SUPPLEMENTARY INFORMATION

Egg-laying ChickenGTEx resource deciphers context-specific regulatory effects on fertility traits

Di Zhu[1,2†], Kai Shi[3†], The ChickenGTEx Consortium, Chong Li[1], Yidan Yan[1], Houcheng Li[2], Zhonghao Bai[2], Lizhi Tan[1], Dailu Guan[4], Yiqiang Zhao[1], Yuzhe Wang[1], Baoliang Fan[5], Ziqin Jiang[6], Zhenqiang Xu[6], Chungang Feng[3*], Lingzhao Fang[2*] & Xiaoxiang Hu[1*]

[1]State Key Laboratory of Animal Biotech Breeding, National Research Facility for Phenotypic and Genotypic Analysis of Model Animals, College of Biological Sciences, China Agricultural University, Beijing, 100193, China.

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

[3]College of Animal Science and Technology, Nanjing Agricultural University, Nanjing, 210095, China.

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

[5]Animal Science and Technology Department, Hebei Agricultural University, Baoding, 071000, China.

[6]Wen's Nanfang Poultry Breeding Co. Ltd., Yunfu, 527400, China.

[†] These authors contributed equally to this work.

Corresponding authors:

Chungang Feng, E-mail: fengchungang@njau.edu.cn

Lingzhao Fang, E-mail: lingzhao.fang@qgg.au.dk

Xiaoxiang Hu, E-mail: huxx@cau.edu.cn

The supplementary information contains:

Supplementary Figures

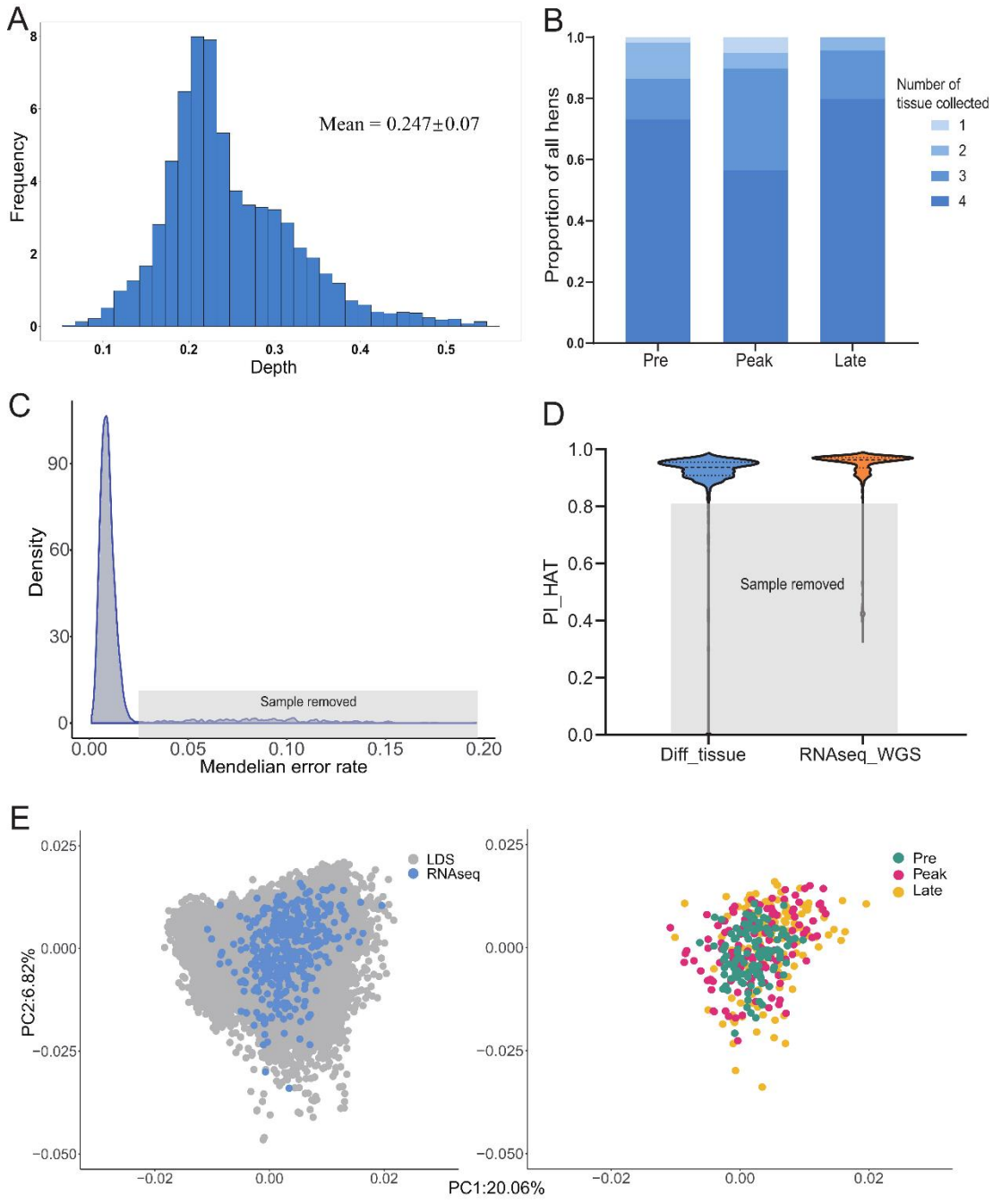

Figure S1 Compilation of chicken egg-laying chickenGTEx datasets.

