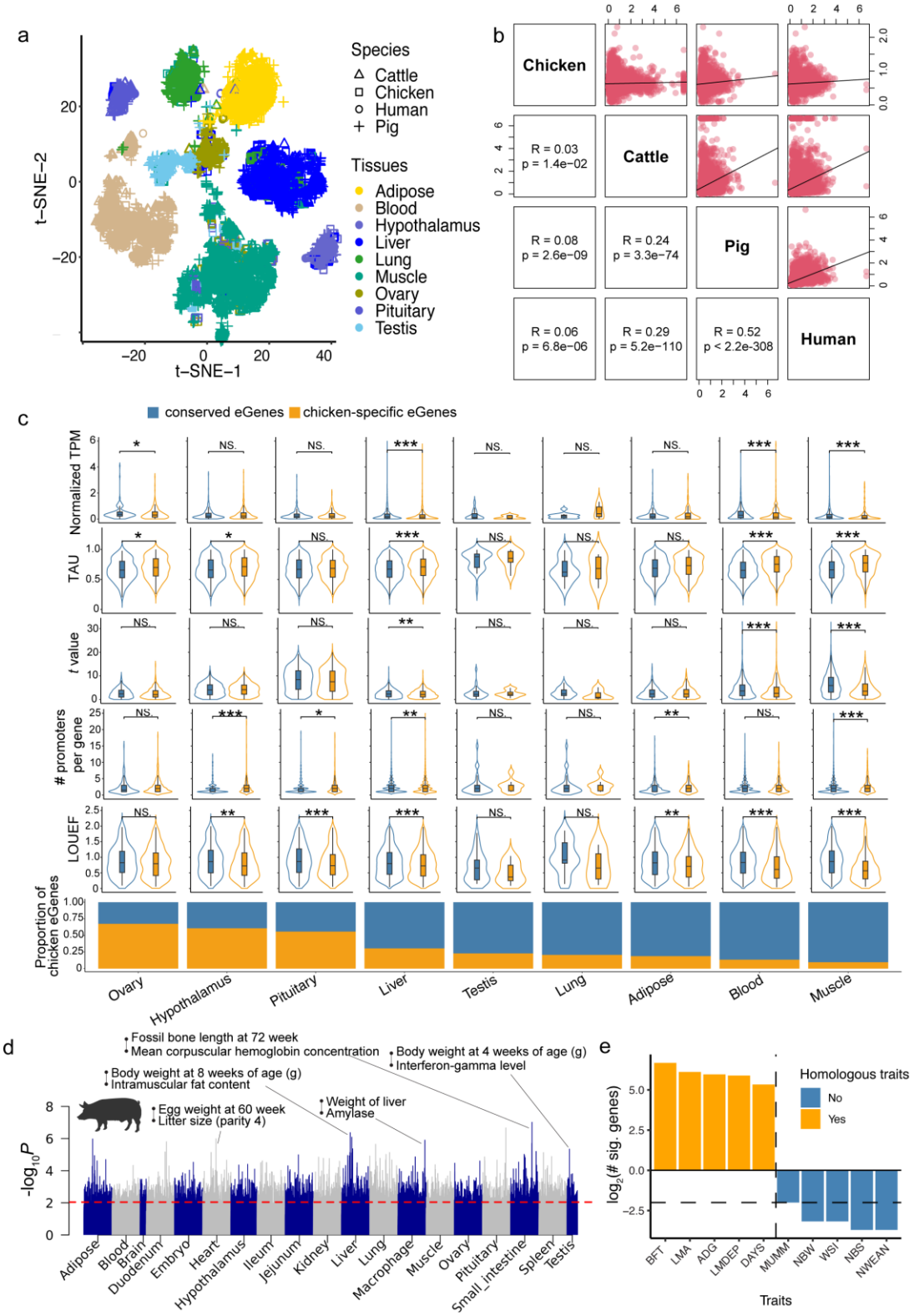


Fig. 6. Comparative analyses of gene regulation and transcriptome-wide associations (TWAS) between chickens and mammals. (**a**) Visualization of variance in gene expression of 14,278 RNA-Seq samples across four species (i.e., Chicken, pig, cattle and human) *via* a *t*-distributed stochastic neighbor embedding (*t*-SNE). Gene expression (Transcripts per Million,

TPM) of 10600 1-1-1-1 orthologous protein-coding genes are normalized between samples using Seurat software (v4.3.0) (*110*). (**b**) Pearson's correlation of averaged effect size of lead eVariants of 5,513 orthologous eGenes between species. (**c**) Comparison of chicken-specific eGenes and conserved eGenes (i.e., eGenes that are shared with at least one mammalian species) across nine tissues. The bottom barplot depicts the proportion of chicken-specific eGenes and conserved eGenes in each tissue. The violin plots from top to bottom depict expression level, TAU (tissue-specificity of expression), *t*-value (measuring the degree of gene differential expression between species), number of promoters per gene, and loss-of-function intolerance (quantified by LOEUF), respectively. Statistical tests were done by employing two-sided Wilcox-test. $^{***}P \leq 0.001$; $^{**}0.001 < P \leq 0.01$; $^{*}0.01 < P \leq 0.05$; NS: not significant ($P > 0.05$). (**d**) Significance (at $\log_{10}$ transformed, *y*-axis) for TWAS-based correlations calculated using one-to-one orthologous gene effect between chicken and pig. The red dashed line depicts the threshold of significance (Permutation-based *P* value < 0.01, corresponding to nominal *P* value < 9.11×10^{-3}). (**e**) The number of genes newly detected (FDR < 0.05) for body weight in chickens by using cross-species meta-TWAS analysis in muscle. The dashed horizontal line indicates 0 before the log transformation. Orange and blue bars represent homologous and nonhomologous traits in pigs for chicken body weight, respectively. BFT: backfat thickness, LMA: loin muscle area, ADG: average daily gain, LMDEP: loin muscle depth, DAYS: days, MUMM: number of mummified pigs, NBW: number of weak pigs, WSI: weaning to estrus interval, NBS: number of stillborn pigs, NWEAN: number of weaned piglets.

Methods and Materials

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)(*111*) 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/)(*112*), using featureCounts (v2.0.1)(*113*) and StringTie (v2.1.5)(*114*), respectively. Using the same software(*113*), we counted the total number of reads as a function of annotated exons, which were further transformed into TPM using TBtools(*115*). We performed the tree clustering of all the RNA-Seq samples using the GGTREE package(*116*). 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(*117*).

We quantified alternative splicing variation from RNA-Seq data using the LeafCutter package(*118*), which took into account spliced reads so that both novel and known alternative splicing events could be identified and quantified(*118*). 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(*118*). 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 (*112*) using the script "map_clusters_to_genes.R" (https://github.com/broadinstitute/gtex-pipeline). Afterward, we filtered out introns 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: $\sum_i(|z_i| < 0.25) \geq n\text{-}3$ 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(*118*). 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)(*119*). 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)(*120*). 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.

Single-cell RNA-Seq data analyses

We retrieved single-cell RNA-Seq data from the chicken heart (n = 7) (*121*) and muscle (n = 2) (*122*) from the public database. Raw sequencing data was processed by using the "count" function after preparing the genome annotation .gtf file (Ensembl v102) with the *mkgtf* tool of the Cell Ranger pipeline(*123*). The Seurat R package (v4.0.5)(*124*) was used for subsequent cell-type identification. We first created the Seurat object based on the raw read count of each sample in a tissue using the *CreateSeuratObject* function. In this step, we filtered out cells with unique gene counts < 200 and with mitochondrial counts > 20% of the total counts. We then normalized raw counts of gene expression using the *LogNormalize* algorithm and further identified highly variable genes (HVG) using the *FindVariableFeatures* algorithm with default parameters. The HVG count matrices of all samples for a given tissue were integrated and combined to form a single *Seurat* object using the *FindIntegrationAnchors* and *IntegrateData* functions. We scaled the integrated dataset using the *ScaleData* function, which was further used to run principal components analysis (PCA) with the *RunPCA* function. The top 15 PCs, where the percentage of variance explained tended to be constant based on the elbow plot by the *JackStraw* function, were selected for running Uniform Manifold Approximation and Projection (UMAP) analysis for cell clustering using the *RunUMAP* function. The nearest neighbors between cells were constructed using the *FindNeighbors* function and cell clusters were thus determined using the *FindClusters* function at a resolution of 0.05. We manually assigned cell names based on original

publications (*121*, *122*) and the PanglaoDB database(*125*). Finally, cell clusters were visualized using the UMAP algorithm with the *DimPlot* function. To further deconvolute bulk RNA-Seq data using single-cell RNA-Seq data, we first created a signature matrix using the CIBERSORTx tool (*126*) with default parameters. Using the "Impute Cell Fractions" from the same tool, we imputed cell fractions with the custom mode and 1000 permutations after uploading the gene expression matrix for the bulk RNA-Seq data.

Tissue-specificity of gene expression

We employed tspex(*127*), a tissue-specificity calculator, to compute 12 tissue-specificity metrics, including 8 general scoring metrics (i.e. Couts, Tau, Gini coefficient, Simpson index, Shannon entropy specificity, ROKU specificity, Specificity measure dispersion, and Jensen-Shannon specificity dispersion) and 4 individualized scoring metrics (i.e. Tissue-specificity index, Z-score, Specificity measure, and Jensen-Shannon specificity). To identify tissue-specifically expressed genes in a tissue, we applied another *t*-statistic approach as described previously(*128*). Briefly, for a given tissue, we carried out differential gene expression analysis between the target tissue and the rest but excluded those from the same biological system using the limma package(*129*). Subsequently, tissue-specific genes were identified when FDR corrected *P*-value(*130*) > 0.05 and fold change > 2. Functional enrichment analysis of tissue-specific genes with Biological Process (BP) terms in the Gene Ontology (GO) database was performed using the clusterProfiler package(*103*).

Sex-biased gene expression

To identify genes with sex-biased expression, we employed DESeq2 software(*131*) to carry out differential expression analysis between male and female samples in 18 tissues, where sample size of each sex was > 10. We fitted a generalized linear model for the differential expression analysis while correcting for factors, including BioProject, year when RNA-Seq data was generated, age, breed, sequencing platform, library layout and selection method. After multiple testing correction by the FDR approach(*130*), the set of differentially expressed genes was identified when FDR corrected *P*-value < 0.01.

Reference-guided transcript assembly

Based on STAR alignment files, we assembled transcripts with the guidance of the Ensembl annotation (GRCg6a v102) using the StringTie2 software tool(*114*). To increase the computational efficiency, transcript assembly was run by tissue. Then, the generated assembly files from all tissues were merged by using the "merge" function of the StringTie2 software(*114*). After quantifying the expression of assembled transcripts, we only retained single-exon transcripts with TPM >1 in at least half of samples in a tissue, and multi-exon transcripts with TPM >0.1 in at least half of samples in a tissue. Moreover, we compared our prediction to Ensembl and NCBI annotations using GffCompare (version 0.11), and classified them into 14 classes as described previously(*95*, *132*). The coding potential of predicted transcripts was predicted by using CPP2 software(*133*), with lncRNA loci predicted using FEElnc(*134*).

