BMC Genomics

RESEARCH Open Access

An mRNA expression atlas for the duck with public RNA-seq datasets

Qiuyu Tao[1,2], Anqi Huang[1,2], Jingjing Qi[1,2], Zhao Yang[1,2], Shihao Guo[1,2], Yinjuan Lu[1,2], Xinxin He[1,2], Xu Han[1,2], Shuaixue Jiang[1,2], Mengru Xu[1,2], Yuan Bai[1,2], Tao Zhang[1,2], Shenqiang Hu[1,2], Liang Li[1,2], Lili Bai[1,2] and HeHe Liu[1,2*]

Abstract

Background Ducks are globally important poultry species and a major source of farm animal products, including meat, eggs, and feathers. A thorough understanding of the functional genomic and transcriptomic sequences is crucial for improving production efficiency.

Result This study constructed the largest duck mRNA expression atlas among all waterfowl species to date. The atlas encompasses 1,257 tissue samples across 30 tissue types, representing all major organ systems. Using advanced clustering analysis, we established co-expression network clusters to describe the transcriptional features in the duck mRNA expression atlas and, when feasible, assign these features to unique tissue types or pathways. Additionally, we identified 27 low-variance, highly expressed housekeeping genes suitable for gene expression experiments. Furthermore, in-depth analysis revealed potential sex-biased gene expression patterns within tissues and specific gene expression profiles in meat-type and egg-type ducks, providing valuable resources to understand the genetic basis of sex differences and particular phenotypes. This research elucidates the biological processes affecting duck productivity.

Conclusion This study presents the most extensive gene expression atlas for any waterfowl species to date. These findings are of significant value for advancing duck biological research and industrial applications.

Keywords RNA-seq, Duck, Expression atlas, Co-expression network, Housekeeping genes, Differentially expressed genes

Background

Ducks are a significant global poultry species, providing essential animal products such as meat, eggs, and feathers. A thorough understanding of the functional genomic and transcriptomic sequences, and their regulation in complex traits, e.g., disease resistance, reproductive capacity, and growth potential, is crucial for improving production efficiency. The publication of the pekin duck genome in 2020 has facilitated the exploration of the genetic basis underlying various duck phenotypes [1]. Recently, population resequencing data and high-quality genomes of meat ducks, egg ducks, and wild ducks, generated through de novo assembly, have improved the

*Correspondence:
HeHe Liu
liuee1985@sicau.edu.cn
[1]State Key Laboratory of Swine and Poultry Breeding Industry, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, Sichuan 611130, P.R. China
[2]Farm Animal Genetic Resources Exploration and Innovation Key Laboratory of Sichuan Province, Sichuan Agricultural University, Chengdu, Sichuan 611130, P.R. China

sequencing quality of whole-genome assemblies in poultry [2].

In addition to genomic analyses, transcriptomic studies play a vital role in diverse areas of biological research [3]. RNA sequencing (RNA-seq) has revolutionized gene expression analysis, transitioning from single-gene studies to whole-genome analysis, enabling comprehensive transcriptome visualization, and reshaping our understanding of transcriptionally regulated complex traits [4]. Initially, large-scale gene expression mapping projects utilized microarray technology to define the transcriptomes of various species [5–9]. In recent years, studies have utilized full-length transcript sequencing, such as those of the functional annotation of animal genomes (FAANG) consortium [10, 11], and farm animal genotype-tissue expression (FarmGTEx) projects, including chicken [12, 13], cattle [14], pigs [15], and sheep [16]. RNA-seq quantification provides comprehensive transcriptome data without prior transcript knowledge, making it beneficial for non-model species without high-quality reference genomes.

Studying gene expression patterns across tissues helps elucidate organisms' evolutionary mechanisms and biological functions [17]. In particular, understanding varying gene expression across tissues is crucial to understanding the genetic foundation of duck breed formation, such as egg-laying and meat ducks. Previous transcriptomic studies have documented gene expression maps for various livestock and poultry, including chickens [9], pigs [5], cattle [18], and sheep [19]. In transcriptome-wide studies, differentially expressed genes (DEGs) are considered potential candidate genes engaged in critical functions, including growth [20, 21], meat quality [22, 23], feather follicle formation [24, 25], and disease resistance [26, 27]. Some tissue-specific gene expression pattern studies [6, 28–30] provide insights into the relationships between gene expression, tissues, and organs. However, comprehensive studies on global transcriptome profiles in ducks are limited. Previous studies on ducks have primarily focused on constructing transcriptome maps for 16 tissues in Pekin ducks and wild ducks, elucidating transcriptome changes during domestication [31]. Currently, there is no publicly available large-scale gene expression map for duck tissues. Comprehensive studies on the gene expression landscape of different duck tissues are lacking, highlighting the urgent need for large-scale gene expression profiling to conduct comparative analyses of tissue-specific expression patterns in ducks.

This study presents the most extensive gene expression atlas for any waterfowl species to date, including RNA-seq libraries from embryonic, mid, and late developmental stages, covering all major organ systems. We aim to provide a model transcriptome for ducks and gain an understanding of the molecular foundation of genes and tissue functions. We investigated sex-biased gene expression in tissues, specific gene expression in meat and egg-laying ducks, and gene expression affecting early embryonic development in ducks. Additionally, identifying housekeeping genes of experimental value was a key focus of this study. Finally, we employed advanced clustering techniques to establish networks of co-expressed genes and identify tissue-specific expression genes. These data will provide deeper insights into the biological processes underlying complex traits that affect duck productivity.

Result

Scope of the Duck gene expression atlas dataset

This study includes transcriptome data from 914 duck tissue samples (Supplementary Table S1) downloaded from the public database (NCBI database and GSA) and 343 unpublished duck tissue samples (Supplementary Table S2). These samples cover 16 different breeds and include samples from embryonic, mid, and late-growth stages. To maximize transcriptome diversity, the atlas includes data from 30 different tissues, selected to provide a comprehensive overview of organ systems related to phenotypic traits such as growth, metabolism, and reproduction. Detailed information can be found in Supplementary Table S3. After quality filtering, the number of reads per sample ranged from 12.7 million to 189 million, with mapping rates between 81.3% and 98.9%, reflecting variability among duck breeds. The dataset was generated using various Illumina sequencing platforms (including HiSeq 2000/2500/4000/6000/× TEN, and NovaSeq 6000), and all samples were paired-end sequenced. We aligned the transcripts to the reference genome (ZJU1.0) (https://api.ncbi.nlm.nih.gov/datasets/v2/genome/accession/GCF_015476345.1/), resulting in the identification of 23,512 genes. For each tissue, gene expression levels were quantified as Transcripts Per Million (TPM) using StringTie [32]. A complete table summarizing gene expression levels across all tissues (measured in TPM) is provided as Supplementary Table S4. This table includes data from both public datasets and additional samples generated in this study. Figure 1A shows the distribution of gene TPM values across different tissues, revealing that 37.92–43.03% of genes exhibit very low expression ($0 < TPM \leq 0.1$), 18.33–20.52% show low expression ($0.1 < TPM \leq 1$), 22.12–26.05% have moderate expression ($1 < TPM \leq 10$), 30.10–35.28% have relatively high expression ($10 < TPM \leq 100$), and 1.21–5.56% display high expression ($TPM > 100$).

To explore global transcriptome patterns and tissue similarities, we analyzed the pairwise correlation heatmap using Spearman correlation across the 30 tissue types. The heatmap (Fig. 1B) reveals that skeletal muscle and various brain regions exhibit the most distinct global

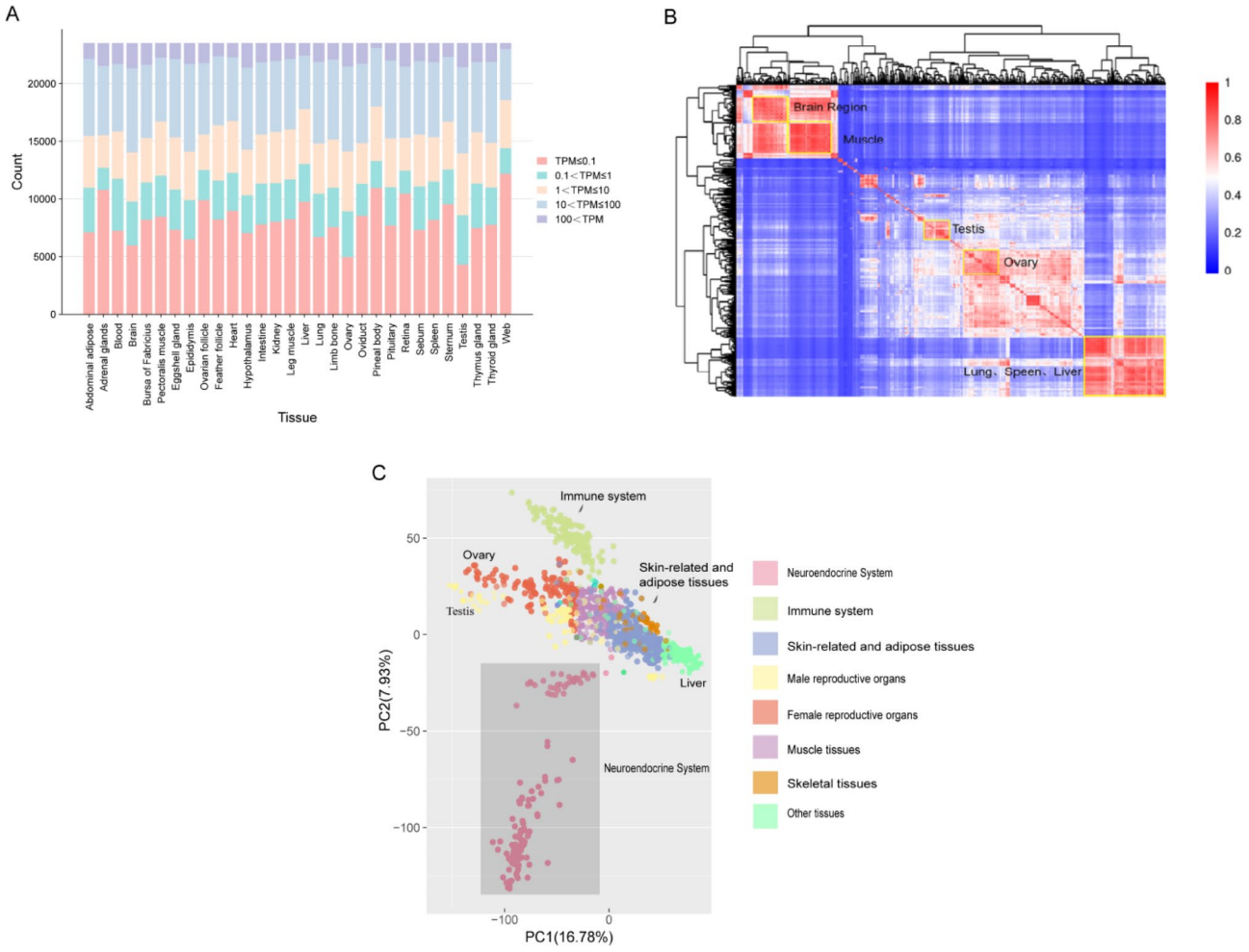

Fig. 1 Gene expression profile among 30 tissue types. (**A**) Statistical distribution chart of gene TPM values across different tissue samples. (**B**) Unbiased hierarchical clustering heat map based on Pearsons's correlation coefficient for all genes. Color intensity indicates the correlation between tissues, red indicates a high correlation (1), and blue indicates a low correlation (0.5). (**C**) The principal component analysis (PCA) plot shows the relation and clustering of all duck tissue samples. (CNS = central nervous system)