(A) Distribution of sequencing depth across low-depth sequencing samples. The mean and standard deviation are indicated in the plot. (B) Stacked bar plot showing the proportion of RNA-seq samples with varying numbers of tissues collected across the three laying stages. (C) Comparison with pedigree records: Distribution of Mendelian error rates between low-depth sequencing sample pairs. The gray area indicates sample pairs with erroneous labels, which were excluded from further analysis. (D) "PI_HAT" values between RNA-seq samples from different tissues of the same hen and between corresponding RNA-seq and WGS samples were calculated using the Plink --genome function. The

gray area indicates sample pairs with erroneous labels, which were excluded from subsequent analyses. A total of 358 WGS samples and 1,272 RNA-seq samples were included in this analysis. (E) Principal component analysis (PCA) comparing low-depth sequencing and laying GTEx samples (left), and different stages within the laying GTEx samples (right).

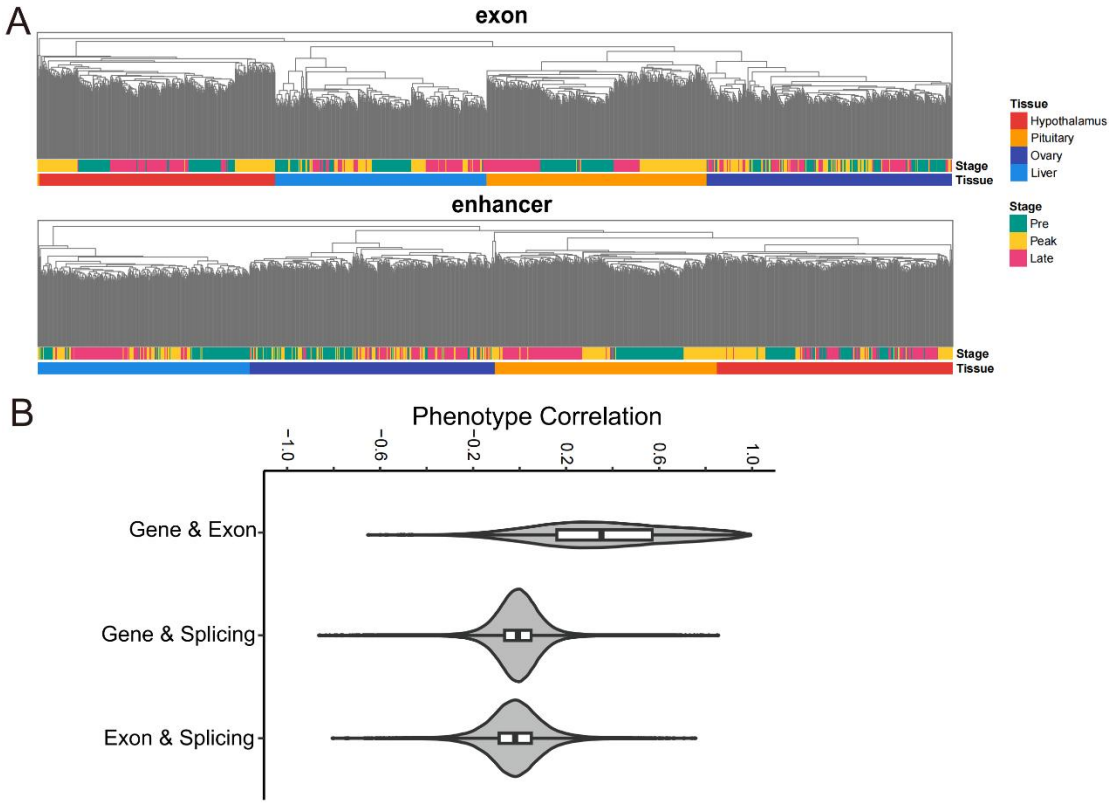

Figure S2 Global patterns and relationships among molecular phenotypes across tissues. (A) Hierarchical clustering was performed using the expression levels (quantified as transcripts per million, TPM) of the 4,000 exons and enhancers showing the highest variance across all samples.This analysis illustrates the overall similarity and tissue-specific clustering patterns among the 1,272 RNA-seq datasets. (B) Violin and box plots showing the distribution of Pearson correlation coefficients among different molecular phenotypes within the same gene. The center line of each box represents the median, and the box bounds represent the interquartile range. In total, 25,016 genes, 78,530 exons, and 130,911 splicing events were included in this analysis.