Construction of gene co-expression networks

To build gene co-expression networks in each tissue, we employed five complementary methods with default parameters, including WGCNA (v1.69)(*135*), ICA (v1.0.2)(*136*), PEER (v1.3)(*137*), MEGENA (v1.3.7)(*138*), and CEMiTool (v1.8.3)(*139*). The input gene expression values were adjusted for hidden confounding factors by regressing out 10 PEER factors and 5 genotypic PCs (see "**Preparation for molQTL mapping**" section). Functional enrichment analysis of gene co-expression modules was conducted by using clusterProfiler (v4.0)(*103*), and the following visualization was done using Gephi (v0.9.2)(*140*).

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)(*45*), 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(*45*). Then, we carried out joint-calling of all GVCF samples using the *GenotypeGVCFs* module from the GATK tool(*45*). 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 S6**). 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(*141*). 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)(*142*). The alignment files in Binary Alignment Map (BAM) format were sorted using SAMtools (v1.9)(*143*), and were further passed for the removal of PCR duplicates using GATK (v4.1.9.0)(*45*). The obtained BAM files were then used for variant discovery to generate individual GVCF files using the *HaplotypeCaller* function of the GATK tool(*45*). The joint-calling of all 2,869 GVCF samples was further done using the *GenotypeGVCFs* module from the GATK tool(*45*). 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(*144*), and then imputed missing genotypes using the Beagle 5.1 program(*145*). 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(*144*) 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(*145*).

Preparation for molQTL mapping

Sample deduplication. After assigning RNA-Seq samples into 28 tissue types, we calculated the identity-by-state (IBS) distance between samples within each tissue based on the imputed SNP genotypes using PLINK v1.9(*146*). The formula of IBS calculation is as follows:

$$IBS = \frac{(IBS2 + 0.5 \times IBS1)}{(IBS0 + IBS1 + IBS2)}$$

where IBS0, IBS1 and IBS2 are the number of non-missing variants when IBS = 0, IBS = 1, and IBS = 2, respectively. If the IBS distance of a pair of samples is higher than 0.9, they were deemed as duplicated so that the samples with the higher sequencing depth were kept. The deduplication process was run until all IBS values per pair of samples were less than 0.9. Finally, a total of 28 tissues with sample sizes ranging from 44 (testis) to 741 (liver) were kept for subsequent molQTL mapping.

Principal component analysis. Within each of those 28 tissues, we first LD-pruned imputed genotypes with the option of "--indep-pairwise 200 100 0.1" using PLINK v1.9(*146*). Principal component analysis (PCA) of samples was then carried out, based on the LD-pruned genotypes using the smartpca tool of the EIGENSOFT v8.0.0 package(*147*). The top 5 principal components (PCs) were selected as covariates for heritability estimation and molQTL mapping.

Estimating PEER confounder factors. To correct for confounders and other unwanted technical or biological variations in RNA-Seq samples, we estimated the Probabilistic Estimation of Expression Residuals (PEER) in each of the tissues using the PEER software package(*137*). The top 10 PEER factors showing highly relative contributions (i.e., factor weight variance) to gene expression variation were selected for subsequent heritability estimation and molQTL mapping.

Phenotype preparation. For protein-coding genes, lncRNAs and exons, we filtered out features with TPM < 0.1 and raw read counts < 6 in > 20% of samples within a tissue. Raw read counts were normalized using the Trimmed Mean of M-value (TMM) algorithm of the edgeR package(*148*). The generated TMM matrix was then further normalized with an inverse normal transformation for subsequent molQTL mapping. For splicing and APA, the preparation of molecular phenotypes was described in the "**RNA-Seq data analysis and molecular phenotype definition**" section.

Estimating *cis*-heritability of gene expression

We leveraged the GCTA program v1.93.2(*149*) to estimate *cis*-heritability (*cis*-h^2) of molecular phenotypes by fitting a mixed linear model:

$$y = X\beta + g + \varepsilon$$

where y is a vector of phenotypic values (i.e., gene expressions) of all samples, β is a vector of corresponding coefficients of quantitative covariates X of all samples, which included 5 genotypic PCs and 10 PEER factors, g is a vector of the genetic values of SNPs around ±1Mb of the transcription start sites (TSS) of a gene, and ε is a vector of residuals. The genetic value g

followed a normal distribution with mean of 0 and variance of $\boldsymbol{A}\sigma_g^2$, where $\boldsymbol{A}$ was the genetic relationship matrix (GRM) between individuals(*150*). Thus, we can estimate σ_g^2, i.e., the variance explained by SNP genotypes (i.e., *cis*-h^2), using the restricted maximum likelihood (REML) approach(*150*, *151*) implemented in GCTA software(*149*). The *cis*-h^2 was finally defined when the significance level was lower than 5% based on the likelihood ratio test(*149*, *150*).

Molecular QTL mapping

In this study, we only intended to map *cis*-molQTL of each feature, i.e., SNPs distributed around 1 Mb upstream and downstream of the TSS of the gene, using tensorQTL v1.0.4(*152*). This utilized graphics processing units (GPUs) with the scalability to increase runtime and reduce the time cost. Initializing with the option of "--mode cis_nominal" of the tensorQTL v1.0.4(*152*), 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(*153*) using the false discovery rate (FDR) approach(*130*), 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)(*105*), where beta_shape1 and beta_shape2 were computed by tensorQTL v1.0.4(*152*). The significant molQTL were tested SNPs whose nominal *P*-values were lower than the gene-level significance threshold.

Fine-mapping analysis of molQTL

We employed four 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 (*32*, *47*, *48*). This analysis was done by using the tensorQTL v1.0.4 with "--mode cis_independent" option(*152*). 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) (*46*). 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 transcribed from the same gene. We then ran the *coloc.abf* function in the coloc package(*154*), 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(*146*).

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)(*109*) and METASOFT v2.0.0(*155*). 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(*156*). 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(*155*), and MashR (v0.2.57)(*109*), as well as $\pi 1$ statistic in the qvalue package(*51*, *157*). The METASOFT and MashR were run as described above, and the $\pi 1$ statistic (i.e. replication rate)(*157*) was used to assess if an eQTL detected in one breed can be replicated in another breed.

Detection of context-dependent QTL

Sex-biased eQTL. To identify eQTL that is significantly associated with gender, we focused on eight tissues that had at least 30 samples for each sex. In this study, we only considered conditionally independent eQTL identified above to reduce the computational burden. We fitted a linear model $y = g + s + s \times g + c + e$, where y is phenotypic values of gene expression; g is genotype (0 for homozygous ref, 1 for heterozygous, and 2 for homozygous alt); s is sex information (0 for female and 1 for male); c is quantitative covariates including 5 genotypic PCs and 10 PEER factors as we used in eQTL mapping, while e is for the residuals. The same parameters were also used for computing the null model but excluding the $s \times g$ term. We then calculated *P* values by comparing the linear interaction model to the null model using analysis of variance. The *lm()* function in R v3.6.3 (*105*) was used for model fitting.

Transcription factor (TF) interacting eQTL. To detect eQTL that may interact with the expression of transcription factors, we retrieved 956 putative transcription factors from the AnimalTFDB (v3.0). As was done for sex-biased QTL detection, we only considered conditionally independent eQTL but excluded eGenes that were TF. Likewise, we fitted the same

interaction model, but the interaction term was TF expression. A total of 15 quantitative covariates including 5 genotypic PCs and 10 PEER factors were also fitted in the model to control confounding factors. The significance threshold was set as FDR(*130*) corrected *P*-value < 0.01.

Cell-type interacting eQTL. We mapped cell-type interaction QTLs by fitting a linear model but included an interaction term implemented in the tensorQTL v1.0.4 (*152*): y = g + i + g × i + e, where y is gene expression, i is the estimated abundance of cell types, and g is genetic effects estimated from SNPs within ±1Mb of the TSS of a gene, while e is for the residuals. To control confounding factors, we also included a total of 15 quantitative covariates including 5 genotypic PCs and 10 PEER factors as described above. We defined genes that had at least one significant SNP after carrying out a multiple testing correction on eigenMT-based *P*-values(*158*) using the FDR approach(*130*). We defined the threshold of significance as FDR < 0.01.

Breed interacting eQTL. To demonstrate breed effects on gene regulation, we ran breed interaction eQTL mapping using the tensorQTL (v1.0.4) tool (*152*) in brain, where the sample size of each breed was > 40, including Leghorn (n=78), Red Jungle Fowl (n = 46) and Ross (n = 157). This interaction eQTL mapping fitted the same model as described for "*Cell-type interaction eQTL"* while the interaction term was breed information. The breed origins were coded as 0 for Red Jungle Fowl and 1 for Leghorn/Ross. After a multiple testing correction using the FDR approach(*130*), gene-variant pairs with FDR < 0.01 were deemed as significant.

Estimating effect sizes of molQTL

We estimated the allelic fold change (aFC) of molQTL by employing the aFC (v0.3) Python script(*159*). The estimation was based on genotypes and molecular phenotypes (the same as molQTL mapping), as well as covariates including 5 genotypic PCs and 10 PEER factors. The 95% confidence interval of aFC was estimated by using the bootstrap method (--boot 100).

Allele specific expression (ASE)

We conducted a haplotype-based ASE analysis through the phASER (v1.1.1) software(*160*). To exclude genomic regions with high mapping error rates, we first computed the genome mappability using GenMap (v1.3.0)(*161*) with parameters: -K 75 -E 2. The generated blacklist was fitted to the phASER (v1.1.1) tool(*160*) to phase variants from the STAR alignment BAM and VCF files with options of "--paired_end 1 --mapq 255 --baseq 10". Using the script "phaser_expr_matrix.py", we measured gene-level haplotypic expression for all samples with default parameters. The generated haplotypic counts files for individual samples were further aggregated by tissue by using "phaser_expr_matrix.py". Finally, we used the "phaser_cis_var.py" script to estimate the effect size of eQTLs based on aggregated haplotypic counts. The correlation of ASE-level effect size (ASE aFC) and eQTL effect size (aFC estimated above) was computed using the Spearman's correlation approach in R v3.6.3 (*105*).

Replication of molQTL discovery

To assess the replication rate of molQTL discovery, we employed the π_1 statistic embedded in the qvalue package(*51*, *157*). Briefly, we randomly split RNA-Seq samples into two groups - QTL discovery and validation population, when the tissue sample size was greater than 100. We

ran QTL mapping in each group separately using tensorQTL (v1.0.4) (*152*) as described above. Based on replicated eQTL *P*-values, we calculated π_0 value that measured the overall proportion of true null hypotheses using the *pi0est* function within the qvalue package(*157*). The π_1 was thus obtained by 1- π_0.