expression patterns, similar to findings in human, pig, and chicken tissues [33–35]. Tissues that cluster closely typically share a common embryonic origin, function, or tissue type. For example, neuroectoderm-derived tissues such as the brain, hypothalamus, pituitary gland, pineal gland, and retina cluster together, with the thyroid gland also showing a high correlation, likely due to its role in hormone regulation. Similarly, reproductive organs (testis, epididymis, ovaries, and oviducts) and immune-related organs (liver, spleen, and lungs) exhibit high correlations within their respective clusters. We also conducted a principal component analysis (PCA) of 1,257 samples, which confirmed that tissues with related functions share similar global expression patterns. Brain region samples, in particular, display unique expression patterns, clustering together distinctly (Fig. 1C). These findings demonstrate the robustness of aggregating data from multiple tissues to create a comprehensive and informative gene expression atlas.

Network cluster analysis

The duck tissue gene expression atlas, comprising 1,257 samples, was analyzed using Graphia Enterprise, a tool for visualizing and analyzing network graphs from large datasets [36, 37]. We filtered out low and stably expressed genes to focus on variable, tissue-specific genes. A pairwise Pearson correlation matrix was generated for gene expression across samples, applying a correlation threshold of $r = 0.75$ and an MCL (Markov Cluster Algorithm) inflation value of 2.0 to form co-expression clusters. The resulting network contained 10,800 nodes (genes) and 1.3 million edges (correlations). Clusters were numbered by size, with cluster 1 being the largest, comprising 1,241 genes (Supplementary Table S5). The network graph (Fig. 2A) features one large component with 8,546

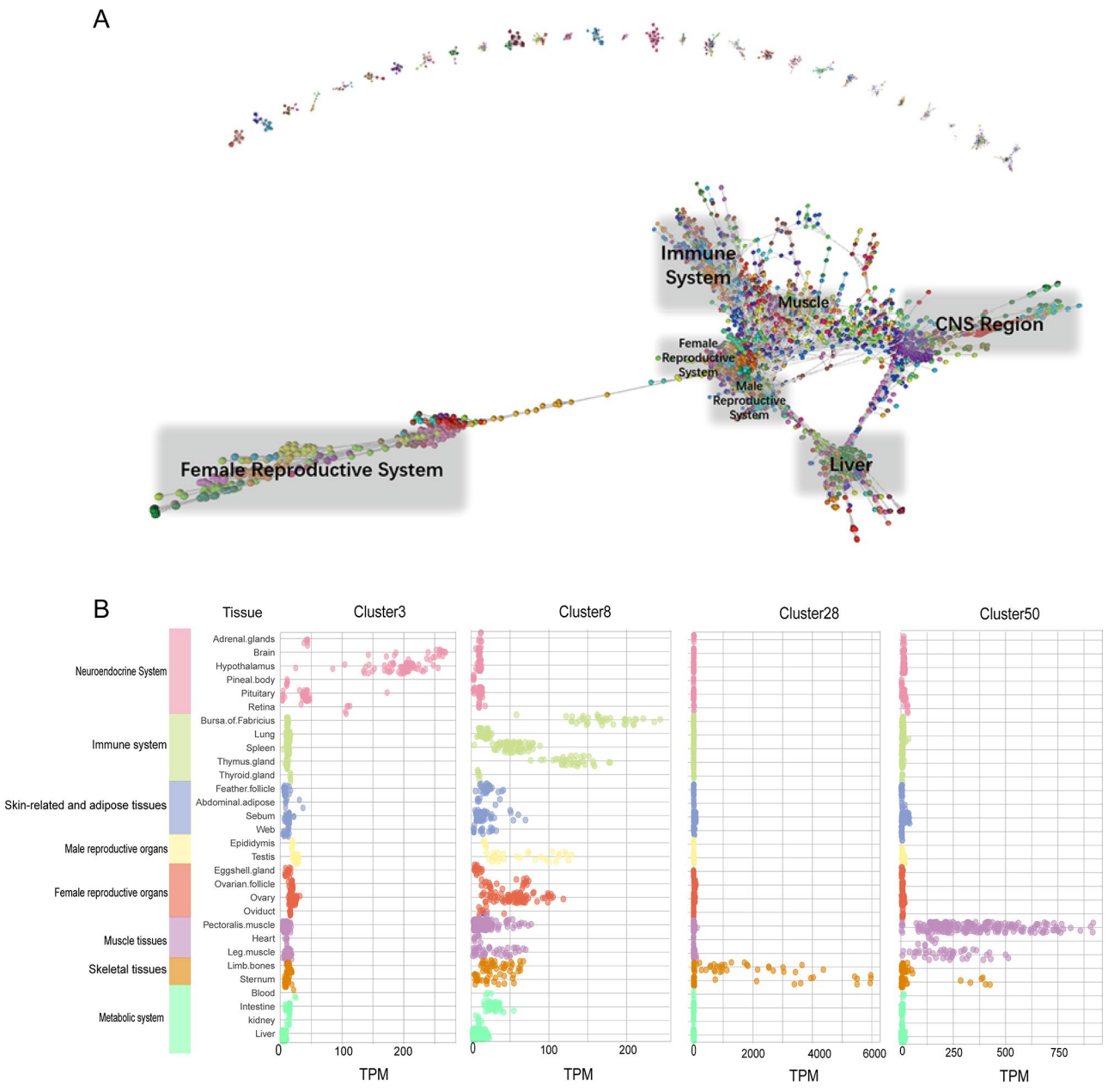

Fig. 2 Interrogation of the underlying expression profiles allows regions of the graph ($r = 0.75$, MCL = 2.0) to be associated with specific tissues or tissue types. (**A**) A three-dimensional visualization of a Pearson correlation gene-to-gene graph of expression levels derived from RNA-Seq data from analysis of all duck tissues. (**B**) It shows scatter plots of gene expression profiles in the selected cluster (Cluster 3, 8, 28, and 50)

nodes and 38 smaller components. Clusters 1 to 50, ranked by size, were manually annotated (Supplementary Table S6). Most co-expression clusters consist of genes expressed predominantly in specific tissues, such as the liver (cluster 1), ovaries (cluster 2), testis (cluster 4), and retina (cluster 6). However, some clusters include genes co-expressed across multiple organ systems, reflecting shared biological functions (Fig. 2B). For example, the third co-expression cluster, comprising 770 nodes, is highly expressed in brain and hypothalamus tissues. Additionally, we identified many genes highly expressed in the bursa of Fabricius, ovaries, and testis, as well as genes expressed explicitly in the sternum and limb bones. Scatter plots of gene expression for all mentioned expression clusters can be found in Supplementary Figure S1.

Strict co-expression clustering requires each transcript to be quantified across multiple conditions to correlate strongly with other co-expressed transcripts, indicating shared functions or pathways. This method is particularly effective for analyzing region-specific gene expression within tissues, as demonstrated in gene expression atlases of pigs [38], sheep [19], and chickens [35]. Some clusters

exhibit strong immune characteristics. For example, clusters 11, 16, 18, 22, and 23 contain genes associated with immune responses and are specifically expressed in the thymus, bursa of Fabricius, lungs, and spleen. For instance, genes in clusters 11, 16, and 22 are specifically expressed in the thymus, with significantly enriched GO terms including immune system process (GO:0002376), T-cell differentiation (GO:0030217), and lymphocyte differentiation (GO:0030098), among other immune-related molecular functions (Fig. 3A; Supplementary Table S7). Some highly expressed unannotated genes in these clusters might also play important roles in immunity, such as *LOC101802873*, *LOC101802866*, *LOC101798744*, and *LOC101789908*. Additionally, genes in clusters 8 and 34 are specifically expressed in the bursa of Fabricius and reproductive organs (ovaries and testis). GO term analysis of these genes indicates they are involved in processes such as the mitotic cell cycle process (GO:1903047), chromosome organization (GO:0051276), mitotic cell cycle (GO:0000278), and protein-DNA complex assembly (GO:0065004) (Fig. 3B; Supplementary Table S8). This result suggests that the ovaries and testis, as reproductive organs, undergo rapid cell division and proliferation to produce germ cells (ova and sperm). Similarly, the bursa of Fabricius, as a primary immune organ, is involved in the extensive generation and maturation of B cells during early development in birds, which also requires high levels of cell division and proliferation [39]. Furthermore, this may imply that these tissues may have evolutionarily conserved regulatory mechanisms for certain biological functions, particularly in cell proliferation and the maintenance of genetic information.