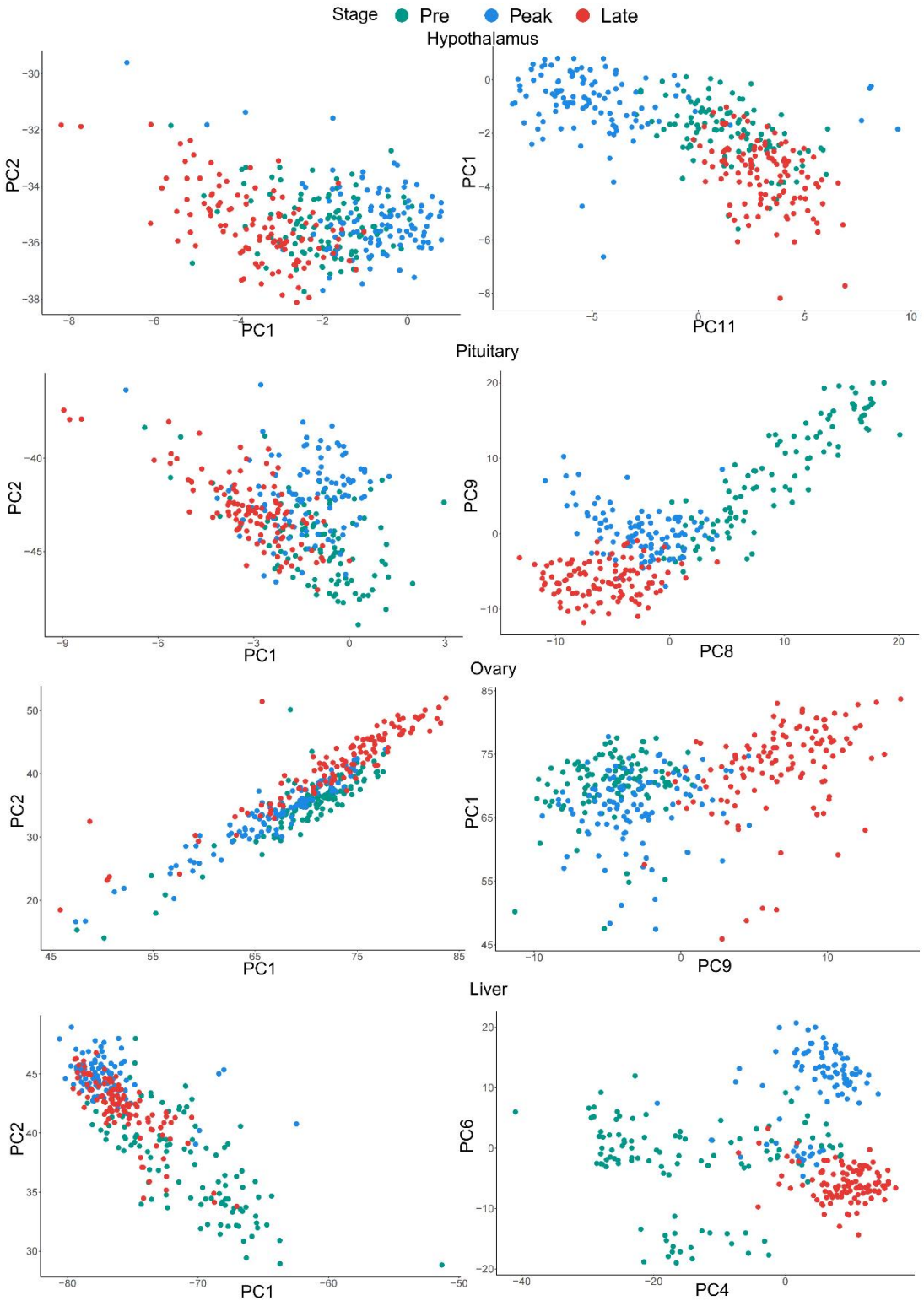

Figure S3 Principal component analysis (PCA) of gene expression within tissues.

PC1 and PC2 do not group the samples by laying stage, whereas higher-order principal components effectively differentiate the samples across stages.