DeepSEA model training and variant effect prediction

DeepSEA is a deep learning model initially trained for predicting variant effects in human(*50*), while in this study, it was retrained by utilizing 310 epigenomic peaks generated by the chicken FAANG consortium (*30*) and by Zhu et al. (*162*) (**Table S7**). According to sequencing type and histone marks, we categorized all 310 epigenomic peaks into groups, including ATAC, CTCF, DNaseSeq, H3K27ac, H3K23me3, H3K4me1 and H3K4me3, which were used as input for the model training using the Selene, a PyTorch-based package (*163*). Briefly, we grouped the genome into 200-bp bins and then labeled the bins according to input features. A genomic bin will be labeled 1 if half of the bin overlaps with an epigenomic peak, otherwise labeled as 0. The model was then trained based on a sequence region of 1,000 bp (i.e., input feature), where the 200-bp bin was placed at the center. We created validation and testing datasets by grouping chromosomes, specifically grouping chromosomes 8 and 9 to the training set and chromosomes 6 and 7 to the validation set. We computed the area under the receiver operating characteristic (AUROC) to evaluate the performance of the DeepSEA model. After that, we computed variant effect/score of two alleles for a given molQTL, i.e. 2 × 310 predicted chromatin variant scores, by inputting a 1000-bp sequence with the center being the Ref or Alt allele. The score is defined as the relative log fold changes of odds between predicted scores of the Ref and Alt. For each feature, SNPs with a score greater than 0.7 were identified as variants affecting the feature.

$$varaint\ score = \left|\log\frac{p(reference)}{1-p(reference)} - \log\frac{p(alternative)}{1-p(alternative)}\right|$$

Functional enrichment of molQTL

To understand the enrichment of molQTL in sequence ontology (i.e., SNP functional types annotated by SnpEff v5.0e) and regulatory elements (i.e., 15 chromatin states annotated in(*30*)), we employed the formula:

$$E = \frac{(C\ /\ A)}{(B\ /\ D)}$$

where A and D are the length of feature annotations and the total genome length, respectively. C is the length of molQTL overlapped with feature annotations, and B is the length of molQTL overlapped with the total genome length. To further uncover the regulatory mechanism of molQTL, we retrieved predicted pairs of regulatory elements-target genes from(*30*). We then overlapped them with our molQTL-regulated genes but at the same time required the molQTL to be located within regulatory elements. Moreover, we also performed the enrichment analysis of molQTL-regulated genes and HiC TAD with data retrieved from a previous study (*52*) using the SnakeHiC pipeline (https://github.com/FarmOmics/SnakeHiC).

Integrating molQTL with GWAS results

GWAS summary statistics. To investigate the regulatory mechanisms underpinning complex traits in pigs, we systematically integrated the identified molQTL with GWAS from 108 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 S17**. 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 1,176 GWAS loci.

Enrichment of molQTL and trait-associated variants. To examine whether molQTL were significantly enriched among the significant GWAS loci, we applied QTLEnrich (v2) (*32*) to quantify the enrichment degree between significant molQTL and GWAS loci.

Transcriptome-wide association study (TWAS). We conducted single- and multi-tissue TWAS with S-PrediXcan(*164*) and S-MultiXcan(*165*) 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(*130*), 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) (*166*), 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(*130*). The gene-trait pairs with corrected *P*-value < 0.05 were 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) (*167*). 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)(*168*, *169*). Leveraging the DAP-G (v1.0.0) (*168*, *169*) outcome, we generated a probabilistic annotation of molQTL using the "summarize_dap2enloc.pl" script. We then calculated approximate LD blocks using PLINK v1.9 (*146*) 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 (*170*) with the options: --load_zval -dump_pip. By integrating GWAS PIP values, we ran the final colocalization analysis with the fastENLOC (v2.0) tool (*171*) and obtained the regional colocalization probability (GRCP). The GRCP > 0.1 was defined as the threshold of significance.

Enrichment analysis of eQTL in selective sweeps

To determine whether domestication could be acting on regulatory variants, we retrieved selective sweeps measured by locus-specific branch length (LSBL) statistics (*14*, *65*). Briefly, we

first calculated F_{ST} for genomic windows with 20 consecutive SNPs between broilers (n=40) and Red Jungle Fowls (RJF, n=35) using VCFtools and also between layers (n=50) and Red Jungle Fowls (n=35). The LSBL values were further computed with the formula: LSBL = ($F_{ST(AB)}$ + $F_{ST(AC)}$ - $F_{ST(BC)}$) / 2. We deemed the top 0.1% of genomic windows ranked by LSBL values as significant. We examined whether eQTL were overrepresented in genomic windows under position selection, *i.e.*, whether genomic windows with at least one eQTL had higher LSBL values than the background, which has an equivalent number of windows as those of eQTL.

Comparative analysis of gene expression

To comparatively analyze gene expression across species, we collected gene expression and regulation data from multiple sources. Specifically, we obtained data for 15,044 samples from the Human GTEx web portal (v8) available at https://gtexportal.org. Additionally, we gathered gene expression data for 7,095 pig samples and 8,742 cattle samples from the FarmGTEx resource accessible at https://www.farmgtex.org/. Furthermore, as part of this study, we included gene expression and regulation data for 7,015 chicken samples. In this study, we focused on protein-coding genes based on the annotation of the Ensembl (v102), and we considered the genes with TPM > 0.1 as expressed. Specifically, we grouped chicken genes into "1-1-1-1 orthologous gene" (1-1 orthologous across species, n = 10600), "complex orthologous genes" ("1 to many", "many to 1" and "many to many", n=3644), "no homology" (without any homologous counterpart in mammals, n=2535). In total, 9 tissues in common across species (*i.e.*, adipose, blood, hypothalamus, liver, lung, muscle, ovary, pituitary, and testis) were included to conduct a comparative analysis of gene expression.

Gene expression (TPM matrix) retrieved for each was normalized using Seurat (v4.3.0) (*110*) to decrease the bias introduced by dynamic data across species. We evaluated transcriptome outcomes by counting the number of reads (reads = TPM × the length of genes (bp)) in each tissue. Samples were by performing dimensionality reduction on the normalized expression data (including 10,600 1-1-1-1 orthologous genes) with the *t*-SNE approach (*117*). Moreover, we explored the conservation of cis-heritability (h^2) and effect size (aFC) of lead eVariants across species. To do so, we selected 1-1-1-1 orthologous genes (n=5,384 for h^2) and eGenes related eVariants (n=5,513 for aFC) that are in common across species, and grouped genes into conserved eGenes (that have at least 1 homology with mammals) and chicken-specific eGenes (that didn't have homology with any mammal species).

Cross-species TWAS comparison

We performed comparative analyses of single-tissue TWAS results of 108 traits in chickens with 9,112, 1,032 and 6,480 single-tissue TWAS results in three mammalian species (*i.e.*, pigs(*47*), cattle(*48*) and humans(*32*)), representing 268, 43 and 135 complex traits, respectively. Within a shared tissue between two species, we computed Pearson's correlations between any pair of 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 carried out permutation analysis by randomly calculating Pearson's correlations between any two of all single-tissue TWAS 1000,000 times. The within-species correlations were then excluded, resulting in 609,861 Pearson's correlations and corresponding *P* values. We set the cutoff of significance as top 0.1% of permuted $-\log_{10}P$, corresponding to the *P* value of 9.11×10^{-3}. We thus conducted cross-species meta-TWAS analysis by combining TWAS results from different

species based on orthologous genes. For meta-TWAS analysis, we applied a sample-size weighting (SSW) strategy (*172*) 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 metaTWAS. 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 (*130*) by replacing original *P* value (TWAS) with *P* value (meta-TWAS) of orthologous genes, the threshold of significance was defined as FDR < 0.05.

References and Notes

1. L. W. Hillier, W. Miller, Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature. 432, 695–716 (2004).
2. D. W. Burt, Emergence of the chicken as a model organism: implications for agriculture and biology. Poultry Science. 86, 1460–1471 (2007).
3. T. H. Beacon, J. R. Davie, The chicken model organism for epigenomic research. Genome. 64, 476–489 (2021).
4. P. Garcia, Y. Wang, J. Viallet, Z. Macek Jilkova, The chicken embryo model: a novel and relevant model for immune-based studies. Frontiers in Immunology. 12, 791081 (2021).
5. D. Wright, C.-J. Rubin, A. Martinez Barrio, K. Schütz, S. Kerje, H. Brändström, A. Kindmark, P. Jensen, L. Andersson, The genetic architecture of domestication in the chicken: effects of pleiotropy and linkage. Molecular Ecology. 19, 5140–5156 (2010).
6. J. Flores-Santin, W. W. Burggren, Beyond the chicken: alternative avian models for developmental physiological research. Front Physiol. 12, 712633 (2021).
7. W. R. A. Brown, S. J. Hubbard, C. Tickle, S. A. Wilson, The chicken as a model for large-scale analysis of vertebrate gene function. Nat Rev Genet. 4, 87–98 (2003).
8. Z. Wu, C. Bortoluzzi, M. F. L. Derks, L. Liu, M. Bosse, S. J. Hiemstra, M. A. M. Groenen, R. P. M. A. Crooijmans, Heterogeneity of a dwarf phenotype in Dutch traditional chicken breeds revealed by genomic analyses. Evolutionary Applications. 14, 1095–1108 (2021).
9. M.-S. Wang, N. O. Otecko, S. Wang, D.-D. Wu, M.-M. Yang, Y.-L. Xu, R. W. Murphy, M.-S. Peng, Y.-P. Zhang, An evolutionary genomic perspective on the breeding of dwarf chickens. Molecular Biology and Evolution. 34, 3081–3088 (2017).
10. M. Lillie, Z. Y. Sheng, C. F. Honaker, L. Andersson, P. B. Siegel, Ö. Carlborg, Genomic signatures of 60 years of bidirectional selection for 8-week body weight in chickens. Poultry Science. 97, 781–790 (2018).
11. J. A. J. van der Eijk, M. B. Verwoolde, G. de Vries Reilingh, C. A. Jansen, T. B. Rodenburg, A. Lammers, Chicken lines divergently selected on feather pecking differ in immune characteristics. Physiology & Behavior. 212, 112680 (2019).
12. Z. Huang, Z. Xu, H. Bai, Y. Huang, N. Kang, X. Ding, J. Liu, H. Luo, C. Yang, W. Chen, Q. Guo, L. Xue, X. Zhang, L. Xu, M. Chen, H. Fu, Y. Chen, Z. Yue, T. Fukagawa, S. Liu, G. Chang, L. Xu, Evolutionary analysis of a complete chicken genome. Proceedings of the National Academy of Sciences. 120, e2216641120 (2023).