The duck gene expression atlas includes libraries from early developmental stages. Cluster 28 is the largest co-expression cluster associated with developmental skeletal tissue. The KEGG pathways significantly enriched in this cluster include ECM-receptor interaction ($P = 0.0002$), glycosaminoglycan biosynthesis ($P = 0.0004$), and TGF-β signaling pathway ($P = 0.0025$), all of which are related to skeletal development (Supplementary Table S9). Genes highly expressed during embryonic stages compared to growth stages within cluster 28 include *CHST11*, *IL17D*,

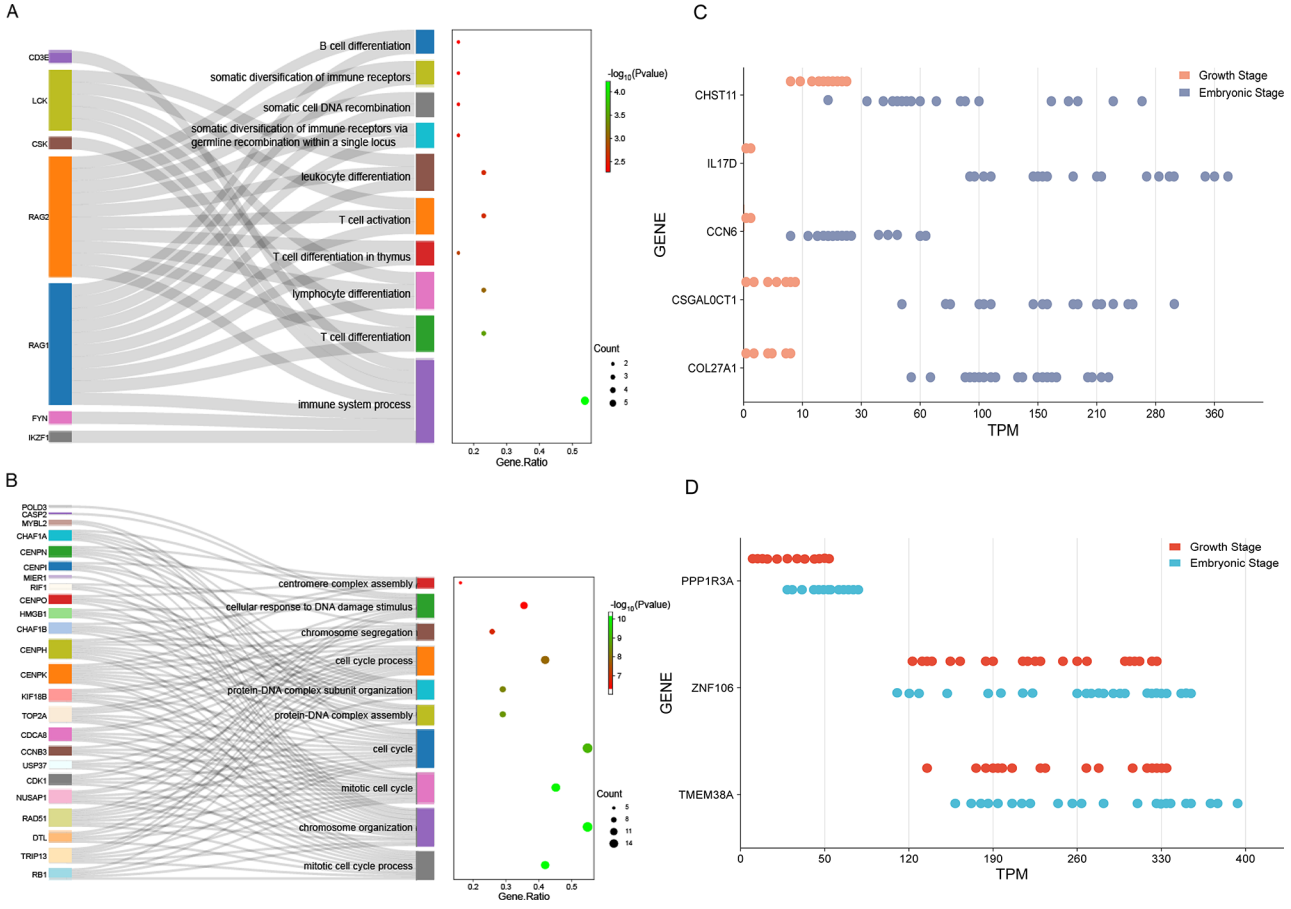

Fig. 3 Functional annotation and details of selected cluster genes. GO term analysis of (**A**) cluster 11, 16, 22 and (**B**) cluster 8, 34 (Sankey diagram: The nodes on the left represent individual genes; The nodes on the right represent enriched biological pathways; The width of the links between genes and processes indicates the contribution of the gene to a particular biological pathways). Genes enriched in (**C**) cluster 28 and (**D**) cluster 50 are more expressed in the embryonic stage than in the growth stage

CCN6, *CSGALNACT1*, and *COL27A1* (Fig. 3C). In skeletal muscle development, clusters 27 and 50 are significantly associated with pathways like calcium signaling ($P = 0.0055$) and myocardial contraction ($P = 0.0710$), which are essential for muscle development and protein synthesis (Supplementary Table S10). Similarly, cluster 50 also contains genes highly expressed during the embryonic stage compared to the growth stage, playing a crucial role in embryonic development. These genes include *PPP1R3A*, *ZNF106*, and *TMEM38A* (Fig. 3D).

Housekeeping genes

Housekeeping genes (HKGs) are defined as genes with conserved expression across all tissues, essential for maintaining fundamental biological processes and cellular functions (e.g., cellular transport) [40]. We identified 1,088 HKGs with an average TPM value greater than 100 across all tissues (Supplementary Table S11). By calculating the coefficient of variation (CV) for these genes, we determined that the first quartile was 0.58, and the third quartile was 1.07. Based on these CV thresholds, HKGs were classified into low-variation ($n = 272$), medium-variation ($n = 544$), and high-variation ($n = 272$) categories (Fig. 4A). To investigate the roles of HKGs in tissue development, we performed GO functional enrichment analysis on these categories (Fig. 4B; Supplementary Table S12). Low-variation HKGs are primarily involved in translation initiation, protein synthesis, and cellular metabolism, while medium-variation HKGs are associated with intracellular transport, localization and distribution. High-variation HKGs are mainly linked to metabolic processes and cellular structure organization, consistent with findings in cattle studies [18]. We visualized the expression of five common HKGs (*GAPDH*, *ACTB*, *HPRT1*, *GUSB*, *B2M*) and the five HKGs with the lowest CVs (*EEF2*, *RPL24*, *RPL7*, *RPL21*, *RPL26L1*) across tissues (Fig. 4C), finding that common HKGs tend to have relatively high CVs. For example, *GAPDH* exhibits significantly higher expression in skeletal muscle than in other tissues. Among the low-variation HKGs, we identified 27 genes with TPM > 1000 (13% of low-variation HKGs) (Fig. 4A; Supplementary Table S13). Remarkably, 85% of these genes belong to the *RP* (Ribosomal Proteins, RPs) gene family, and 7.5% belong to the *EEF* (Eukaryotic Elongation Factors, EEFs) gene family, making them valuable candidate reference genes for gene expression experiments (Fig. 4D).

Sex-biased gene expression in ducks

Sex-biased genes are classified into female-biased genes and male-biased genes. Genes with FDR < 0.05 and log2FC > 1 are defined as female-biased genes, whereas those with FDR < 0.05 and log2FC < -1 are defined as male-biased genes. To detect sex-biased gene expression across 8 tissues (Pectoralis muscle (13 males vs. 13 females), thymus gland (29 males vs. 6 females), hypothalamus (10 males vs. 10 females), pituitary (6 males vs. 6 females), liver (13 males vs. 13 females), bursa of Fabricius (28 males vs. 6 females), sternum (13 males vs. 5 females), and gonad (20 males vs. 34 females)) in ducks, DESeq2 was used with a global false discovery rate (FDR) of 5%. The differential expression of each tissue is shown in Supplementary Figure S2. In the sex-biased genes identified across these tissues, the gonad, pituitary, and hypothalamus exhibited the most pronounced sex differences, with 11,074, 6,295, and 5,250 differentially expressed genes, respectively (Fig. 5A). Notably, in the pituitary and hypothalamus, the number of male-biased genes ($n = 5,299$ and 4,137) significantly exceeded that of female-biased genes ($n = 995$ and 1,124). Moreover, the results indicated that most sex-biased genes across these tissues were located on autosomes (Fig. 5B). KEGG pathway enrichment analysis of genes with at least 5-fold sex-biased expression in one or more tissues revealed that female-biased genes were primarily enriched in metabolic pathways, such as lipid metabolic processes (sphingolipid metabolism, PPAR signaling pathway, terpenoid backbone biosynthesis) and amino acid metabolism (phenylalanine metabolism, tyrosine metabolism, cysteine, and methionine metabolism). In contrast, male-biased genes were mainly enriched in pathways related to tissue development, including skeletal system development, muscle tissue development, and connective tissue development (Supplementary Table S14; Supplementary Figure S3).

Ducks have no global chromosomal dosage compensation (males ZZ, females ZW) [41]. It has been observed that Z-linked genes show higher expression in males than in females, indicating a dosage effect [42–44]. However, in birds, this gene dosage effect on sex chromosomes is typically incomplete [45, 46]. In this study, we analyze the F: M ratio (sex-biased expression of autosomal and Z-linked genes) and the Z(Z): AA ratio (relative expression of Z-linked genes to autosomal genes). To verify incomplete dosage compensation in ducks, we compared the $\log_2$ F: M (female: male) ratios of autosomal genes with those on the Z chromosome. In these eight tissues, the $\log_2$ F: M ratio for autosomal genes was close to 0. In contrast, the $\log_2$ F: M ratios for Z-linked genes in five tissues showed significant male bias. For example, compared to females, the expression levels in males were 1.69-fold higher in the pituitary gland, 2.1-fold higher in the hypothalamus, 1.8-fold higher in the sternum, 1.2-fold higher in the thymus, and 1.28 -fold higher in the bursa of Fabricius. Additionally, we found no significant dosage effect of Z-linked gene expression in the pectoralis muscle, liver, and gonad (Fig. 5C). This finding suggests that Z-linked genes in the pituitary, hypothalamus,