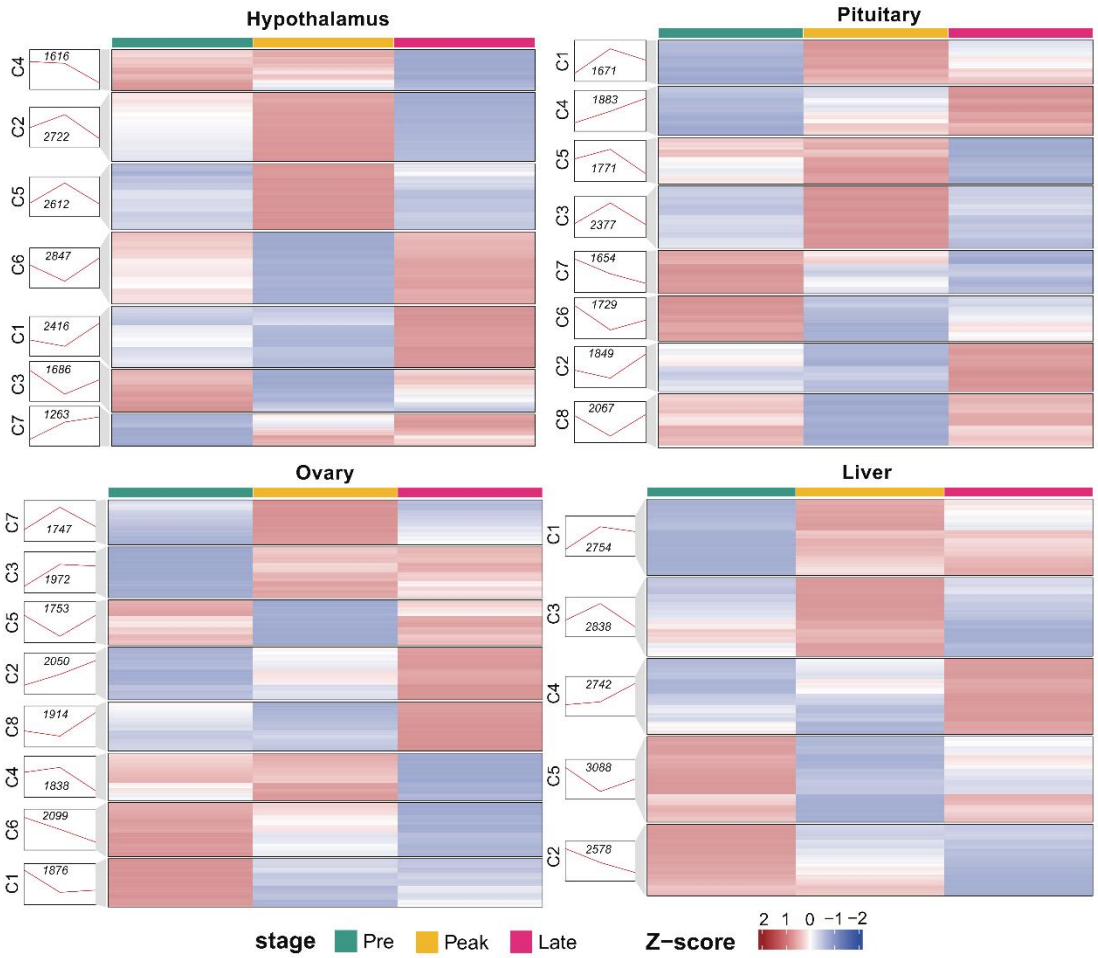

Figure S4 Time-course analysis across laying stages.

The number of clusters was determined using the elbow method. The red lines in the left panel represent the gene expression patterns within each cluster, with the numbers indicating gene size. The heatmap displays the expression levels standardized to Z-scores.