13. R. A. Lawal, S. H. Martin, K. Vanmechelen, A. Vereijken, P. Silva, R. M. Al-Atiyat, R. S. Aljumaah, J. M. Mwacharo, D.-D. Wu, Y.-P. Zhang, P. M. Hocking, J. Smith, D. Wragg, O. Hanotte, The wild species genome ancestry of domestic chickens. BMC Biology. 18, 13 (2020).
14. M.-S. Wang, M. Thakur, M.-S. Peng, Y. Jiang, L. A. F. Frantz, M. Li, J.-J. Zhang, S. Wang, J. Peters, N. O. Otecko, C. Suwannapoom, X. Guo, Z.-Q. Zheng, A. Esmailizadeh, N. Y. Hirimuthugoda, H. Ashari, S. Suladari, M. S. A. Zein, S. Kusza, S. Sohrabi, H. Kharrati-Koopaee, Q.-K. Shen, L. Zeng, M.-M. Yang, Y.-J. Wu, X.-Y. Yang, X.-M. Lu, X.-Z. Jia, Q.-H. Nie, S. J. Lamont, E. Lasagna, S. Ceccobelli, H. G. T. N. Gunwardana, T. M. Senasige, S.-H. Feng, J.-F. Si, H. Zhang, J.-Q. Jin, M.-L. Li, Y.-H. Liu, H.-M. Chen, C. Ma, S.-S. Dai, A. K. F. H. Bhuiyan, M. S. Khan, G. L. L. P. Silva, T.-T. Le, O. A. Mwai, M. N. M. Ibrahim, M. Supple, B. Shapiro, O. Hanotte, G. Zhang, G. Larson, J.-L. Han, D.-D. Wu, Y.-P. Zhang, 863 genomes reveal the origin and domestication of chicken. Cell Res. 30, 693–701 (2020).
15. M.-S. Wang, J.-J. Zhang, X. Guo, M. Li, R. Meyer, H. Ashari, Z.-Q. Zheng, S. Wang, M.-S. Peng, Y. Jiang, M. Thakur, C. Suwannapoom, A. Esmailizadeh, N. Y. Hirimuthugoda, M. S. A. Zein, S. Kusza, H. Kharrati-Koopaee, L. Zeng, Y.-M. Wang, T.-T. Yin, M.-M. Yang, M.-L. Li, X.-M. Lu, E. Lasagna, S. Ceccobelli, H. G. T. N. Gunwardana, T. M. Senasig, S.-H. Feng, H. Zhang, A. K. F. H. Bhuiyan, M. S. Khan, G. L. L. P. Silva, L. T. Thuy, O. A. Mwai, M. N. M. Ibrahim, G. Zhang, K.-X. Qu, O. Hanotte, B. Shapiro, M. Bosse, D.-D. Wu, J.-L. Han, Y.-P. Zhang, Large-scale genomic analysis reveals the genetic cost of chicken domestication. BMC Biology. 19, 118 (2021).
16. C.-J. Rubin, M. C. Zody, J. Eriksson, J. R. S. Meadows, E. Sherwood, M. T. Webster, L. Jiang, M. Ingman, T. Sharpe, S. Ka, F. Hallböök, F. Besnier, Ö. Carlborg, B. Bed'hom, M. Tixier-Boichard, P. Jensen, P. Siegel, K. Lindblad-Toh, L. Andersson, Whole-genome resequencing reveals loci under selection during chicken domestication. Nature. 464, 587–591 (2010).
17. S. Qanbari, C.-J. Rubin, K. Maqbool, S. Weigend, A. Weigend, J. Geibel, S. Kerje, C. Wurmser, A. T. Peterson, I. L. B. Jr, R. Preisinger, R. Fries, H. Simianer, L. Andersson, Genetics of adaptation in modern chicken. PLOS Genetics. 15, e1007989 (2019).
18. C. Zhang, D. Lin, Y. Wang, D. Peng, H. Li, J. Fei, K. Chen, N. Yang, X. Hu, Y. Zhao, N. Li, Widespread introgression in Chinese indigenous chicken breeds from commercial broiler. Evolutionary Applications. 12, 610–621 (2019).
19. M. Y. Wu, G. W. Low, G. Forcina, H. van Grouw, B. P. Y.-H. Lee, R. R. Y. Oh, F. E. Rheindt, Historic and modern genomes unveil a domestic introgression gradient in a wild red junglefowl population. Evolutionary Applications. 13, 2300–2315 (2020).
20. M. Johnsson, E. Gering, P. Willis, S. Lopez, L. Van Dorp, G. Hellenthal, R. Henriksen, U. Friberg, D. Wright, Feralisation targets different genomic loci to domestication in the chicken. Nat Commun. 7, 12950 (2016).
21. E. Gering, M. Johnsson, P. Willis, T. Getty, D. Wright, Mixed ancestry and admixture in Kauai's feral chickens: invasion of domestic genes into ancient Red Junglefowl reservoirs. Molecular Ecology. 24, 2112–2124 (2015).
22. A. A. Gheyas, A. Vallejo-Trujillo, A. Kebede, M. Lozano-Jaramillo, T. Dessie, J. Smith, O. Hanotte, Integrated environmental and genomic analysis reveals the drivers of local adaptation in african indigenous chickens. Molecular Biology and Evolution. 38, 4268–4285 (2021).

23. Y. Zan, Z. Sheng, M. Lillie, L. Rönnegård, C. F. Honaker, P. B. Siegel, Ö. Carlborg, Artificial selection response due to polygenic adaptation from a multilocus, multiallelic genetic architecture. Molecular Biology and Evolution. 34, 2678–2689 (2017).
24. S. Shi, D. Shao, L. Yang, Q. Liang, W. Han, Q. Xue, L. Qu, L. Leng, Y. Li, X. Zhao, P. Dong, M. Walugembe, B. B. Kayang, A. P. Muhairwa, H. Zhou, H. Tong, Whole genome analyses reveal novel genes associated with chicken adaptation to tropical and frigid environments. Journal of Advanced Research. 47, 13-25 (2022).
25. Z.-L. Hu, C. A. Park, J. M. Reecy, Bringing the Animal QTLdb and CorrDB into the future: meeting new challenges and providing updated services. Nucleic Acids Research. 50, D956–D961 (2022).
26. M. E. Goddard, B. J. Hayes, Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet. 10, 381–391 (2009).
27. L. Andersson, M. Georges, Domestic-animal genomics: deciphering the genetics of complex traits. Nat Rev Genet. 5, 202–212 (2004).
28. J. Smith, J. M. Alfieri, N. Anthony, P. Arensburger, G. N. Athrey, J. Balacco, A. Balic, P. Bardou, P. Barela, Y. Bigot, H. Blackmon, P. M. Borodin, R. Carroll, M. C. Casono, M. Charles, H. Cheng, M. Chiodi, L. Cigan, L. M. Coghill, R. Crooijmans, N. Das, S. Davey, A. Davidian, F. Degalez, J. M. Dekkers, M. Derks, A. B. Diack, A. Djikeng, Y. Drechsler, A. Dyomin, O. Fedrigo, S. R. Fiddaman, G. Formenti, L. A. F. Frantz, J. E. Fulton, E. Gaginskaya, S. Galkina, R. A. Gallardo, J. Geibel, A. Gheyas, C. J. P. Godinez, A. Goodell, J. A. M. Graves, D. K. Griffin, B. Haase, J.-L. Han, O. Hanotte, L. J. Henderson, Z.-C. Hou, K. Howe, L. Huynh, E. Ilatsia, E. Jarvis, S. M. Johnson, J. Kaufman, T. Kelly, S. Kemp, C. Kern, J. H. Keroack, C. Klopp, S. Lagarrigue, S. J. Lamont, M. Lange, A. Lanke, D. M. Larkin, G. Larson, J. K. N. Layos, O. Lebrasseur, L. P. Malinovskaya, R. J. Martin, M. L. M. Cerezo, A. S. Mason, F. M. McCarthy, M. J. McGrew, J. Mountcastle, C. K. Muhonja, W. Muir, K. Muret, T. Murphy, I. Ng'ang'a, M. Nishibori, R. E. O'Connor, M. Ogugo, R. Okimoto, O. Ouko, H. R. Patel, F. Perini, M. I. Pigozzi, K. C. Potter, P. D. Price, C. Reimer, E. S. Rice, N. Rocos, T. F. Rogers, P. Saelao, J. Schauer, R. Schnabel, V. Schneider, H. Simianer, A. Smith, M. P. Stevens, K. Stiers, C. K. Tiambo, M. Tixier-Boichard, A. A. Torgasheva, A. Tracey, C. A. Tregaskes, L. Vervelde, Y. Wang, W. C. Warren, P. D. Waters, D. Webb, S. Weigend, A. Wolc, A. E. Wright, D. Wright, Z. Wu, M. Yamagata, C. Yang, Z.-T. Yin, M. C. Young, G. Zhang, B. Zhao, H. Zhou, Fourth Report on Chicken Genes and Chromosomes 2022. CGR (2023), doi:10.1159/000529376.
29. C. Kern, Y. Wang, X. Xu, Z. Pan, M. Halstead, G. Chanthavixay, P. Saelao, S. Waters, R. Xiang, A. Chamberlain, I. Korf, M. E. Delany, H. H. Cheng, J. F. Medrano, A. L. Van Eenennaam, C. K. Tuggle, C. Ernst, P. Flicek, G. Quon, P. Ross, H. Zhou, Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. Nat Commun. 12, 1821 (2021).
30. Z. Pan, An atlas of regulatory elements in chicken: a resource for chicken genetics and genomics. Science Advances. 9, eade120 (2023).
31. F. W. Albert, L. Kruglyak, The role of regulatory variation in complex traits and disease. Nat Rev Genet. 16, 197–212 (2015).
32. The GTEx Consortium, The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science. 369, 1318–1330 (2020).
33. The GTEx Consortium, Genetic effects on gene expression across human tissues. Nature. 550, 204–213 (2017).
34. The GTEx Consortium, The Genotype-Tissue Expression (GTEx) pilot analysis: Multitissue gene regulation in humans. Science. 348, 648–660 (2015).

35. N. Kerimov, J. D. Hayhurst, K. Peikova, J. R. Manning, P. Walter, L. Kolberg, M. Samoviča, M. P. Sakthivel, I. Kuzmin, S. J. Trevanion, T. Burdett, S. Jupp, H. Parkinson, I. Papatheodorou, A. D. Yates, D. R. Zerbino, K. Alasoo, A compendium of uniformly processed human gene expression and splicing quantitative trait loci. Nat Genet. 53, 1–10 (2021).
36. M. Johnsson, K. B. Jonsson, L. Andersson, P. Jensen, D. Wright, Genetic regulation of bone metabolism in the chicken: similarities and differences to mammalian systems. PLOS Genetics. 11, e1005250 (2015).
37. A. Höglund, K. Strempfl, J. Fogelholm, D. Wright, R. Henriksen, The genetic regulation of size variation in the transcriptome of the cerebrum in the chicken and its role in domestication and brain size evolution. BMC Genomics. 21, 518 (2020).
38. C. Falker-Gieske, J. Bennewitz, J. Tetens, Structural variation and eQTL analysis in two experimental populations of chickens divergently selected for feather-pecking behavior. Neurogenetics. 24, 29–41 (2023).
39. A. C. Mott, A. Mott, S. Preuß, J. Bennewitz, J. Tetens, C. Falker-Gieske, eQTL analysis of laying hens divergently selected for feather pecking identifies KLF14 as a potential key regulator for this behavioral disorder. Frontiers in Genetics. 13, 969752 (2022).
40. A. Höglund, R. Henriksen, J. Fogelholm, A. M. Churcher, C. M. Guerrero-Bosagna, A. Martinez-Barrio, M. Johnsson, P. Jensen, D. Wright, The methylation landscape and its role in domestication and gene regulation in the chicken. Nat Ecol Evol. 4, 1713–1724 (2020).
41. S. Liu, L. Fang, The CattleGTEx atlas reveals regulatory mechanisms underlying complex traits. Nat Genet. 54, 1273–1274 (2022).
42. H. Ellegren, L. Hultin-Rosenberg, B. Brunström, L. Dencker, K. Kultima, B. Scholz, Faced with inequality: chicken do not have a general dosage compensation of sex-linked genes. BMC Biology. 5, 40 (2007).
43. F. W. Nicholas, 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).
44. Z. Wang, L. Qu, J. Yao, X. Yang, G. Li, Y. Zhang, J. Li, X. Wang, J. Bai, G. Xu, X. Deng, N. Yang, C. Wu, An EAV-HP insertion in 5′ flanking region of SLCO1B3 causes blue eggshell in the chicken. PLOS Genetics. 9, e1003183 (2013).
45. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010).
46. Y. Zou, P. Carbonetto, G. Wang, M. Stephens, Fine-mapping from summary data with the "Sum of Single Effects" model. PLOS Genetics. 18, e1010299 (2022).
47. The FarmGTEx-PigGTEx Consortium, J. Teng, Y. Gao, H. Yin, Z. Bai, S. Liu, H. Zeng, L. Bai, Z. Cai, B. Zhao, X. Li, Z. Xu, Q. Lin, Z. Pan, W. Yang, X. Yu, D. Guan, Y. Hou, B. N. Keel, G. A. Rohrer, A. K. Lindholm-Perry, W. T. Oliver, M. Ballester, D. Crespo-Piazuelo, R. Quintanilla, O. Canela-Xandri, K. Rawlik, C. Xia, Y. Yao, Q. Zhao, W. Yao, L. Yang, H. Li, H. Zhang, W. Liao, T. Chen, P. Karlskov-Mortensen, M. Fredholm, M. Amills, A. Clop, E. Giuffra, J. Wu, X. Cai, S. Diao, X. Pan, C. Wei, J. Li, H. Cheng, S. Wang, G. Su, G. Sahana, M. S. Lund, J. C. M. Dekkers, L. Kramer, C. K. Tuggle, R. Corbett, M. A. M. Groenen, O. Madsen, M. Gòdia, D. Rocha, M. Charles, C. Li, H. Pausch, X. Hu, L. Frantz, Y. Luo, L. Lin, Z. Zhou, Z. Zhang, Z. Chen, L. Cui, R. Xiang, X. Shen, P. Li, R. Huang, G. Tang, M. Li, Y. Zhao, G. Yi, Z. Tang, J. Jiang, F. Zhao, X. Yuan, X. Liu, Y. Chen, X. Xu, S. Zhao, P. Zhao, C. Haley, H. Zhou, Q. Wang, Y. Pan, X. Ding, L. Ma, J. Li, P. Navarro, Q.