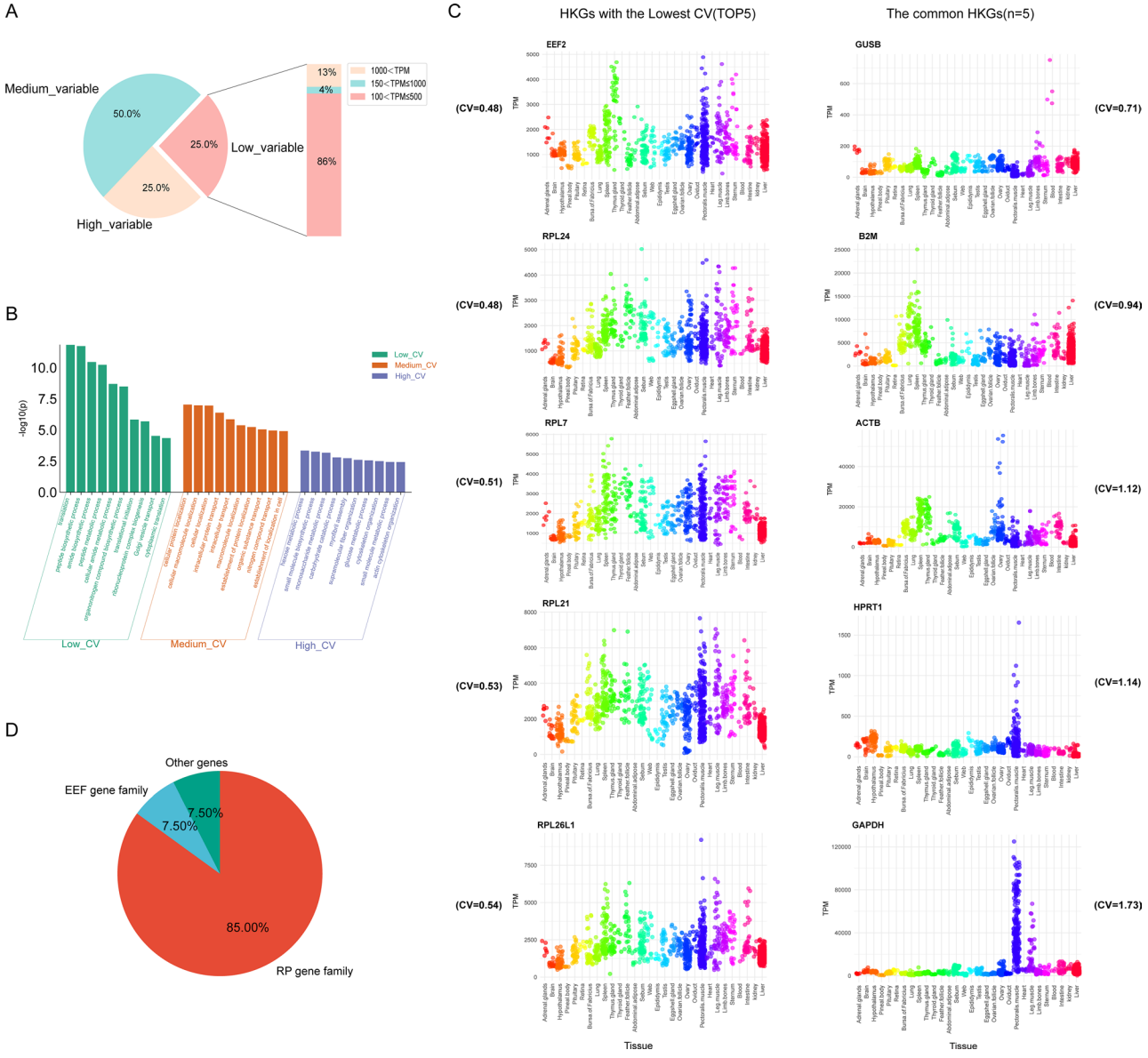

Fig. 4 The expression pattern and hierarchical clustering of 1088 HKGs across 30 duck tissues. (**A**) The HKGs are variably expressed, and only 25% are constantly expressed. Among those constant HK genes, only 13% are highly expressed with TPM larger than 100. (**B**) GO Enrichment Analysis of Housekeeping Genes with High, medium, and low CV (Top 10). (**C**) Expression of the Top 5 HKGs with the Lowest CV and 5 Common HKGs of Variation Across Various Tissues (CV = Coefficients of Variation). (**D**) Proportion of genes in 27 HKGs

sternum, thymus gland, and bursa of Fabricius are not well balanced and compensated between males and females, supporting a dosage effect between the sexes.

Furthermore, the Z (Z): AA ratio analysis showed that the expression of Z-linked genes was significantly higher than that of autosomal genes in the male pituitary ($P = 4.6e-12$, Wilcoxon rank-sum test), male hypothalamus ($P = 2.22e-16$), male bursa of Fabricius ($P = 5.1e-16$), male liver ($P = 6.8e-08$), and male gonads ($P = 9.2e-10$). In contrast, our results confirmed that the expression differences between Z-linked and autosomal genes in these tissues were relatively lower in females (Fig. 5D). Combining these results, we identified tissues with an F: M ratio less than 0 and significant differences in the Z(Z): AA ratio as the pituitary, hypothalamus, and bursa of Fabricius. Intersecting the male-biased genes in these three tissues, we identified 11 Z-chromosome dosage effect genes: *KIAA1958*, *LOC119714541*, *LOC101791319*, *REM1*, *LOC119714512*, *LOC113840422*, *LOC101793244*, *CER1*, *LOC113840306*, *LOC113840423*, and *LOC113840287* (Fig. 5E).

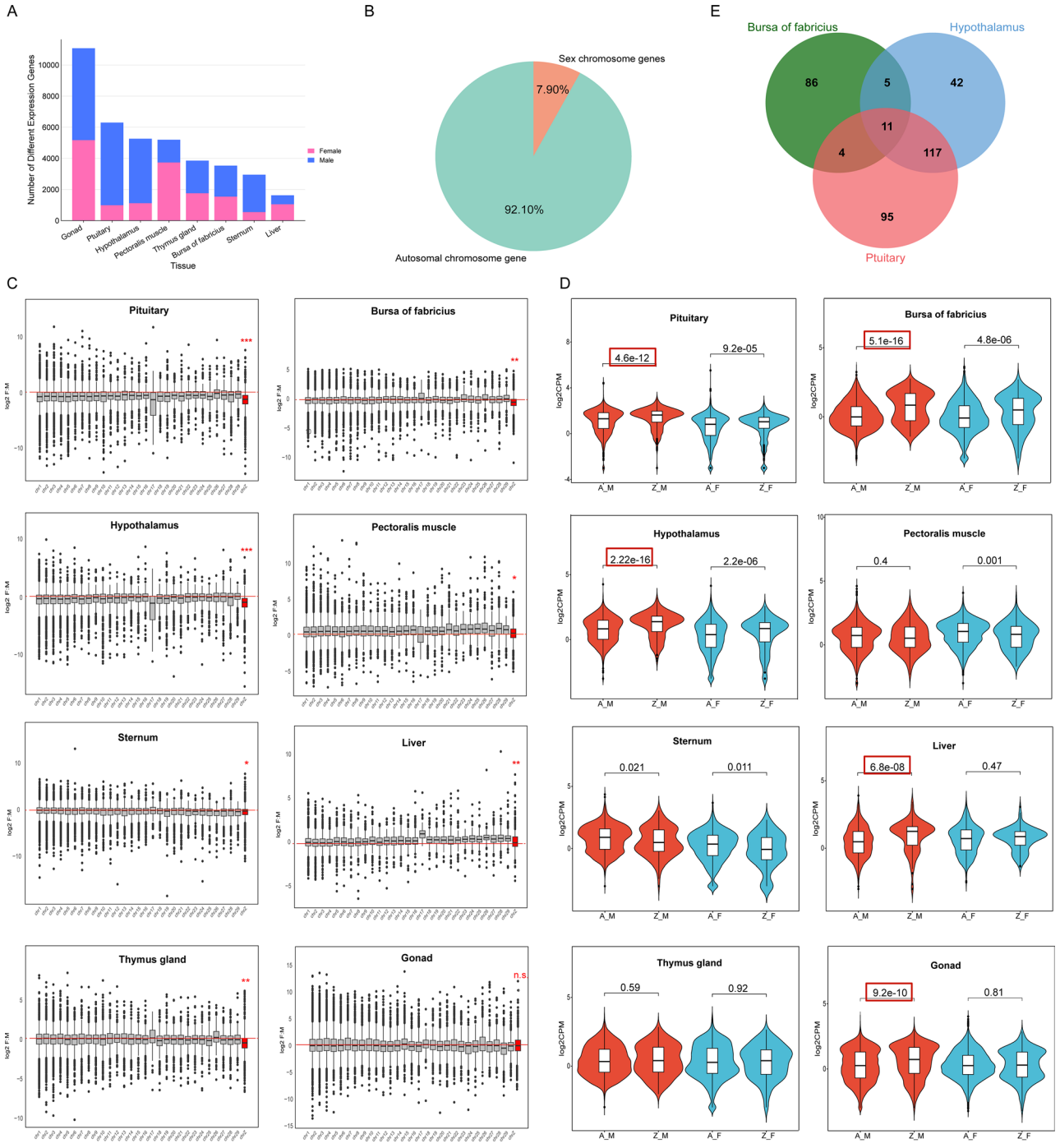

Fig. 5 Expression patterns of sex-biased genes in ducks. (**A**) The distribution of sex-specific genes in the eight tissues (female-biased genes in red; Blue for the male-bias genes) (**B**) The location of the distribution ratio of differentially expressed genes (green as autosomal gene number proportion; Orange is the ratio of sex chromosome genes) (**C**) Fold-change map (***indicates $P \leq 0.001$; **indicates $0.001 < P \leq 0.01$; *indicates $0.01 < P \leq 0.05$; n.s. indicates $P > 0.05$) and (**D**) zither map of the duck Z chromosome for pituitary, hypothalamus, sternum, thymus gland, bursa of Fabricius, pectoralis muscle, liver and gonads. (**E**) Venn diagram of Z chromosome of dose-effect genes in the bursa of Fabricius, pituitary, and hypothalamus

Differential gene expression between meat ducks and laying ducks

Domestic ducks originated from the mallard (Anas platyrhynchos) over 2,000 years ago [47–49]. Over the past century, domestic ducks have been selectively bred into egg-laying, meat, and dual-purpose types to meet the high demand for eggs and meat. Domestication and subsequent selective breeding have rapidly evolved phenotypes and production traits [50]. Phenotypic differences between populations can result from genetic