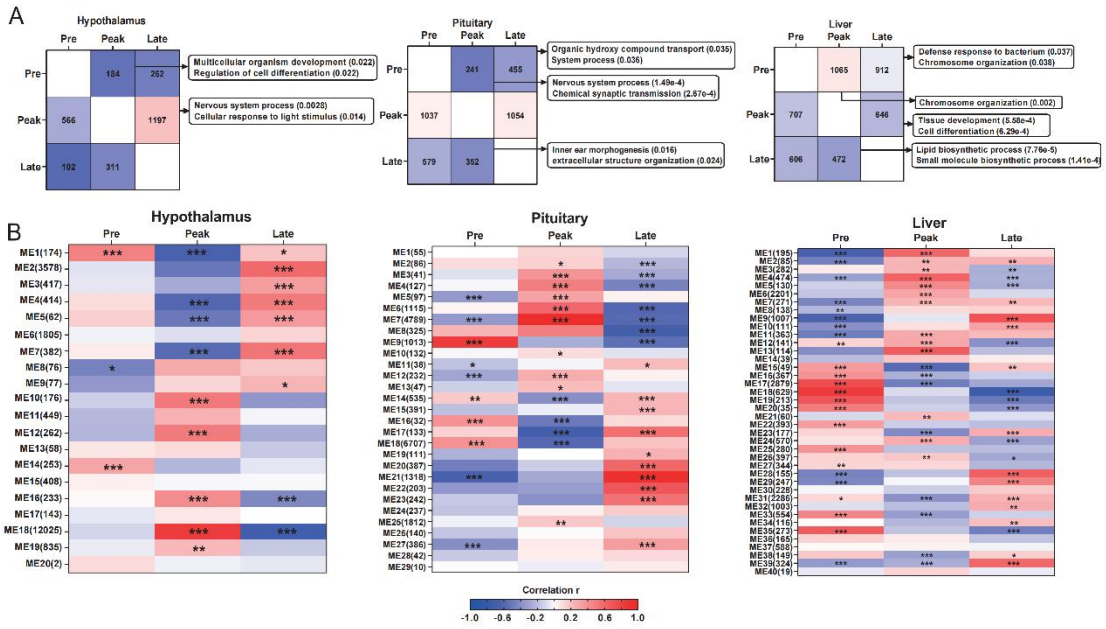

Figure S5 (A) The number of differentially expressed genes (DEGs) across laying stages in three tissues and the corresponding Gene Ontology (GO) enrichment are shown. The heatmap displays the number of DEGs with higher expression in the query stage compared to the reference stage. Row labels indicate the query stages, and column labels indicate the reference stages. For example, the cell at row "Pre" and column "Peak" represents the number of genes upregulated in the pre-laying stage relative to the peak-laying stage. (B) Co-expression network modules identified in three tissues and their correlations with egg-laying stages. Numbers in parentheses indicate the number of genes within each module. Asterisks denote the significance of correlations (* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$).

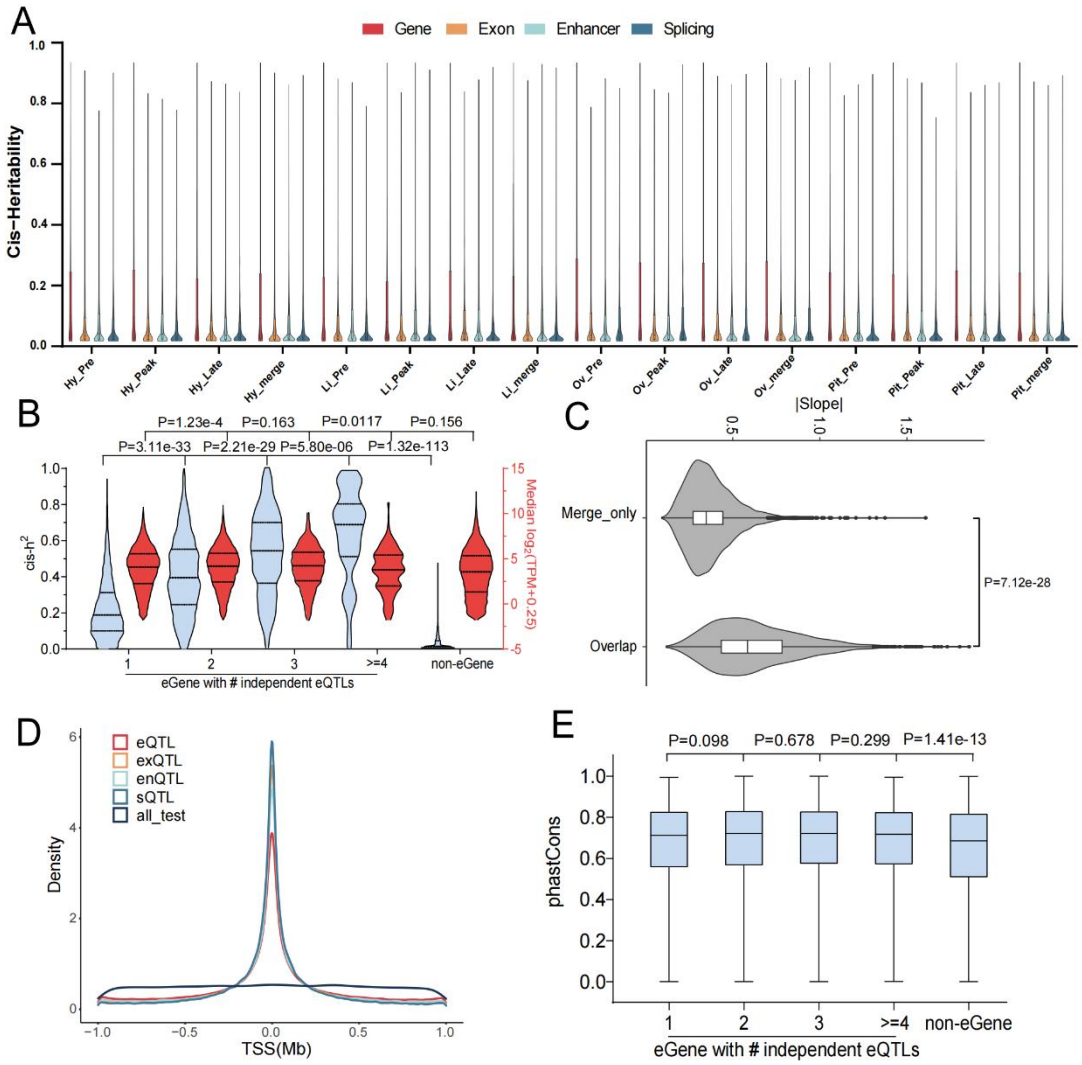

Figure S6 Compilation of molQTL mapping.