Zhang, B. Li, A. Tenesa, K. Li, G. E. Liu, Z. Zhang, L. Fang, A compendium of genetic regulatory effects across pig tissues (2022), doi:10.1101/2022.11.11.516073.

48. S. Liu, Y. Gao, O. Canela-Xandri, S. Wang, Y. Yu, W. Cai, B. Li, R. Xiang, A. J. Chamberlain, E. Pairo-Castineira, K. D'Mellow, K. Rawlik, C. Xia, Y. Yao, P. Navarro, D. Rocha, X. Li, Z. Yan, C. Li, B. D. Rosen, C. P. Van Tassell, P. M. Vanraden, S. Zhang, L. Ma, J. B. Cole, G. E. Liu, A. Tenesa, L. Fang, A multi-tissue atlas of regulatory variants in cattle. Nat Genet, 1–10 (2022).

49. E. Axelsson, M. T. Webster, N. G. C. Smith, D. W. Burt, H. Ellegren, Comparison of the chicken and turkey genomes reveals a higher rate of nucleotide divergence on microchromosomes than macrochromosomes. Genome Res. 15, 120–125 (2005).

50. J. Zhou, O. G. Troyanskaya, Predicting effects of noncoding variants with deep learning–based sequence model. Nat Methods. 12, 931–934 (2015).

51. J. D. Storey, R. Tibshirani, Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences. 100, 9440–9445 (2003).

52. D. Guan, Ying Wang, S. E. Aggrey, R. Okimoto, R. Hawken, H. Zhou, Profiling chromatin contacts at micro-scale in the chicken genome. International Plant and Animal Genome Conference 30. San Diego, USA (2022).

53. C. Robins, Y. Liu, W. Fan, D. M. Duong, J. Meigs, N. V. Harerimana, E. S. Gerasimov, E. B. Dammer, D. J. Cutler, T. G. Beach, E. M. Reiman, P. L. D. Jager, D. A. Bennett, J. J. Lah, A. P. Wingo, A. I. Levey, N. T. Seyfried, T. S. Wingo, Genetic control of the human brain proteome. The American Journal of Human Genetics. 108, 400–410 (2021).

54. X. Sun, Z. He, L. Guo, C. Wang, C. Lin, L. Ye, X. Wang, Y. Li, M. Yang, S. Liu, X. Hua, W. Wen, C. Lin, Z. Long, W. Zhang, H. Li, Y. Jian, Z. Zhu, X. Wu, H. Lin, ALG3 contributes to stemness and radioresistance through regulating glycosylation of TGF-β receptor II in breast cancer. Journal of Experimental & Clinical Cancer Research. 40, 149 (2021).

55. J. Chen, X. Zhao, H. Wang, Y. Chen, W. Wang, W. Zhou, X. Wang, J. Tang, Y. Zhao, X. Lu, S. Chen, L. Wang, C. Shen, S. Yang, Common variants in TGFBR2 and miR-518 genes are associated with hypertension in the Chinese population. American Journal of Hypertension. 27, 1268–1276 (2014).

56. G. Abou-Ezzi, T. Supakorndej, J. Zhang, B. Anthony, J. Krambs, H. Celik, D. Karpova, C. S. Craft, D. C. Link, TGF-β signaling plays an essential role in the lineage specification of mesenchymal stem/progenitor cells in fetal bone marrow. Stem Cell Reports. 13, 48–60 (2019).

57. M. H. Gouveia, A. R. Bentley, H. Leonard, K. A. C. Meeks, K. Ekoru, G. Chen, M. A. Nalls, E. M. Simonsick, E. Tarazona-Santos, M. F. Lima-Costa, A. Adeyemo, D. Shriner, C. N. Rotimi, Trans-ethnic meta-analysis identifies new loci associated with longitudinal blood pressure traits. Sci Rep. 11, 4075 (2021).

58. D. Noda, S. Itoh, Y. Watanabe, M. Inamitsu, S. Dennler, F. Itoh, S. Koike, D. Danielpour, P. ten Dijke, M. Kato, ELAC2, a putative prostate cancer susceptibility gene product, potentiates TGF-β/Smad-induced growth arrest of prostate cells. Oncogene. 25, 5591–5600 (2006).

59. H. Hu, Y.-R. Miao, L.-H. Jia, Q.-Y. Yu, Q. Zhang, A.-Y. Guo, AnimalTFDB 3.0: a comprehensive resource for annotation and prediction of animal transcription factors. Nucleic Acids Research. 47, D33–D38 (2019).

60. A. S. Sowa, E. Martin, I. M. Martins, J. Schmidt, R. Depping, J. J. Weber, F. Rother, E. Hartmann, M. Bader, O. Riess, H. Tricoire, T. Schmidt, 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).

61. J. Pan, Y. Yang, B. Yang, Y. Yu, Artificial Polychromatic Light Affects Growth and Physiology in Chicks. PLOS ONE. 9, e113595 (2014).

62. J. Cao, Z. Wang, Y. Dong, Z. Zhang, J. Li, F. Li, Y. Chen, Effect of combinations of monochromatic lights on growth and productive performance of broilers. Poultry Science. 91, 3013–3018 (2012).

63. M. A. Estermann, A. T. Major, C. A. Smith, DMRT1 regulation of TOX3 modulates expansion of the gonadal steroidogenic cell lineage (2022), p. 2022.07.29.502037, doi:10.1101/2022.07.29.502037.

64. X. Zhang, J. Li, X. Wang, Y. Jie, C. Sun, J. Zheng, J. Li, N. Yang, S. Chen, ATAC-seq and RNA-seq analysis unravel the mechanism of sex differentiation and infertility in sex reversal chicken. Epigenetics & Chromatin. 16, 2 (2023).

65. M.-S. Wang, Y.-X. Huo, Y. Li, N. O. Otecko, L.-Y. Su, H.-B. Xu, S.-F. Wu, M.-S. Peng, H.-Q. Liu, L. Zeng, D. M. Irwin, Y.-G. Yao, D.-D. Wu, Y.-P. Zhang, Comparative population genomics reveals genetic basis underlying body size of domestic chickens. Journal of Molecular Cell Biology. 8, 542–552 (2016).

66. K.-Y. Park, H.-S. Hwang, K.-H. Cho, K. Han, G. E. Nam, Y. H. Kim, Y. Kwon, Y.-G. Park, Body weight fluctuation as a risk factor for type 2 diabetes: results from a nationwide cohort study. J Clin Med. 8, 950 (2019).

67. D. P. Berry, J. M. Lee, K. A. Macdonald, K. Stafford, L. Matthews, J. R. Roche, Associations among body condition score, body weight, somatic cell count, and clinical mastitis in seasonally calving dairy cattle. Journal of Dairy Science. 90, 637–648 (2007).

68. C. S. Fox, N. Heard-Costa, L. A. Cupples, J. Dupuis, R. S. Vasan, L. D. Atwood, Genome-wide association to body mass index and waist circumference: the Framingham Heart Study 100K project. BMC Medical Genetics. 8, S18 (2007).

69. A. Rajendran, K. Vaidya, J. Mendoza, J. Bridwell-Rabb, S. S. Kamat, Functional annotation of ABHD14B, an orphan serine hydrolase enzyme. Biochemistry. 59, 183–196 (2020).

70. A. Rajendran, A. Soory, N. Khandelwal, G. Ratnaparkhi, S. S. Kamat, A multi-omics analysis reveals that the lysine deacetylase ABHD14B influences glucose metabolism in mammals. J Biol Chem. 298, 102128 (2022).

71. A. Ragvin, E. Moro, D. Fredman, P. Navratilova, Ø. Drivenes, P. G. Engström, M. E. Alonso, E. de la C. Mustienes, J. L. G. Skarmeta, M. J. Tavares, F. Casares, M. Manzanares, V. van Heyningen, A. Molven, P. R. Njølstad, F. Argenton, B. Lenhard, T. S. Becker, Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. Proceedings of the National Academy of Sciences. 107, 775–780 (2010).

72. S. C. Collins, H. W. Do, B. Hastoy, A. Hugill, J. Adam, M. V. Chibalina, J. Galvanovskis, M. Godazgar, S. Lee, M. Goldsworthy, A. Salehi, A. I. Tarasov, A. H. Rosengren, R. Cox, P. Rorsman, Increased expression of the diabetes gene SOX4 reduces insulin secretion by impaired fusion pore expansion. Diabetes. 65, 1952–1961 (2016).

73. J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, J. P. Mesirov, Integrative genomics viewer. Nat Biotechnol. 29, 24–26 (2011).

74. H. J. Taylor, Y.-H. Hung, N. Narisu, M. R. Erdos, M. Kanke, T. Yan, C. M. Grenko, A. J. Swift, L. L. Bonnycastle, P. Sethupathy, F. S. Collins, D. L. Taylor, Human pancreatic islet microRNAs implicated in diabetes and related traits by large-scale genetic analysis (2022), doi:10.1101/2022.04.21.489048.

75. T. Qi, Y. Wu, H. Fang, F. Zhang, S. Liu, J. Zeng, J. Yang, Genetic control of RNA splicing and its distinct role in complex trait variation. Nat Genet, 1–9 (2022).