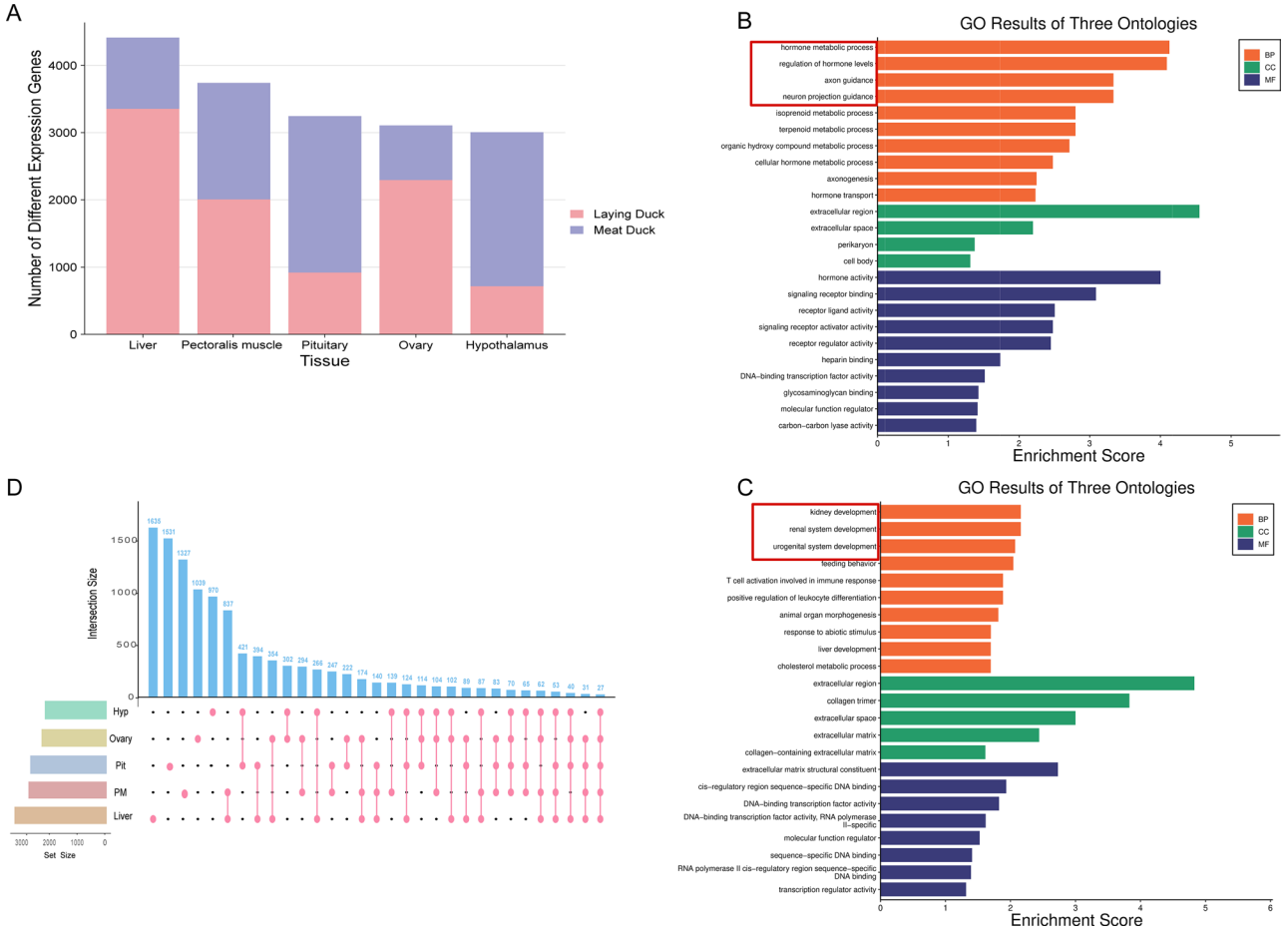

Fig. 6 Differentially expressed genes between egg-type and meat-type ducks. (**A**) The number of differentially expressed genes in the five tissues (red indicates genes highly expressed in laying ducks, blue indicates genes highly expressed in meat ducks). KEGG enrichment plot of highly expressed genes (**B**) in egg-type ducks and (**C**) in meat-type ducks. (**D**) Venn diagram shows the shared and unique differentially expressed genes among the hypothalamus, ovary, pituitary, pectoralis muscle, and liver tissues between meat and laying ducks. (Hyp = hypothalamus, Pit = pituitary, PM = pectoralis muscle)

changes affecting the function and expression of gene products [51]. However, the genetic basis underlying the significant differences between laying and meat ducks remains unclear. Therefore, in this study, we compared gene expression across tissues between laying and meat ducks to explore the genetic changes underlying these differences.

The distribution of differentially expressed genes (DEGs) ($|\log_2 FC| > 1$) between laying ducks (Jinding, Shaoxing, and Gaoyou) and meat ducks (Peking and Cherry Valley) across various tissues (liver (13 laying ducks vs. 13 meat ducks), pectoralis muscle (3 laying ducks vs. 3 meat ducks), ovary (9 laying ducks vs. 9 meat ducks), pituitary gland (12 laying ducks vs. 12 meat ducks), hypothalamus (10 laying ducks vs. 12 meat ducks)) is shown in Fig. 6A. Excluding the liver ($n = 4358$), the pectoralis muscle ($n = 3741$) and pituitary ($n = 3246$) exhibited the highest number of DEGs. Furthermore, functional enrichment analysis revealed that the DEGs with higher expression in laying ducks compared to meat ducks ($\log_2 FC > 4$; $n = 992$) were enriched in biological processes related to hormone metabolic processes (GO:0042445), regulation of hormone levels (GO:0010817) among others (Fig. 6B). This indicates an upregulation of gene expression in hormone regulation, neurodevelopment, and metabolic processes, possibly related to laying ducks' reproductive and hormonal regulation needs. In contrast, the DEGs with higher expression in meat ducks compared to laying ducks (log2FC < -4; $n = 630$) were enriched in biological processes such as kidney development (GO:0001822), T cell activation (GO:0002286), and liver development (GO:0001889) among others (Fig. 6C). This indicates an upregulation of gene expression involved in organ development, immune response, and gene regulation, which may be related to meat ducks' growth, immune, and metabolic needs. To better understand tissue-specific DEGs, we performed a comparative analysis across the five tissues (Fig. 6D). The results showed that 27 common DEGs are shared among these five tissues.

Discussion

A comprehensive tissue transcript survey reveals the biological roles and regulatory mechanisms of genetic variants underlying complex traits [52]. Gene expression analysis has shown that transcriptome variations contribute to phenotypic diversity among species [53]. However, ducks have lacked such comprehensive studies. This study generated the first extensive duck transcriptome atlas, encompassing 1,257 samples from 30 tissue types across 10 organ systems (e.g., neuroendocrine, immune). By investigating and comparing transcriptome profiles across these tissues, we aim to enhance our understanding of how gene expression variability influences phenotypic diversity.

This study integrates data from various sources. To minimize potential biases introduced by different data origins, we applied a unified quality control and analysis workflow to all RNA-seq data. Specifically, low-quality reads were filtered using the FASTX Toolkit with consistent parameters, and all reads were aligned to the same reference genome (ZJU1.0) using Hisat2 with standardized settings. Gene expression levels for all samples were normalized to TPM values using StringTie. However, this study has some limitations. While TPM normalization reduces the impact of sequencing depth, it may not completely eliminate biases and could affect low-expression genes [54]. Nonetheless, this has minimal impact on our study, as the focus is primarily on highly expressed genes.

Analyzing tissue-specific gene expression patterns is essential for elucidating the relationships between gene expression, tissues, and organs. We employed Graphia [36] for network clustering analysis on the gene expression atlas in duck tissues. Genes highly expressed in specific tissues and exhibiting consistent expression patterns formed distinct clusters. Detailed examination of these clusters can provide important biological insights. For instance, Cluster 11 and 16, which are highly expressed in the thymus, are related to the innate immune response. The unannotated gene *LOC101802873*, highly expressed in the thymus (cluster 16), may encode a T-cell receptor-associated transmembrane adaptor protein that regulates T-cell signaling. Another example is Cluster 23, where functional annotation revealed that the genes within the cluster also play roles in immune response processes. The significantly enriched KEGG pathways include Influenza A ($P = 1.68 \times 10^{-5}$), RIG-I-like receptor signaling pathway ($P = 0.0025$), and Herpes simplex infection (HSV-1) ($P = 8.11 \times 10^{-3}$), all of which are related to viral infection pathways. Manual annotation of unannotated genes within this cluster identified *PARP12* (*LOC101802866*), *PARP14* (*LOC101798744*), and *LOC101789908*. Related studies have shown that in wild ducks infected with the influenza A virus, the expression levels of *PARP12* (*LOC101802866* and *LOC101796889*) and *PARP14* (*LOC101789908*) are upregulated in the lungs and ileum [55]. Moreover, the duck gene expression atlas dataset includes multiple libraries from early developmental stages. For example, the development skeletal co-expression cluster is Cluster 28, which is related to skeletal development. We observed that genes enriched in Cluster 28 with higher expression during embryonic stages compared to growth stages include *CHST11*, *IL17D*, *CCN6*, *CSGALNACT1*, and *COL27A1*. Studies have shown that these genes regulate embryonic cartilage development in animals, and their absence can lead to cartilage dysplasia or even embryonic lethality, indicating their crucial roles in the growth and development of embryonic skeletal structures [56–63]. Regarding the development of skeletal muscles, Cluster 50 is of interest, which is related to muscle contraction development and protein synthesis. Genes enriched in this cluster with higher expression during embryonic stages than growth stages include *PPP1R3A*, *ZNF106*, and *TMEM38A*. *TMEM38A* is a member of the transmembrane protein family, and related studies have shown it plays an important role in the growth and development of bovine embryos [64].

Housekeeping genes (HKGs) are broadly expressed across tissues, essential for basic cellular functions, and serve as reliable experimental controls [65]. This study identified 1,088 HKGs, including 27 with high expression levels suitable for use as experimental controls. The ribosomal protein gene family, responsible for the large ribosomal subunit and protein synthesis within cells, constituted the largest proportion [66–68]. Related studies also suggest that the *RP* gene family can be used as reference genes [69]. Common HKGs (*GAPDH*, *ACTB*, *HPRT1*, *GUSB*, *B2M*) were expressed across all duck tissues at varying levels, consistent with previous reports [70, 71]. We also obtained a list of HKGs that maintain basic cell functions and energy metabolism (e.g., translational initiation, protein synthesis, and cellular substance metabolism).