(A) Violin plot showing the distribution of cis-heritability estimates for each molecular phenotype across different groups. In the hypothalamus, pituitary, ovary, and liver, 21,829, 20,783, 21,770, and 18,050 genes; 57,981, 56,137, 53,488, and 47,568 exons; 8,921, 8,074, 7,391, and 6,893 enhancers; and 64,969, 59,251, 87,063, and 48,364 alternative splicing events were analyzed, respectively. (B) Comparison of cis-heritability and expression levels between eGenes regulated by different numbers of independent eQTLs. P-values were calculated using two-sided, unpaired Student's t-tests. A total of 20,815 eGenes were included in this analysis. Lines within each violin plot indicate the median and interquartile ranges. (C) Violin and box plots showing the comparison of effect sizes between eQTLs detected only in the combined laying stages and those overlapping across stages. A total of 20,815 eGenes were included in this analysis. (D) Distribution of four types of molecular QTL and all background variants around the transcription start site (TSS). (E) Comparison of phastCons scores

between eGenes controlled by different numbers of independent eQTLs and non-eGenes. P-values were calculated using two-sided, unpaired Student's t-tests. A total of 25,016 genes were included in this analysis.

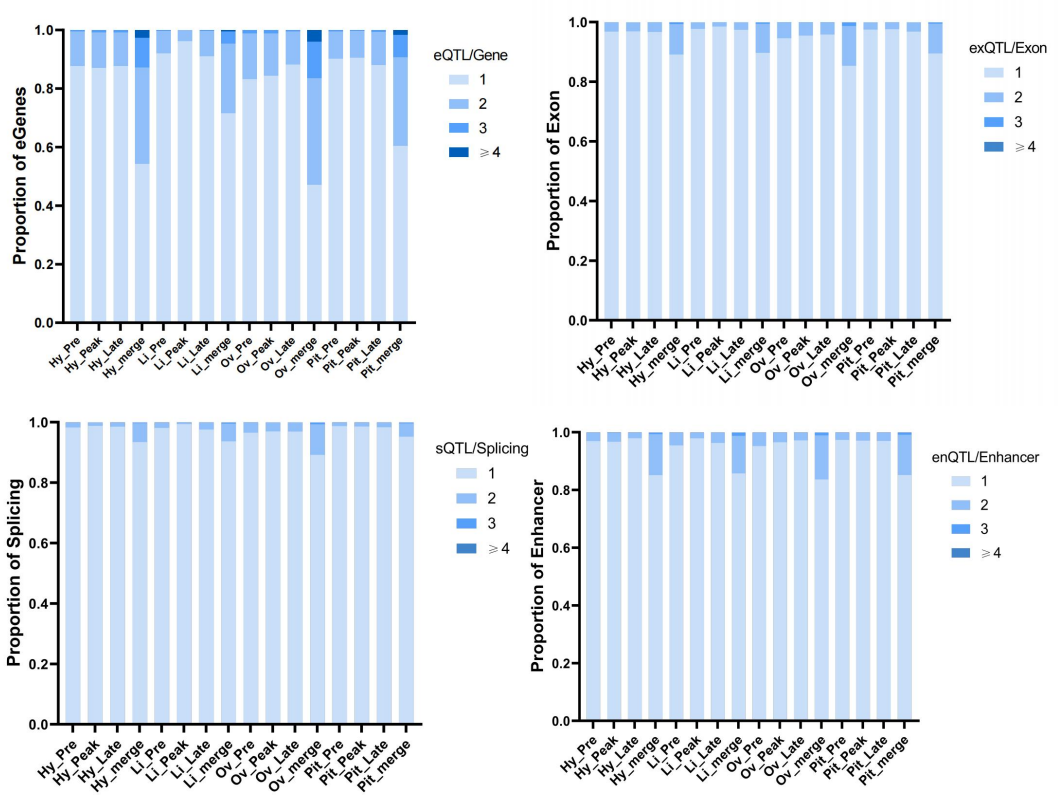

Figure S7 Proportion of molecule phenotype with different numbers of independent molQTL being detected