76. L. Li, K.-L. Huang, Y. Gao, Y. Cui, G. Wang, N. D. Elrod, Y. Li, Y. E. Chen, P. Ji, F. Peng, W. K. Russell, E. J. Wagner, W. Li, An atlas of alternative polyadenylation quantitative trait loci contributing to complex trait and disease heritability. Nat Genet. 53, 994–1005 (2021).
77. D. W. Burt, Chicken genome: Current status and future opportunities. Genome Res. 15, 1692–1698 (2005).
78. M. Bergman, N. Ringertz, Gene expression pattern of chicken erythrocyte nuclei in heterokaryons. Journal of Cell Science. 97, 167–175 (1990).
79. C. Désert, E. Merlot, T. Zerjal, B. Bed'hom, S. Härtle, A. Le Cam, P.-F. Roux, E. Baeza, F. Gondret, M. J. Duclos, S. Lagarrigue, Transcriptomes of whole blood and PBMC in chickens. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics. 20, 1–9 (2016).
80. N. J. Connally, S. Nazeen, D. Lee, H. Shi, J. Stamatoyannopoulos, S. Chun, C. Cotsapas, C. A. Cassa, S. R. Sunyaev, The missing link between genetic association and regulatory function. eLife. 11, e74970 (2022).
81. Y. Hasin, M. Seldin, A. Lusis, Multi-omics approaches to disease. Genome Biology. 18, 83 (2017).
82. U. Võsa, A. Claringbould, H.-J. Westra, M. J. Bonder, P. Deelen, B. Zeng, H. Kirsten, A. Saha, R. Kreuzhuber, S. Yazar, H. Brugge, R. Oelen, D. H. de Vries, M. G. P. van der Wijst, S. Kasela, N. Pervjakova, I. Alves, M.-J. Favé, M. Agbessi, M. W. Christiansen, R. Jansen, I. Seppälä, L. Tong, A. Teumer, K. Schramm, G. Hemani, J. Verlouw, H. Yaghootkar, R. Sönmez Flitman, A. Brown, V. Kukushkina, A. Kalnapenkis, S. Rüeger, E. Porcu, J. Kronberg, J. Kettunen, B. Lee, F. Zhang, T. Qi, J. A. Hernandez, W. Arindrarto, F. Beutner, J. Dmitrieva, M. Elansary, B. P. Fairfax, M. Georges, B. T. Heijmans, A. W. Hewitt, M. Kähönen, Y. Kim, J. C. Knight, P. Kovacs, K. Krohn, S. Li, M. Loeffler, U. M. Marigorta, H. Mei, Y. Momozawa, M. Müller-Nurasyid, M. Nauck, M. G. Nivard, B. W. J. H. Penninx, J. K. Pritchard, O. T. Raitakari, O. Rotzschke, E. P. Slagboom, C. D. A. Stehouwer, M. Stumvoll, P. Sullivan, P. A. C. 't Hoen, J. Thiery, A. Tönjes, J. van Dongen, M. van Iterson, J. H. Veldink, U. Völker, R. Warmerdam, C. Wijmenga, M. Swertz, A. Andiappan, G. W. Montgomery, S. Ripatti, M. Perola, Z. Kutalik, E. Dermitzakis, S. Bergmann, T. Frayling, J. van Meurs, H. Prokisch, H. Ahsan, B. L. Pierce, T. Lehtimäki, D. I. Boomsma, B. M. Psaty, S. A. Gharib, P. Awadalla, L. Milani, W. H. Ouwehand, K. Downes, O. Stegle, A. Battle, P. M. Visscher, J. Yang, M. Scholz, J. Powell, G. Gibson, T. Esko, L. Franke, Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. Nat Genet. 53, 1300–1310 (2021).
83. N. de Klein, E. A. Tsai, M. Vochteloo, D. Baird, Y. Huang, C.-Y. Chen, S. van Dam, R. Oelen, P. Deelen, O. B. Bakker, O. El Garwany, Z. Ouyang, E. E. Marshall, M. I. Zavodszky, W. van Rheenen, M. K. Bakker, J. Veldink, T. R. Gaunt, H. Runz, L. Franke, H.-J. Westra, Brain expression quantitative trait locus and network analyses reveal downstream effects and putative drivers for brain-related diseases. Nat Genet, 1–12 (2023).
84. B. Zeng, J. Bendl, R. Kosoy, J. F. Fullard, G. E. Hoffman, P. Roussos, Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. Nat Genet, 1–9 (2022).
85. J. Lonsdale, J. Thomas, M. Salvatore, R. Phillips, E. Lo, S. Shad, R. Hasz, G. Walters, F. Garcia, N. Young, B. Foster, M. Moser, E. Karasik, B. Gillard, K. Ramsey, S. Sullivan, J. Bridge, H. Magazine, J. Syron, J. Fleming, L. Siminoff, H. Traino, M. Mosavel, L. Barker, S. Jewell, D. Rohrer, D. Maxim, D. Filkins, P. Harbach, E. Cortadillo, B. Berghuis, L. Turner, E. Hudson, K. Feenstra, L. Sobin, J. Robb, P. Branton, G. Korzeniewski, C. Shive, D. Tabor, L. Qi, K. Groch, S. Nampally, S. Buia, A. Zimmerman, A. Smith, R. Burges, K.

Robinson, K. Valentino, D. Bradbury, M. Cosentino, N. Diaz-Mayoral, M. Kennedy, T. Engel, P. Williams, K. Erickson, K. Ardlie, W. Winckler, G. Getz, D. DeLuca, D. MacArthur, M. Kellis, A. Thomson, T. Young, E. Gelfand, M. Donovan, Y. Meng, G. Grant, D. Mash, Y. Marcus, M. Basile, J. Liu, J. Zhu, Z. Tu, N. J. Cox, D. L. Nicolae, E. R. Gamazon, H. K. Im, A. Konkashbaev, J. Pritchard, M. Stevens, T. Flutre, X. Wen, E. T. Dermitzakis, T. Lappalainen, R. Guigo, J. Monlong, M. Sammeth, D. Koller, A. Battle, S. Mostafavi, M. McCarthy, M. Rivas, J. Maller, I. Rusyn, A. Nobel, F. Wright, A. Shabalin, M. Feolo, N. Sharopova, A. Sturcke, J. Paschal, J. M. Anderson, E. L. Wilder, L. K. Derr, E. D. Green, J. P. Struewing, G. Temple, S. Volpi, J. T. Boyer, E. J. Thomson, M. S. Guyer, C. Ng, A. Abdallah, D. Colantuoni, T. R. Insel, S. E. Koester, A. R. Little, P. K. Bender, T. Lehner, Y. Yao, C. C. Compton, J. B. Vaught, S. Sawyer, N. C. Lockhart, J. Demchok, H. F. Moore, The Genotype-Tissue Expression (GTEx) project. Nat Genet. 45, 580–585 (2013).

86. M. Georges, C. Charlier, B. Hayes, Harnessing genomic information for livestock improvement. Nat Rev Genet. 20, 135–156 (2019).

87. J. R. Xue, A. Mackay-Smith, K. Mouri, M. F. Garcia, M. X. Dong, J. F. Akers, M. Noble, X. Li, ZOONOMIA CONSORTIUM, K. Lindblad-Toh, E. K. Karlsson, J. P. Noonan, T. D. Capellini, K. J. Brennand, R. Tewhey, P. C. Sabeti, S. K. Reilly, The functional and evolutionary impacts of human-specific deletions in conserved elements. Science. 380, eabn2253 (2023).

88. I. M. Kaplow, A. J. Lawler, D. E. Schäffer, C. Srinivasan, M. E. Wirthlin, B. N. Phan, X. Zhang, K. Foley, K. Prasad, A. R. Brown, Z. Consortium, W. K. Meyer, A. R. Pfenning, Relating enhancer genetic variation across mammals to complex phenotypes using machine learning (2022), doi:10.1101/2022.08.26.505436.

89. G. Andrews, K. Fan, H. E. Pratt, N. Phalke, ZOONOMIA CONSORTIUM, E. K. Karlsson, K. Lindblad-Toh, S. Gazal, J. E. Moore, Z. Weng, Mammalian evolution of human cis-regulatory elements and transcription factor binding sites. Science. 380, eabn7930 (2023).

90. F. W. Albert, L. Kruglyak, The role of regulatory variation in complex traits and disease. Nat Rev Genet. 16, 197–212 (2015).

91. G. Sella, N. H. Barton, Thinking about the evolution of complex traits in the era of genome-wide association studies. Annual Review of Genomics and Human Genetics. 20, 461–493 (2019).

92. M. Li, C. Sun, N. Xu, P. Bian, X. Tian, X. Wang, Y. Wang, X. Jia, R. Heller, M. Wang, F. Wang, X. Dai, R. Luo, Y. Guo, X. Wang, P. Yang, D. Hu, Z. Liu, W. Fu, S. Zhang, X. Li, C. Wen, F. Lan, A. Z. Siddiki, C. Suwannapoom, X. Zhao, Q. Nie, X. Hu, Y. Jiang, N. Yang, De novo assembly of 20 chicken genomes reveals the undetectable phenomenon for thousands of core genes on micro-chromosomes and sub-telomeric regions. Molecular Biology and Evolution, msac066 (2022).

93. K. Wang, H. Hu, Y. Tian, J. Li, A. Scheben, C. Zhang, Y. Li, J. Wu, L. Yang, X. Fan, G. Sun, D. Li, Y. Zhang, R. Han, R. Jiang, H. Huang, F. Yan, Y. Wang, Z. Li, G. Li, X. Liu, W. Li, D. Edwards, X. Kang, The chicken pan-genome reveals gene content variation and a promoter region deletion in IGF2BP1 affecting body size. Molecular Biology and Evolution (2021), doi:10.1093/molbev/msab231.

94. A. Rhie, S. A. McCarthy, O. Fedrigo, J. Damas, G. Formenti, S. Koren, M. Uliano-Silva, W. Chow, A. Fungtammasan, J. Kim, C. Lee, B. J. Ko, M. Chaisson, G. L. Gedman, L. J. Cantin, F. Thibaud-Nissen, L. Haggerty, I. Bista, M. Smith, B. Haase, J. Mountcastle, S. Winkler, S. Paez, J. Howard, S. C. Vernes, T. M. Lama, F. Grutzner, W. C. Warren, C. N. Balakrishnan, D. Burt, J. M. George, M. T. Biegler, D. Iorns, A. Digby, D. Eason, B. Robertson, T. Edwards, M. Wilkinson, G. Turner, A. Meyer, A. F. Kautt, P. Franchini, H.