Sex-specific differences in gene expression have been reported in human [72], mice [73], cattle [74], pigs [75] and chicken [35]. While some studies have examined these differences in ducks, the genetic basis remains incomplete. Therefore, this study compares gene expression between male and female ducks in eight tissues: pectoralis muscle, liver, pituitary gland, hypothalamus, gonads, thymus gland, and bursa of Fabricius. KEGG pathway enrichment analysis revealed that female-biased genes are enriched in metabolic pathways, such as lipid and amino acid metabolism, while male-biased genes are enriched in pathways related to tissue development, including skeletal, muscle, and connective tissue development. This is consistent with studies on sex-biased genes in sheep and chickens [19, 35], which may reflect the intrinsic differences between sexes in allocating resources

for growth or reproduction. Additionally, we focused on some highly expressed genes identified as potential regulators for establishing sexual identity (Supplementary Table S15). *LHX9*, a *LIM* homeobox gene family member, plays a crucial role in animal embryonic development. In chickens, the *LHX9* motif is enriched only in female-biased DARs. *POMC* encodes pro-opiomelanocortin, and studies in chickens have indicated that *POMC* methylation patterns and gene expression are associated with sex differences [76]. Some sex-differentially expressed genes identified in this study have been analyzed in previous research, but we discovered many novel sex-differentially expressed genes, providing clues for further analysis of sex differences in ducks. Furthermore, ducks (ZZ males, ZW females) lack global chromosome dosage compensation, with studies showing higher expression levels of male Z-linked genes than females, indicating a dosage effect [41–44]. By utilizing the F: M ratio and the Z (Z): AA ratio for further analysis, we identified three tissues—pituitary gland, hypothalamus, and bursa of Fabricius—that showed significant differences with F: M < 0 and Z (Z): AA ratios. By intersecting the male-biased genes in these tissues, we identified 11 Z chromosome dosage effect genes.

The genetic basis for the differences between laying and meat ducks remains unclear. By comparing differentially expressed genes in the liver, ovary, breast muscle, hypothalamus, and pituitary tissues of laying and meat ducks, we found that laying duck-biased genes are primarily involved in hormone metabolism and regulation processes, while meat duck-biased genes are associated with tissue growth processes. By intersecting the differentially expressed genes in the five tissues, 27 genes were differentially expressed in all five tissues. Among them, the expression levels of the *FGA* and *FGB* genes in the ovary, breast muscle, hypothalamus, and pituitary were biased towards egg ducks (Supplementary Table S16). *FGA* and *FGB* are critical during blastocyst implantation and embryo transfer following in vitro fertilization [77, 78]. Studies have confirmed the association between genetic variants in *FGA* (rs6050) and *FGB* (rs1800790) and pregnancy outcomes [79]. Therefore, the *FGA* and *FGB* genes may be related to the reproductive capacity of ducks. Additionally, the expression levels of the *SOCS3* gene in the breast muscle, liver, hypothalamus, and pituitary were biased toward meat ducks. *SOCS3* is a SOCS (Suppressor of Cytokine Signaling) protein family member. Studies have shown that *SOCS3* overexpression enhances the mRNA expression of genes related to muscle maturation and hypertrophy [80]. Similarly, during the embryonic development of Pekin ducks, *SOCS3* has been identified to be associated with muscle development and fat deposition [81]. These results suggest that the *SOCS3* gene may be crucial in developing skeletal muscle in meat ducks.

Conclusion

This study assembled the most extensive duck transcriptome collection to date, representing core tissues across all major organ systems. By applying network clustering and correlation analysis, we validated known gene expression-function relationships and identified candidate genes involved in immune response, embryonic skeletal development, and muscle growth. We also provided new insights into sex-biased gene expression and identified genes associated with phenotypic differences between laying and meat ducks. Finally, we identified 27 low-variation, highly expressed housekeeping genes, with the *RP* gene family being the most represented. This transcriptome map is a valuable resource for understanding duck biology and establishes a robust foundation for using ducks as animal models in industrial and research applications. However, this study has some limitations, such as the potential bias introduced by differences in sample sources, and the biological functions of the candidate genes require further validation. Future research can supplement the current findings through experimental validation and broader species coverage.

Materials and methods

RNA-seq dataset collection

The publicly accessible datasets utilized in this study are listed in Supplementary Table S1. These datasets, accessible through the Sequence Read Archive (SRA) and Genome Sequence Archive (GSA), provide the raw data for our analysis. The SRA, available via the National Center for Biotechnology Information (NCBI), stores the data in binary format, which can be converted to FASTQ files using the fast-dump tool from the SRA Toolkit (https://trace.ncbi.nlm.nih.gov/Traces/sra/?view=software). Additionally, this study includes 343 samples collected by the Waterfowl Research Group of Sichuan Agricultural University, which have not yet been deposited in public databases. These samples represent six duck breeds: Nonghua Ma duck, Nonghua White duck, Huaifu Meat duck, Cherry Valley duck, Sichuan Ma duck, and Jian Chang duck, with ages ranging from 56 to 300 days. Tissue samples include the hypothalamus, pineal gland, retina, bursa of Fabricius, thymus, liver, ovary, oviduct, follicle, eggshell gland, testis, sternum, pectoral muscle, and web.

Raw data processing

Low-quality reads were filtered using the FASTX Toolkit (v0.0.13). Specifically, the fastq_quality_filter tool was employed to remove reads where more than 20% of bases had a quality score below 20. After filtering, the

remaining high-quality reads were aligned to the reference genome ZJU1.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_015476345.1) using Hisat2 (v2.1.0). To enhance alignment accuracy, the parameters --known-splicesite-infile and --pen-noncansplice were used to prioritize the alignment of known splice sites [82]. The aligned data were processed with Samtools. StringTie calculates the expression level of each transcript based on the RNA-Seq alignment results and quantitatively outputs the transcript in TPM format [32]. The merged transcripts were compared with the reference annotation ZJU1.0 (https://www.ncbi.nlm.nih.gov/assembly/GCF_015476345.1) using the gffcompare tool (v0.10.1, https://ccb.jhu.edu/software/stringtie/gffcompare.shtml) [83]. Gene expression levels were quantified in transcripts per million (TPM) for most analyses. For differential expression analysis (DEG), raw read counts were used. In addition, log-transformed TPM values were applied for Principal Component Analysis (PCA) and sample-to-sample correlation analyses to ensure normal-like data distribution and improve interpretability.

Network clustering analysis and functional annotation

Duck tissue gene expression profiles were clustered using Graphia [36]. The Pearson correlation matrix retained only relationships with $r \geq 0.75$, forming a network graph where nodes (genes) were connected by edges (correlations above the threshold). The Markov Clustering Algorithm (MCL) with an inflation value of 2.0 was employed to interpret the network's structure. Genes exhibiting strong co-expression patterns were assumed to share related functions [19]. Gene Ontology (GO) functional analysis of the genes within each co-expression cluster was performed using the online tool KOBAS 3.0 (http://www.bioinfo).

Identification of housekeeping genes

Related studies showed housekeeping genes (HKGs) exhibit constitutive expression across all or most tissues [40]. To assess the expression variation of HKGs in the duck expression profile, we calculated the coefficient of variation (CV) for each gene, a method commonly used in similar studies [18, 70]. CV is defined as the ratio of the standard deviation to the mean ($CV = \sigma/\mu$), where σ represents the standard deviation and μ the mean expression level across tissues. We classified HKGs into low ($CV \leq 0.58$), medium ($0.58 < CV \leq 1.07$), and high ($1.07 < CV$) variation expression categories, and performed GO enrichment analysis for each category. Among the low variation HKGs, those with an average expression of $TPM > 1000$ across all duck tissues were identified as candidate reference HKGs.

Differential expression analysis

The DESeq2 package in R (v3.5.1) was used to convert read counts to counts per million (CPM) and to identify differentially expressed genes (DEGs) using the Wald test. Significance thresholds were set at $FDR < 0.05$ and $|\log_2 FC| > 1$.

Data visualization

Data visualization was conducted in R (v4.2.3) with RStudio, primarily using ggplot2 (v3.3.6) package, along with pheatmap (v1.0.12), ggpubr (v0.4.0), and tidygraph (v1.2.0). Figures were assembled and annotated in Adobe Illustrator (Adobe Inc., San Jose, USA).

Statistical analysis

Data analysis was performed using SPSS software (version 20.0, SPSS Inc., Chicago, IL) and Microsoft Excel (Microsoft Corporation, Redmond, WA, USA) for organizing, processing, and analyzing the dataset. ANOVA analysis among different groups was done using SPSS software. Differences at $P < 0.05$ were considered significant. Graphs were created using GraphPad Prism (version 8.0.2) and R (version 4.2.3).

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11385-4.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7
Supplementary Material 8
Supplementary Material 9
Supplementary Material 10
Supplementary Material 11
Supplementary Material 12
Supplementary Material 13
Supplementary Material 14
Supplementary Material 15
Supplementary Material 16
Supplementary Material 17
Supplementary Material 18
Supplementary Material 19

Acknowledgements
This study utilized data from public databases. We extend our gratitude to the authors and institutions who contributed the corresponding data to these public databases.

Author contributions
QT and HL are responsible for constructing article ideas and drafting manuscripts. In addition, QT, JQ, ZY, SG, XH, XH, SJ, MX, YB, TZ, YL conducted data statistics and analysis. AH, SH, HL, LB, LL participated in the writing instruction and revision of the manuscript. All listed authors have made substantial contributions to the research and publication. Thank you to all the authors for their contributions to the study.

Funding
This research was supported by grants from National Key R&D Program of China (2022YFF1000100, 2023YFD1300300, 2024YFF1000900) and China Agriculture Research System of Waterfowl (CARS-42).