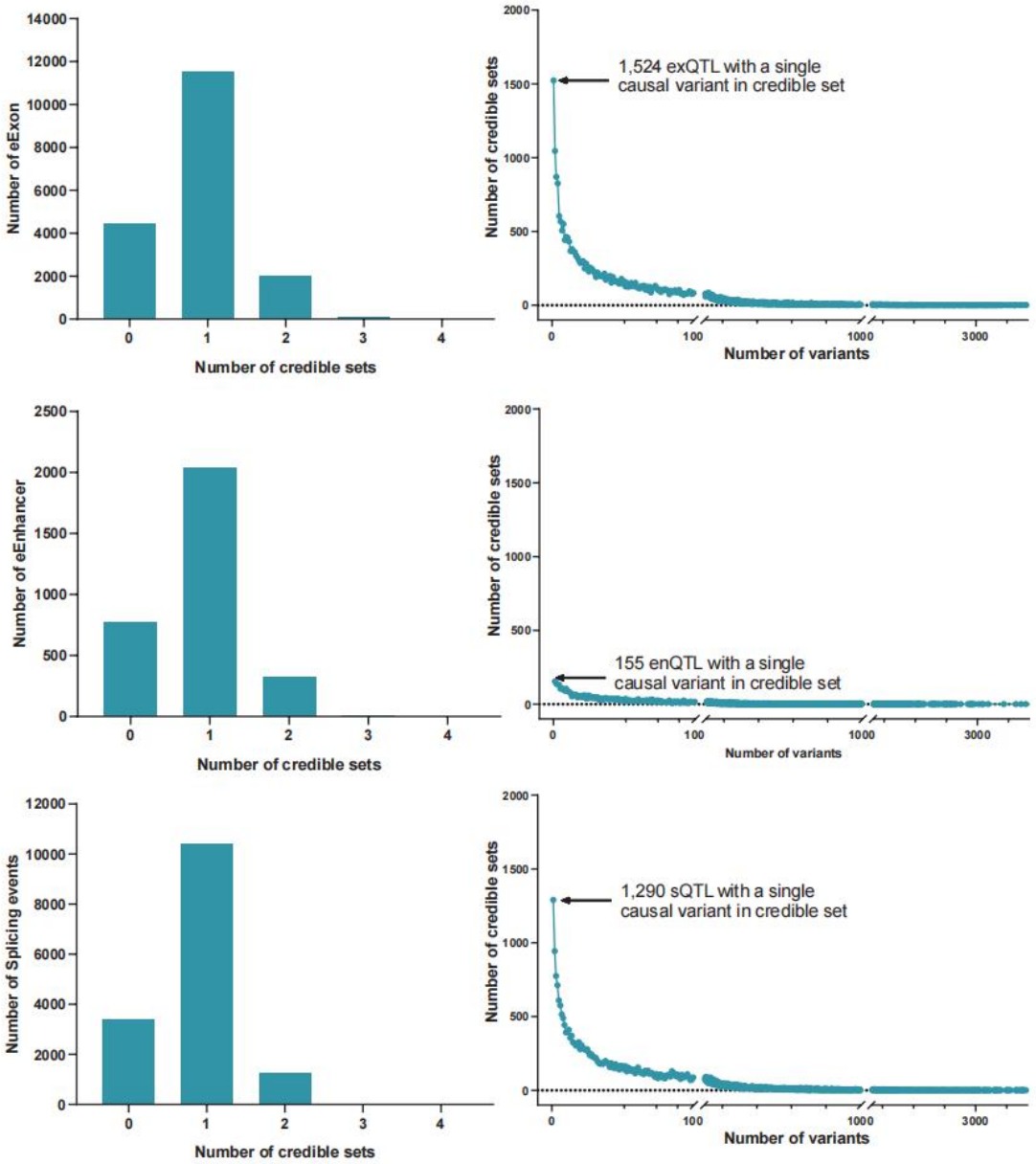

Figure S8 Fine-mapping resolution of molecular QTLs identified using SuSiE. Number of fine-mapped credible sets and the fine-mapping resolution of exQTLs, enQTLs, and sQTLs identified using SuSiE. The x-axis represents the number of SNPs within each credible set, and the y-axis represents the number of credible sets containing the corresponding number of SNPs.