W. Detrich, H. Svardal, M. Wagner, G. J. P. Naylor, M. Pippel, M. Malinsky, M. Mooney, M. Simbirsky, B. T. Hannigan, T. Pesout, M. Houck, A. Misuraca, S. B. Kingan, R. Hall, Z. Kronenberg, I. Sović, C. Dunn, Z. Ning, A. Hastie, J. Lee, S. Selvaraj, R. E. Green, N. H. Putnam, I. Gut, J. Ghurye, E. Garrison, Y. Sims, J. Collins, S. Pelan, J. Torrance, A. Tracey, J. Wood, R. E. Dagnew, D. Guan, S. E. London, D. F. Clayton, C. V. Mello, S. R. Friedrich, P. V. Lovell, E. Osipova, F. O. Al-Ajli, S. Secomandi, H. Kim, C. Theofanopoulou, M. Hiller, Y. Zhou, R. S. Harris, K. D. Makova, P. Medvedev, J. Hoffman, P. Masterson, K. Clark, F. Martin, K. Howe, P. Flicek, B. P. Walenz, W. Kwak, H. Clawson, M. Diekhans, L. Nassar, B. Paten, R. H. S. Kraus, A. J. Crawford, M. T. P. Gilbert, G. Zhang, B. Venkatesh, R. W. Murphy, K.-P. Koepfli, B. Shapiro, W. E. Johnson, F. Di Palma, T. Marques-Bonet, E. C. Teeling, T. Warnow, J. M. Graves, O. A. Ryder, D. Haussler, S. J. O'Brien, J. Korlach, H. A. Lewin, K. Howe, E. W. Myers, R. Durbin, A. M. Phillippy, E. D. Jarvis, Towards complete and error-free genome assemblies of all vertebrate species. Nature. 592, 737–746 (2021).

95. D. Guan, M. M. Halstead, A. D. Islas-Trejo, D. E. Goszczynski, H. H. Cheng, P. J. Ross, H. Zhou, Prediction of transcript isoforms in 19 chicken tissues by Oxford Nanopore long-read sequencing. Frontiers in Genetics. 13, 997460 (2022).

96. R. I. Kuo, E. Tseng, L. Eory, I. R. Paton, A. L. Archibald, D. W. Burt, Normalized long read RNA sequencing in chicken reveals transcriptome complexity similar to human. BMC Genomics. 18, 323 (2017).

97. S. Thomas, J. G. Underwood, E. Tseng, A. K. Holloway, on behalf of the Bench To Basinet CvDC Informatics Subcommittee, Long-read sequencing of chicken transcripts and identification of new transcript isoforms. PLOS ONE. 9, e94650 (2014).

98. J. Zhang, C. Nie, X. Li, X. Zhao, Y. Jia, J. Han, Y. Chen, L. Wang, X. Lv, W. Yang, K. Li, J. Zhang, Z. Ning, H. Bao, C. Zhao, J. Li, L. Qu, Comprehensive analysis of structural variants in chickens using PacBio sequencing. Frontiers in Genetics. 13, 971588 (2022).

99. S. Kojima, S. Koyama, M. Ka, Y. Saito, E. H. Parrish, M. Endo, S. Takata, M. Mizukoshi, K. Hikino, A. Takeda, A. F. Gelinas, S. M. Heaton, R. Koide, A. J. Kamada, M. Noguchi, M. Hamada, Y. Kamatani, Y. Murakawa, K. Ishigaki, Y. Nakamura, K. Ito, C. Terao, Y. Momozawa, N. F. Parrish, Mobile element variation contributes to population-specific genome diversification, gene regulation and disease risk. Nat Genet, 1–13 (2023).

100. T. Wicker, J. S. Robertson, S. R. Schulze, F. A. Feltus, V. Magrini, J. A. Morrison, E. R. Mardis, R. K. Wilson, D. G. Peterson, A. H. Paterson, R. Ivarie, The repetitive landscape of the chicken genome. Genome Res. 15, 126–136 (2005).

101. S. Kim-Hellmuth, F. Aguet, M. Oliva, M. Muñoz-Aguirre, S. Kasela, V. Wucher, S. E. Castel, A. R. Hamel, A. Viñuela, A. L. Roberts, S. Mangul, X. Wen, G. Wang, A. N. Barbeira, D. Garrido-Martín, B. B. Nadel, Y. Zou, R. Bonazzola, J. Quan, A. Brown, A. Martinez-Perez, J. M. Soria, Gte. Consortium§, G. Getz, E. T. Dermitzakis, K. S. Small, M. Stephens, H. S. Xi, H. K. Im, R. Guigó, A. V. Segrè, B. E. Stranger, K. G. Ardlie, T. Lappalainen, Cell type–specific genetic regulation of gene expression across human tissues. Science. 369, 1332 (2020).

102. J. A. Morris, C. Caragine, Z. Daniloski, J. Domingo, T. Barry, L. Lu, K. Davis, M. Ziosi, D. A. Glinos, S. Hao, E. P. Mimitou, P. Smibert, K. Roeder, E. Katsevich, T. Lappalainen, N. E. Sanjana, Discovery of target genes and pathways at GWAS loci by pooled single-cell CRISPR screens. Science. 380, eadh7699 (2023).

103. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation. 0 (2021), doi:10.1016/j.xinn.2021.100141.

104. H. Wickham, ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag New York, 2016).
105. R Core Team, R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. (2022). Available at https://www.R-project.org/.
106. P. Cingolani, A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, D. M. Ruden, A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly. 6, 80–92 (2012).
107. J. Zhou, O. G. Troyanskaya, Predicting effects of noncoding variants with deep learning–based sequence model. Nat Methods. 12, 931–934 (2015).
108. "Complete Linkage Method" in The Concise Encyclopedia of Statistics (Springer, New York, NY, 2008; https://doi.org/10.1007/978-0-387-32833-1_71), pp. 102–102.
109. S. M. Urbut, G. Wang, P. Carbonetto, M. Stephens, Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. Nat Genet. 51, 187–195 (2019).
110. R. Satija, J. A. Farrell, D. Gennert, A. F. Schier, A. Regev, Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 33, 495–502 (2015).
111. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29, 15–21 (2013).
112. F. Jehl, K. Muret, M. Bernard, M. Boutin, L. Lagoutte, C. Désert, P. Dehais, D. Esquerré, H. Acloque, E. Giuffra, S. Djebali, S. Foissac, T. Derrien, F. Pitel, T. Zerjal, C. Klopp, S. Lagarrigue, An integrative atlas of chicken long non-coding genes and their annotations across 25 tissues. Sci Rep. 10, 20457 (2020).
113. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 30, 923–930 (2014).
114. S. Kovaka, A. V. Zimin, G. M. Pertea, R. Razaghi, S. L. Salzberg, M. Pertea, Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biology. 20, 278 (2019).
115. C. Chen, H. Chen, Y. Zhang, H. R. Thomas, M. H. Frank, Y. He, R. Xia, TBtools: An integrative toolkit developed for interactive analyses of big biological data. Molecular Plant. 13, 1194–1202 (2020).
116. G. Yu, D. K. Smith, H. Zhu, Y. Guan, T. T.-Y. Lam, 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).
117. L. van der Maaten, G. Hinton, Visualizing Data using t-SNE. Journal of Machine Learning Research. 9, 2579–2605 (2008).
118. Y. I. Li, D. A. Knowles, J. Humphrey, A. N. Barbeira, S. P. Dickinson, H. K. Im, J. K. Pritchard, Annotation-free quantification of RNA splicing using LeafCutter. Nat Genet. 50, 151–158 (2018).
119. Z. Xia, L. A. Donehower, T. A. Cooper, J. R. Neilson, D. A. Wheeler, E. J. Wagner, W. Li, Dynamic analyses of alternative polyadenylation from RNA-seq reveal a 3′-UTR landscape across seven tumour types. Nat Commun. 5, 5274 (2014).
120. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 26, 841–842 (2010).
121. J. Li, S. Xing, G. Zhao, M. Zheng, X. Yang, J. Sun, J. Wen, R. Liu, Identification of diverse cell populations in skeletal muscles and biomarkers for intramuscular fat of chicken by single-cell RNA sequencing. BMC Genomics. 21, 752 (2020).

122. M. Mantri, G. J. Scuderi, R. Abedini-Nassab, M. F. Z. Wang, D. McKellar, H. Shi, B. Grodner, J. T. Butcher, I. De Vlaminck, Spatiotemporal single-cell RNA sequencing of developing chicken hearts identifies interplay between cellular differentiation and morphogenesis. Nat Commun. 12, 1771 (2021).
123. G. X. Y. Zheng, J. M. Terry, P. Belgrader, P. Ryvkin, Z. W. Bent, R. Wilson, S. B. Ziraldo, T. D. Wheeler, G. P. McDermott, J. Zhu, M. T. Gregory, J. Shuga, L. Montesclaros, J. G. Underwood, D. A. Masquelier, S. Y. Nishimura, M. Schnall-Levin, P. W. Wyatt, C. M. Hindson, R. Bharadwaj, A. Wong, K. D. Ness, L. W. Beppu, H. J. Deeg, C. McFarland, K. R. Loeb, W. J. Valente, N. G. Ericson, E. A. Stevens, J. P. Radich, T. S. Mikkelsen, B. J. Hindson, J. H. Bielas, Massively parallel digital transcriptional profiling of single cells. Nat Commun. 8, 14049 (2017).
124. Y. Hao, S. Hao, E. Andersen-Nissen, W. M. Mauck, S. Zheng, A. Butler, M. J. Lee, A. J. Wilk, C. Darby, M. Zager, P. Hoffman, M. Stoeckius, E. Papalexi, E. P. Mimitou, J. Jain, A. Srivastava, T. Stuart, L. M. Fleming, B. Yeung, A. J. Rogers, J. M. McElrath, C. A. Blish, R. Gottardo, P. Smibert, R. Satija, Integrated analysis of multimodal single-cell data. Cell. 184, 3573-3587.e29 (2021).
125. O. Franzén, L.-M. Gan, J. L. M. Björkegren, PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database. 2019, baz046 (2019).
126. A. M. Newman, C. B. Steen, C. L. Liu, A. J. Gentles, A. A. Chaudhuri, F. Scherer, M. S. Khodadoust, M. S. Esfahani, B. A. Luca, D. Steiner, M. Diehn, A. A. Alizadeh, Determining cell type abundance and expression from bulk tissues with digital cytometry. Nat Biotechnol. 37, 773–782 (2019).
127. A. P. Camargo, A. A. Vasconcelos, M. B. Fiamenghi, G. A. G. Pereira, M. F. Carazzolle, tspex: a tissue-specificity calculator for gene expression data. Research Square (2020), doi:10.21203/rs.3.rs-51998/v1.
128. L. Fang, W. Cai, S. Liu, O. Canela-Xandri, Y. Gao, J. Jiang, K. Rawlik, B. Li, S. G. Schroeder, B. D. Rosen, C. Li, T. S. Sonstegard, L. J. Alexander, C. P. V. Tassell, P. M. VanRaden, J. B. Cole, Y. Yu, S. Zhang, A. Tenesa, L. Ma, G. E. Liu, Comprehensive analyses of 723 transcriptomes enhance genetic and biological interpretations for complex traits in cattle. Genome Res. 30, 790-801 (2020).
129. M. E. Ritchie, B. Phipson, D. Wu, Y. Hu, C. W. Law, W. Shi, G. K. Smyth, limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Research. 43, e47 (2015).
130. Y. Benjamini, Y. Hochberg, 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).
131. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 15, 550 (2014).
132. G. Pertea, M. Pertea, GFF Utilities: GffRead and GffCompare. F1000Research (2020), , doi:10.12688/f1000research.23297.2.
133. X. Tong, S. Liu, CPPred: coding potential prediction based on the global description of RNA sequence. Nucleic Acids Research. 47, e43 (2019).
134. V. Wucher, F. Legeai, B. Hédan, G. Rizk, L. Lagoutte, T. Leeb, V. Jagannathan, E. Cadieu, A. David, H. Lohi, S. Cirera, M. Fredholm, N. Botherel, P. A. J. Leegwater, C. Le Béguec, H. Fieten, J. Johnson, J. Alföldi, C. André, K. Lindblad-Toh, C. Hitte, T. Derrien, FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. Nucleic Acids Research. 45, e57 (2017).

135. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 9, 559 (2008).
136. A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications. Neural Networks. 13, 411–430 (2000).
137. O. Stegle, L. Parts, M. Piipari, J. Winn, R. Durbin, Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc. 7, 500–507 (2012).
138. W.-M. Song, B. Zhang, Multiscale embedded gene co-expression network analysis. PLOS Computational Biology. 11, e1004574 (2015).
139. P. S. T. Russo, G. R. Ferreira, L. E. Cardozo, M. C. Bürger, R. Arias-Carrasco, S. R. Maruyama, T. D. C. Hirata, D. S. Lima, F. M. Passos, K. F. Fukutani, M. Lever, J. S. Silva, V. Maracaja-Coutinho, H. I. Nakaya, CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses. BMC Bioinformatics. 19, 56 (2018).
140. M. Leonard, S. Graham, D. Bonacum, The human factor: the critical importance of effective teamwork and communication in providing safe care. BMJ Quality & Safety. 13, i85–i90 (2004).
141. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30, 2114–2120 (2014).
142. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (2013), doi:10.48550/arXiv.1303.3997.
143. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25, 2078–2079 (2009).
144. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. GigaScience. 10, giab008 (2021).
145. B. L. Browning, Y. Zhou, S. R. Browning, A one-penny imputed genome from next-generation reference panels. The American Journal of Human Genetics. 103, 338–348 (2018).
146. S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. de Bakker, M. J. Daly, P. C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics. 81, 559–575 (2007).
147. A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, D. Reich, Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 38, 904–909 (2006).
148. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26, 139–140 (2010).
149. J. Yang, S. H. Lee, M. E. Goddard, P. M. Visscher, GCTA: A Tool for Genome-wide Complex Trait Analysis. The American Journal of Human Genetics. 88, 76–82 (2011).
150. J. Yang, B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, P. M. Visscher, Common SNPs explain a large proportion of the heritability for human height. Nat Genet. 42, 565–569 (2010).
151. H. D. PATTERSON, R. THOMPSON, Recovery of inter-block information when block sizes are unequal. Biometrika. 58, 545–554 (1971).

152. A. Taylor-Weiner, F. Aguet, N. J. Haradhvala, S. Gosai, S. Anand, J. Kim, K. Ardlie, E. M. Van Allen, G. Getz, Scaling computational genomics to millions of individuals with GPUs. Genome Biology. 20, 228 (2019).
153. H. Ongen, A. Buil, A. A. Brown, E. T. Dermitzakis, O. Delaneau, Fast and efficient QTL mapper for thousands of molecular phenotypes. Bioinformatics. 32, 1479–1485 (2016).
154. C. Giambartolomei, D. Vukcevic, E. E. Schadt, L. Franke, A. D. Hingorani, C. Wallace, V. Plagnol, Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLOS Genetics. 10, e1004383 (2014).
155. D. Duong, L. Gai, S. Snir, E. Y. Kang, B. Han, J. H. Sul, E. Eskin, 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).
156. J. S. Speagle, A conceptual introduction to Markov Chain Monte Carlo methods (2020), doi:10.48550/arXiv.1909.12313.
157. Storey JD, Bass AJ, Dabney A, Robinson D, qvalue: Q-value estimation for false discovery rate control (2022), (available at https://github.com/StoreyLab/qvalue).
158. J. R. Davis, L. Fresard, D. A. Knowles, M. Pala, C. D. Bustamante, A. Battle, S. B. Montgomery, An efficient multiple-testing adjustment for eqtl studies that accounts for linkage disequilibrium between variants. The American Journal of Human Genetics. 98, 216–224 (2016).
159. P. Mohammadi, S. E. Castel, A. A. Brown, T. Lappalainen, Quantifying the regulatory effect size of cis-acting genetic variation using allelic fold change. Genome Res. (2017), doi:10.1101/gr.216747.116.
160. S. E. Castel, F. Aguet, P. Mohammadi, F. Aguet, S. Anand, K. G. Ardlie, S. Gabriel, G. A. Getz, A. Graubert, K. Hadley, R. E. Handsaker, K. H. Huang, S. Kashin, X. Li, D. G. MacArthur, S. R. Meier, J. L. Nedzel, D. T. Nguyen, A. V. Segrè, E. Todres, F. Aguet, S. Anand, K. G. Ardlie, B. Balliu, A. N. Barbeira, A. Battle, R. Bonazzola, A. Brown, C. D. Brown, S. E. Castel, D. F. Conrad, D. J. Cotter, N. Cox, S. Das, O. M. de Goede, E. T. Dermitzakis, J. Einson, B. E. Engelhardt, E. Eskin, T. Y. Eulalio, N. M. Ferraro, E. D. Flynn, L. Fresard, E. R. Gamazon, D. Garrido-Martín, N. R. Gay, G. A. Getz, M. J. Gloudemans, A. Graubert, R. Guigó, K. Hadley, A. R. Hame, R. E. Handsaker, Y. He, P. J. Hoffman, F. Hormozdiari, L. Hou, K. H. Huang, H. K. Im, B. Jo, S. Kasela, S. Kashin, M. Kellis, S. Kim-Hellmuth, A. Kwong, T. Lappalainen, X. Li, X. Li, Y. Liang, D. G. MacArthur, S. Mangul, S. R. Meier, P. Mohammadi, S. B. Montgomery, M. Muñoz-Aguirre, D. C. Nachun, J. L. Nedzel, D. T. Nguyen, A. B. Nobel, M. Oliva, Y. S. Park, Y. Park, P. Parsana, A. S. Rao, F. Reverter, J. M. Rouhana, C. Sabatti, A. Saha, A. V. Segrè, A. D. Skol, M. Stephens, B. E. Stranger, B. J. Strober, N. A. Teran, E. Todres, A. Viñuela, G. Wang, X. Wen, F. Wright, V. Wucher, Y. Zou, P. G. Ferreira, G. Li, M. Melé, E. Yeger-Lotem, M. E. Barcus, D. Bradbury, T. Krubit, J. A. McLean, L. Qi, K. Robinson, N. V. Roche, A. M. Smith, L. Sobin, D. E. Tabor, A. Undale, J. Bridge, L. E. Brigham, B. A. Foster, B. M. Gillard, R. Hasz, M. Hunter, C. Johns, M. Johnson, E. Karasik, G. Kopen, W. F. Leinweber, A. McDonald, M. T. Moser, K. Myer, K. D. Ramsey, B. Roe, S. Shad, J. A. Thomas, G. Walters, M. Washington, J. Wheeler, S. D. Jewell, D. C. Rohrer, D. R. Valley, D. A. Davis, D. C. Mash, M. E. Barcus, P. A. Branton, L. Sobin, L. K. Barker, H. M. Gardiner, M. Mosavel, L. A. Siminoff, P. Flicek, M. Haeussler, T. Juettemann, W. J. Kent, C. M. Lee, C. C. Powell, K. R. Rosenbloom, M. Ruffier, D. Sheppard, K. Taylor, S. J. Trevanion, D. R. Zerbino, N. S. Abell, J. Akey, L. Chen, K. Demanelis, J. A. Doherty, A. P. Feinberg, K. D. Hansen, P. F. Hickey, L. Hou, F. Jasmine, L. Jiang, R. Kaul, M. Kellis, M. G. Kibriya, J. B. Li, Q. Li, S. Lin, S. E. Linder, S. B. Montgomery, M. Oliva, Y. Park, B. L.

Pierce, L. F. Rizzardi, A. D. Skol, K. S. Smith, M. Snyder, J. Stamatoyannopoulos, B. E. Stranger, H. Tang, M. Wang, P. A. Branton, L. J. Carithers, P. Guan, S. E. Koester, A. R. Little, H. M. Moore, C. R. Nierras, A. K. Rao, J. B. Vaught, S. Volpi, K. G. Ardlie, T. Lappalainen, GTEx Consortium, A vast resource of allelic expression data spanning human tissues. Genome Biology. 21, 234 (2020).

161. C. Pockrandt, M. Alzamel, C. S. Iliopoulos, K. Reinert, GenMap: ultra-fast computation of genome mappability. Bioinformatics. 36, 3687–3692 (2020).

162. X.-N. Zhu, Y.-Z. Wang, C. Li, H.-Y. Wu, R. Zhang, X.-X. Hu, Chicken chromatin accessibility atlas accelerates epigenetic annotation of birds and gene fine-mapping associated with growth traits. Zool Res. 44, 53–62 (2023).

163. K. M. Chen, E. M. Cofer, J. Zhou, O. G. Troyanskaya, Selene: a PyTorch-based deep learning library for sequence data. Nat Methods. 16, 315–318 (2019).

164. A. N. Barbeira, S. P. Dickinson, R. Bonazzola, J. Zheng, H. E. Wheeler, J. M. Torres, E. S. Torstenson, K. P. Shah, T. Garcia, T. L. Edwards, E. A. Stahl, L. M. Huckins, D. L. Nicolae, N. J. Cox, H. K. Im, Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. Nat Commun. 9, 1825 (2018).

165. A. N. Barbeira, M. Pividori, J. Zheng, H. E. Wheeler, D. L. Nicolae, H. K. Im, Integrating predicted transcriptome from multiple tissues improves association detection. PLOS Genetics. 15, e1007889 (2019).

166. Z. Zhu, F. Zhang, H. Hu, A. Bakshi, M. R. Robinson, J. E. Powell, G. W. Montgomery, M. E. Goddard, N. R. Wray, P. M. Visscher, J. Yang, Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat Genet. 48, 481–487 (2016).

167. X. Wen, R. Pique-Regi, F. Luca, Integrating molecular QTL data into genome-wide genetic association analysis: Probabilistic assessment of enrichment and colocalization. PLOS Genetics. 13, e1006646 (2017).

168. Y. Lee, F. Luca, R. Pique-Regi, X. Wen, Bayesian Multi-SNP genetic association analysis: Control of FDR and use of summary statistics. BioRxiv. 316471 (2018).

169. X. Wen, Y. Lee, F. Luca, R. Pique-Regi, Efficient integrative Multi-SNP association analysis using deterministic approximation of posteriors. Am J Hum Genet. 98, 1114-1129 (2016).

170. X. Wen, Molecular QTL discovery incorporating genomic annotations using Bayesian false discovery rate control. The Annals of Applied Statistics. 10, 1619–1638 (2016).

171. A. Hukku, M. G. Sampson, F. Luca, R. Pique-Regi, X. Wen, 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).

172. A. Bhattacharya, J. B. Hirbo, D. Zhou, W. Zhou, J. Zheng, M. Kanai, B. Pasaniuc, E. R. Gamazon, N. J. Cox, Best practices for multi-ancestry, meta-analytic transcriptome-wide association studies: Lessons from the Global Biobank Meta-analysis Initiative. Cell Genomics. 2 (2022), doi:10.1016/j.xgen.2022.100180.

Supplementary Materials

Materials and Methods

Figs. S1 to S32

Tables S1 to S22