Data availability
The reference genome data was available in NCBI's database (ZJU1.0, https://www.ncbi.nlm.nih.gov/assembly/GCF_015476345.1). The raw genome sequencing data is available in NCBI's SRA database (https://www.ncbi.nlm.nih.gov/sra/; Accession number: PRJNA573985, PRJNA605347, PRJNA640233, PRJNA642918, PRJNA645648, PRJNA609437, PRJNA665025, PRJNA665331, PRJNA665334, PRJNA665337, PRJNA561952, PRJNA704544, PRJNA721551, PRJNA743556, PRJNA759406, PRJNA767080, PRJNA263681, PRJNA785593, PRJNA791517, PRJNA273367, PRJNA828272, PRJNA859501, PRJNA863028, PRJNA878639, PRJNA886426, PRJNA891486, PRJCA004157, PRJNA946269, PRJNA273367, PRJNA349262, PRJNA357037, PRJNA358835, PRJNA397953, PRJNA412507, PRJNA419583, PRJNA437203, PRJNA437527, PRJNA459507, PRJNA476836, PRJNA484659, PRJNA489334, PRJNA194464, PRJNA496401, PRJNA509092, PRJNA526427 and PRJNA530027) and in Genome Sequence Archive (GSA) (https://ngdc.cncb.ac.cn/gsa/; Accession number:PRJCA003535, PRJCA004157, PRJCA023388, PRJCA006648, PRJCA007509 and PRJCA010062).

Declarations

Ethics approval and consent to participate
All methods were carried out following relevant guidelines and regulations. All duck work was conducted following a protocol approved by Sichuan Agricultural University China's animal ethics and welfare committee (AEWC).

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 28 September 2024 / Accepted: 19 February 2025
Published online: 18 March 2025

References

1. Li J. Analysis of the evolution of sex chromosomes in Pekin Duck and monotremes. Zhejiang University; 2020.
2. Zhu F, Yin Z-T, Wang Z, Smith J, Zhang F, Martin F, Ogeh D, Hincke M, Lin F-B, Burt DW. Three chromosome-level Duck genome assemblies provide insights into genomic variation during domestication. Nat Commun. 2021;12(1):5932.
3. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, Austine-Orimoloye O, Azov AG, Barnes I, Bennett R. Ensembl 2022. Nucleic Acids Res. 2022;50(D1):D988–95.
4. Wickramasinghe S, Cánovas A, Rincón G, Medrano JF. RNA-sequencing: a tool to explore new frontiers in animal genetics. Livest Sci. 2014;166:206–16.
5. Freeman TC, Ivens A, Baillie JK, Beraldi D, Barnett MW, Dorward D, Downing A, Fairbairn L, Kapetanovic R, Raza S, et al. A gene expression atlas of the domestic pig. BMC Biol. 2012;10:90.
6. Harhay GP, Smith TP, Alexander LJ, Haudenschild CD, Keele JW, Matukumalli LK, Schroeder SG, Van Tassell CP, Gresham CR, Bridges SM. An atlas of bovine gene expression reveals novel distinctive tissue characteristics and evidence for improving genome annotation. Genome Biol. 2010;11:1–18.
7. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, Wiltshire T, Orth AP, Vega RG, Sapinoso LM, Moqrich A. Large-scale analysis of the human and mouse transcriptomes. *Proceedings of the National Academy of Sciences* 2002, 99(7):4465–4470.
8. Su AI, Wiltshire T, Batalov S, Lapp H, Ching KA, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G. A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences* 2004, 101(16):6062–6067.
9. Bush SJ, Freem L, MacCallum AJ, O'Dell J, Wu C, Afrasiabi C, Psifidi A, Stevens MP, Smith J, Summers KM. Combination of novel and public RNA-seq datasets to generate an mRNA expression atlas for the domestic chicken. BMC Genomics. 2018;19:1–19.
10. Clark EL, Archibald AL, Daetwyler HD, Groenen MA, Harrison PW, Houston RD, Kühn C, Lien S, Macqueen DJ, Reecy JM. From FAANG to fork: application of highly annotated genomes to improve farmed animal production. Genome Biol. 2020;21:1–9.
11. Kern C, Wang Y, Xu X, Pan Z, Halstead M, Chanthavixay G, Saelao P, Waters S, Xiang R, Chamberlain A. Functional annotations of three domestic animal genomes provide vital resources for comparative and agricultural research. Nat Commun. 2021;12(1):1821.
12. Guan D, Bai Z, Zhu X, Zhong C, Hou Y, Consortium C, Lan F, Diao S, Yao Y, Zhao B. The ChickenGTEx pilot analysis: a reference of regulatory variants across 28 chicken tissues. *bioRxiv* 2023:2023.2006. 2027.546670.
13. Fang L, Guan D, Bai Z, Zhu X, Zhong C, Hou Y, Li H, Lan F, Diao S, Yao Y. The ChickenGTEx atlas: the genetic regulation of multi-tissue and single-cell transcriptome signatures in chickens. 2023.
14. Liu S, Gao Y, Canela-Xandri O, Wang S, Yu Y, Cai W, Li B, Xiang R, Chamberlain AJ, Pairo-Castineira E, et al. A multi-tissue atlas of regulatory variants in cattle. Nat Genet. 2022;54(9):1438–47.
15. Teng J, Gao Y, Yin H, Bai Z, Liu S, Zeng H, Consortium P, Bai L, Cai Z, Zhao B. A compendium of genetic regulatory effects across pig tissues. Nat Genet. 2024;56(1):112–23.
16. Zhao B, Luo H, He J, Huang X, Chen S, Fu X, Zeng W, Tian Y, Liu S, Li C-, et al. Comprehensive transcriptome and methylome analysis delineates the biological basis of hair follicle development and wool-related traits in Merino sheep. BMC Biol. 2021;19(1):197.
17. Cannavò E, Koelling N, Harnett D, Garfield D, Casale FP, Ciglar L, Gustafson HE, Viales RR, Marco-Ferreres R, Degner JF, et al. Genetic variants regulating expression levels and isoform diversity during embryogenesis. Nature. 2017;541(7637):402–6.
18. Zhang T, Wang T, Niu Q, Xu L, Chen Y, Gao X, Gao H, Zhang L, Liu GE, Li J. Transcriptional atlas analysis from multiple tissues reveals the expression specificity patterns in beef cattle. BMC Biol. 2022;20(1):79.
19. Clark EL, Bush SJ, McCulloch ME, Farquhar IL, Young R, Lefevre L, Pridans C, Tsang HG, Wu C, Afrasiabi C. A high resolution atlas of gene expression in the domestic sheep (Ovis aries). PLoS Genet. 2017;13(9):e1006997.
20. Wang X, Xiao Y, Yang H, Lu L, Liu X, Lyu W. Transcriptome analysis reveals the genes involved in growth and metabolism in muscovy ducks. Biomed Res Int. 2021;2021:6648435.
21. Wang Z, Liang W, Yan D, Tian H, Dong B, Zhao W, Chang G, Chen G. Identification of genes related to growth traits from transcriptome profiles of Duck breast muscle tissue. Animal Biotechnol. 2023;34(4):1239–46.
22. Zhou H, Yang Y, Wang L, Ye S, Liu J, Gong P, Qian Y, Zeng H, Chen X. Integrated multi-omic data reveal the potential molecular mechanisms of the nutrition and flavor in Liancheng white Duck meat. Front Genet. 2022;13:939585.
23. Ge K, Chen X, Kuang J, Yang L, Geng Z. Comparison of liver transcriptome from high-and low-intramuscular fat Chaohu ducks provided additional candidate genes for lipid selection. 3 Biotech. 2019;9:1–11.
24. Yang J, Qu Y, Huang Y, Lei F. Dynamic transcriptome profiling towards Understanding the morphogenesis and development of diverse feather in domestic Duck. BMC Genomics. 2018;19:1–16.
25. Chen X, Ge K, Wang M, Zhang C, Geng Z. Integrative analysis of the Pekin Duck (Anas anas) MicroRNAome during feather follicle development. BMC Dev Biol. 2017;17:1–12.
26. Cao J, Zhang Y, Chen Y, Liang S, Liu D, Fan W, Xu Y, Liu H, Zhou Z, Liu X. Dynamic transcriptome reveals the mechanism of liver injury caused by DHAV-3 infection in Pekin Duck. Front Immunol. 2020;11:568565.
27. Liu T, Cheng A, Wang M, Jia R, Yang Q, Wu Y, Sun K, Zhu D, Chen S, Liu M. RNA-seq comparative analysis of Peking Ducks spleen gene expression 24 h post-infected with Duck plague virulent or attenuated virus. Vet Res. 2017;48:1–14.
28. Kern C, Wang Y, Chitwood J, Korf I, Delany M, Cheng H, Medrano JF, Van Eenennaam AL, Ernst C, Ross P. Genome-wide identification of tissue-specific

long non-coding RNA in three farm animal species. BMC Genomics. 2018;19:1–14.
29. Chen M, Yao YL, Yang Y, Zhu M, Tang Y, Liu S, Li K, Tang Z. Comprehensive profiles of mRNAs and MiRNAs reveal molecular characteristics of multiple organ physiologies and development in pigs. Front Genet. 2019;10:756.
30. Chamberlain AJ, Vander Jagt CJ, Hayes BJ, Khansefid M, Marett LC, Millen CA, Nguyen TT, Goddard ME. Extensive variation between tissues in allele specific expression in an outbred mammal. BMC Genomics. 2015;16:1–20.
31. Yin Z-T, Li X-Q, Sun Y-X, Smith J, Hincke M, Yang N, Hou Z-C. Selection on the promoter regions plays an important role in complex traits during Duck domestication. BMC Biol. 2023;21(1):303.
32. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT, stringtie and ballgown. Nat Protoc. 2016;11(9):1650–67.
33. Karlsson M, Sjöstedt E, Oksvold P, Sivertsson Å, Huang J, Álvez MB, Arif M, Li X, Lin L, Yu J. Genome-wide annotation of protein-coding genes in pig. BMC Biol. 2022;20(1):25.
34. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, Sivertsson Å, Kampf C, Sjöstedt E, Asplund A. Tissue-based map of the human proteome. Science. 2015;347(6220):1260419.
35. Zhang J, Wang X, Lv C, Wan Y, Zhang X, Li J, Wang Y. A gene expression atlas of Lohmann white chickens. *bioRxiv* 2022:2022.2007. 2030.500160.
36. Freeman TC, Horsewell S, Patir A, Harling-Lee J, Regan T, Shih BB, Prendergast J, Hume DA, Angus T. Graphia: A platform for the graph-based visualisation and analysis of complex data. *bioRxiv* 2020:2020.2009. 2002.279349.
37. Dimonaco NJ, Salavati M, Shih BB. Computational analysis of SARS-CoV-2 and SARS-like coronavirus diversity in human, Bat and Pangolin populations. Viruses. 2020;13(1):49.
38. Freeman TC, Ivens A, Baillie JK, Beraldi D, Barnett MW, Dorward D, Downing A, Fairbairn L, Kapetanovic R, Raza S. A gene expression atlas of the domestic pig. BMC Biol. 2012;10(1):1–22.
39. Ribatti D, Crivellato E, Vacca A. The contribution of Bruce Glick to the definition of the role played by the bursa of Fabricius in the development of the B cell lineage. Clin Experimental Immunol. 2006;145(1):1–4.
40. Zhu J, He F, Hu S, Yu J. On the nature of human housekeeping genes. Trends Genet. 2008;24(10):481–4.
41. Wolf JB, Bryk J. General lack of global dosage compensation in ZZ/ZW systems? Broadening the perspective with RNA-seq. BMC Genomics. 2011;12:1–10.
42. Itoh Y, Melamed E, Yang X, Kampf K, Wang S, Yehya N, Van Nas A, Replogle K, Band MR, Clayton DF. Dosage compensation is less effective in birds than in mammals. J Biol. 2007;6:1–15.
43. Kaiser VB, Ellegren H. Nonrandom distribution of genes with sex-biased expression in the chicken genome. Evolution. 2006;60(9):1945–51.
44. Storchová R, Divina P. Nonrandom representation of sex-biased genes on chicken Z chromosome. J Mol Evol. 2006;63:676–81.
45. Mank J, Ellegren H. All dosage compensation is local: gene-by-gene regulation of sex-biased expression on the chicken Z chromosome. Heredity. 2009;102(3):312–20.
46. Melamed E, Arnold AP. Regional differences in dosage compensation on the chicken Z chromosome. Genome Biol. 2007;8:1–10.
47. Qu L, Liu W, Yang F, Hou Z, Zheng J, Xu G, Yang N. Origin and domestication history of Peking ducks deltermined through microsatellite and mitochondrial marker analysis. Sci China Ser C: Life Sci. 2009;52(11):1030–5.
48. Li H-F, Zhu W-Q, Song W-T, Shu J-T, Han W, Chen K-W. Origin and genetic diversity of Chinese domestic ducks. Mol Phylogenet Evol. 2010;57(2):634–40.
49. Zhang Z, Jia Y, Chen Y, Wang L, Lv X, Yang F, He Y, Ning Z, Qu L. Genomic variation in Pekin Duck populations developed in three different countries as revealed by whole-genome data. Anim Genet. 2018;49(2):132–6.
50. Zhou Z, Li M, Cheng H, Fan W, Yuan Z, Gao Q, Xu Y, Guo Z, Zhang Y, Hu J. An intercross population study reveals genes associated with body size and plumage color in ducks. Nat Commun. 2018;9(1):2648.
51. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ. Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res. 2010;20(6):816–25.
52. Finucane HK, Reshef YA, Anttila V, Slowikowski K, Gusev A, Byrnes A, Gazal S, Loh P-R, Lareau C, Shoresh N, et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. Nat Genet. 2018;50(4):621–9.
53. Brawand D, Soumillon M, Necsulea A, Julien P, Csárdi G, Harrigan P, Weier M, Liechti A, Aximu-Petri A, Kircher M. The evolution of gene expression levels in mammalian organs. Nature. 2011;478(7369):343–8.
54. Gao Y, Wang Y, Zhang L, Dou T, Zhou Y. The impact of transcriptome sequencing depth on gene detection and expression level Estimation. J Yunnan Agricultural University: Nat Sci Ed. 2014;29(5):6.
55. Campbell LK, Magor KE. Pattern recognition receptor signaling and innate responses to influenza A viruses in the Mallard Duck, compared to humans and chickens. Front Cell Infect Microbiol. 2020;10:209.
56. Christiansen HE, Lang MR, Pace JM, Parichy DM. Critical early roles for col27a1a and col27a1b in zebrafish notochord morphogenesis, vertebral mineralization and post-embryonic axial growth. PLoS ONE. 2009;4(12):e8481.
57. Watanabe Y, Takeuchi K, Higa Onaga S, Sato M, Tsujita M, Abe M, Natsume R, Li M, Furuichi T, Saeki M. Chondroitin sulfate N-acetylgalactosaminyltransferase-1 is required for normal cartilage development. Biochem J. 2010;432(1):47–55.
58. Nakamura Y, Weidinger G, Liang JO, Aquilina-Beck A, Tamai K, Moon RT, Warman ML. The CCN family member Wisp3, mutant in progressive pseudorheumatoid dysplasia, modulates BMP and Wnt signaling. J Clin Investig. 2007;117(10):3075–86.
59. Hamazaki N, Uesaka M, Nakashima K, Agata K, Imamura T. Gene activation-associated long noncoding RNAs function in mouse preimplantation development. Development. 2015;142(5):910–20.
60. Klüppel M, Wight TN, Chan C, Hinek A, Wrana JL. Maintenance of chondroitin sulfation balance by chondroitin-4-sulfotransferase 1 is required for chondrocyte development and growth factor signaling during cartilage morphogenesis. 2005.
61. Shabbir RMK, Nalbant G, Ahmad N, Malik S, Tolun A. Homozygous CHST11 mutation in chondrodysplasia, brachydactyly, overriding digits, clino-symphalangism and synpolydactyly. J Med Genet. 2018;55(7):489–96.
62. Mikami T, Kitagawa H. Biosynthesis and function of chondroitin sulfate. Biochim Et Biophys Acta (BBA)-General Subj. 2013;1830(10):4719–33.
63. Izumikawa T, Okuura Y, Koike T, Sakoda N, Kitagawa H. Chondroitin 4-O-sulfotransferase-1 regulates the chain length of chondroitin sulfate in co-operation with chondroitin N-acetylgalactosaminyltransferase-2. Biochem J. 2011;434(2):321–31.
64. Srirattana K. St. John JC: manipulating the mitochondrial genome to enhance cattle embryo development. G3 Genes|Genomes|Genetics. 2017;7(7):2065–80.
65. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. Trends Genet. 2013;29(10):569–74.
66. Odintsova TI, Müller EC, Ivanov AV, Egorov TA, Bienert R, Vladimirov SN, Kostka S, Otto A, Wittmann-Liebold B, Karpova GG. Characterization and analysis of posttranslational modifications of the human large cytoplasmic ribosomal subunit proteins by mass spectrometry and Edman sequencing. J Protein Chem. 2003;22(3):249–58.
67. Anger AM, Armache JP, Berninghausen O, Habeck M, Subklewe M, Wilson DN, Beckmann R. Structures of the human and Drosophila 80S ribosome. Nature. 2013;497(7447):80–5.
68. Liang X, Zuo MQ, Zhang Y, Li N, Ma C, Dong MQ, Gao N. Structural snapshots of human pre-60S ribosomal particles before and after nuclear export. Nat Commun. 2020;11(1):3542.
69. Zhang Y, Zhao L, Zeng Y. Selection and application of reference genes in gene expression studies. J Plant Physiol. 2014;50(08):1119–25.
70. de Jonge HJ, Fehrmann RS, de Bont ES, Hofstra RM, Gerbens F, Kamps WA, de Vries EG, Zee AG, Te Meerman GJ, Ter Elst A: evidence based selection of housekeeping genes. PLoS ONE. 2007;2(9):e898.
71. Zeng J, Liu S, Zhao Y, Tan X, Aljohi HA, Liu W, Hu S. Identification and analysis of house-keeping and tissue-specific genes based on RNA-seq data sets across 15 mouse tissues. Gene. 2016;576(1):560–70.
72. Imahara SD, Jelacic S, Junker CE, O'Keefe GE. The influence of gender on human innate immunity. Surgery. 2005;138(2):275–82.
73. Everhardt Queen A, Moerdyk-Schauwecker M, McKee LM, Leamy LJ, Huet YM. Differential expression of inflammatory cytokines and stress genes in male and female mice in response to a lipopolysaccharide challenge. PLoS ONE. 2016;11(4):e0152289.
74. Forde N, Maillo V, O'Gaora P, Simintiras CA, Sturmey RG, Ealy AD, Spencer TE, Gutierrez-Adan A, Rizos D, Lonergan P. Sexually dimorphic gene expression in bovine conceptuses at the initiation of implantation. Biol Reprod. 2016;95(4):92.
75. Zhang J, Zhou C, Ma J, Chen L, Jiang A, Zhu L, Shuai S, Wang J, Li M, Li X. Breed, sex and anatomical location-specific gene expression profiling of the Porcine skeletal muscles. BMC Genet. 2013;14:1–9.
76. Rancourt RC, Schellong K, Tzschentke B, Henrich W, Plagemann A. DNA methylation and expression of Proopiomelanocortin (POMC) gene in the

hypothalamus of three-week-old chickens show sex-specific differences. FEBS Open Bio. 2018;8(6):932–9.
77. Wang H, Pilla F, Anderson S, Martínez-Escribano S, Herrer I, Moreno-Moya JM, Musti S, Bocca S, Oehninger S, Horcajadas JA. A novel model of human implantation: 3D endometrium-like culture system to study attachment of human trophoblast (Jar) cell spheroids. Mol Hum Reprod. 2012;18(1):33–43.
78. Heymann D, Vidal L, Or Y, Shoham Z. Hyaluronic acid in embryo transfer media for assisted reproductive technologies. Cochrane Database Syst Reviews 2020(9).
79. Schwedler C, Heymann G, Bukreeva L, Hoppe B. Association of genetic polymorphisms of fibrinogen, factor XIII A-Subunit and α2-Antiplasmin with fibrinogen levels in pregnant women. Life. 2021;11(12):1340.
80. Caldow MK, Steinberg GR, Cameron-Smith D. Impact of SOCS3 over-expression on human skeletal muscle development in vitro. Cytokine. 2011;55(1):104–9.
81. Xu T-S, Gu L-H, Huang W, Xia W-L, Zhang Y-S, Zhang Y-G, Rong G, Schachtschneider K, Hou S-S. Gene expression profiling in Pekin Duck embryonic breast muscle. PLoS ONE. 2017;12(5):e0174612.
82. Kim JM, Kim J, Kim YH, Kim KT, Ryu SH, Lee TG, Suh PG. Comparative secretome analysis of human bone marrow-derived mesenchymal stem cells during osteogenesis. J Cell Physiol. 2013;228(1):216–24.
83. Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. *F1000Research* 2020, 9.

Publisher's note