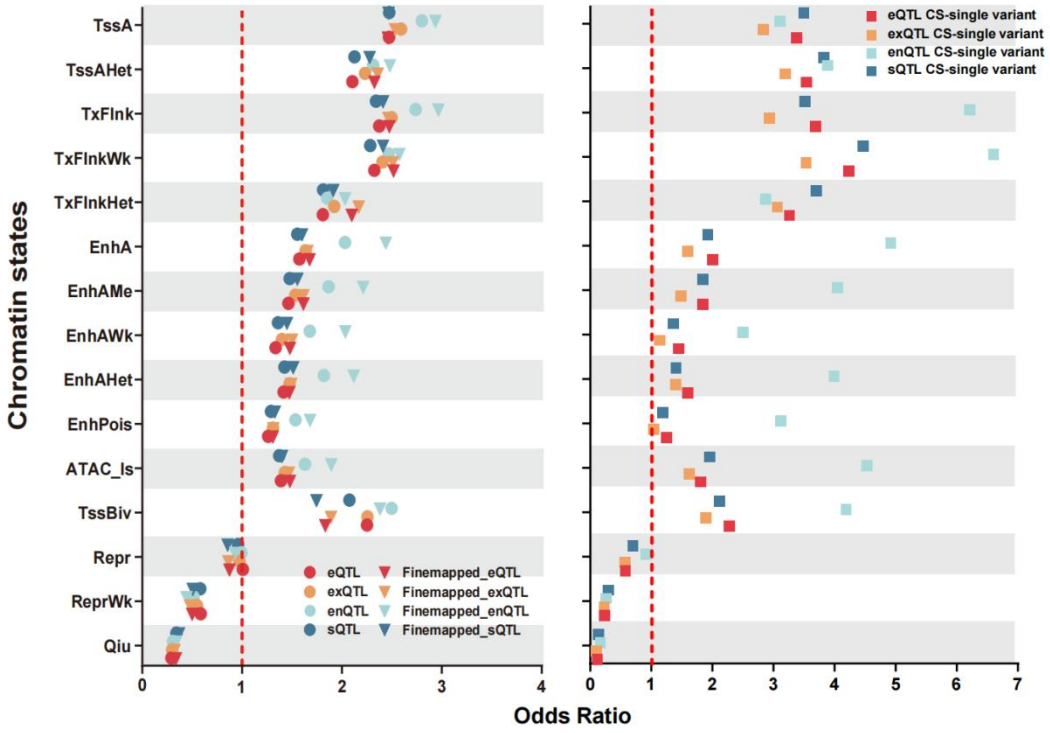

Figure S9 Functional enrichment of molecular QTLs before and after fine-mapping. Enrichment odds ratios of molecular QTLs before and after fine-mapping (left panel) and enrichment of credible sets containing only a single variant (right panel) across 26 chromatin states.

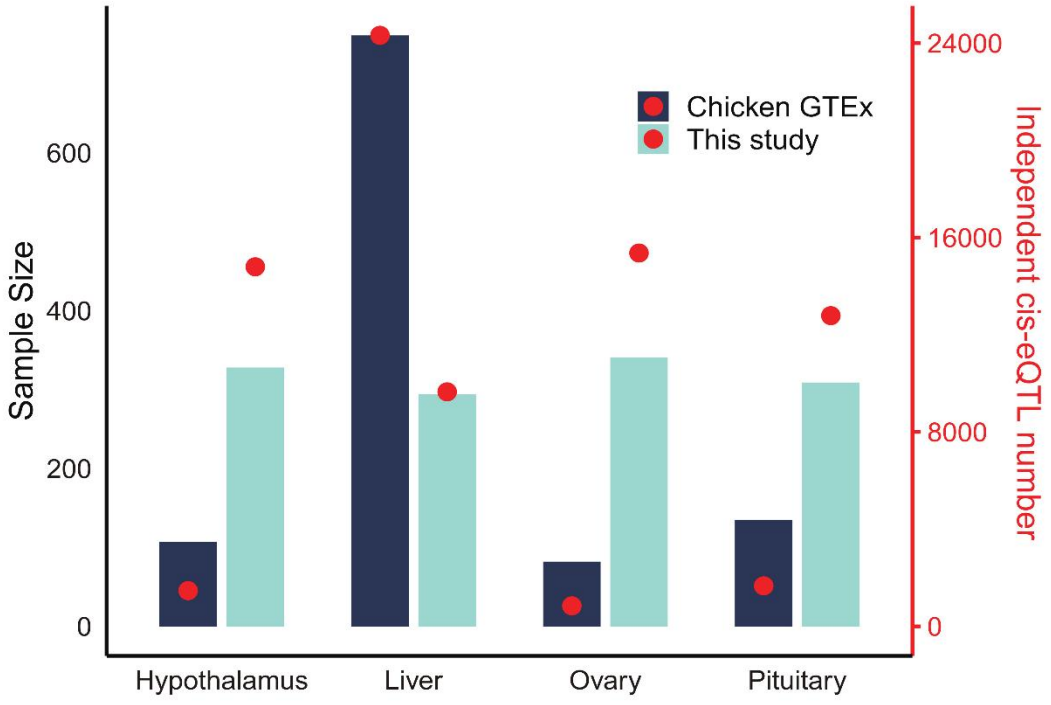

Figure S10 Comparison of eQTL mapping results between egg-laying chicken GTEx and chicken

GTEx in overlapping tissues. The figure displays results for the merged stages group, with the left y-axis representing sample size and the right y-axis representing the number of detected independent eQTL.

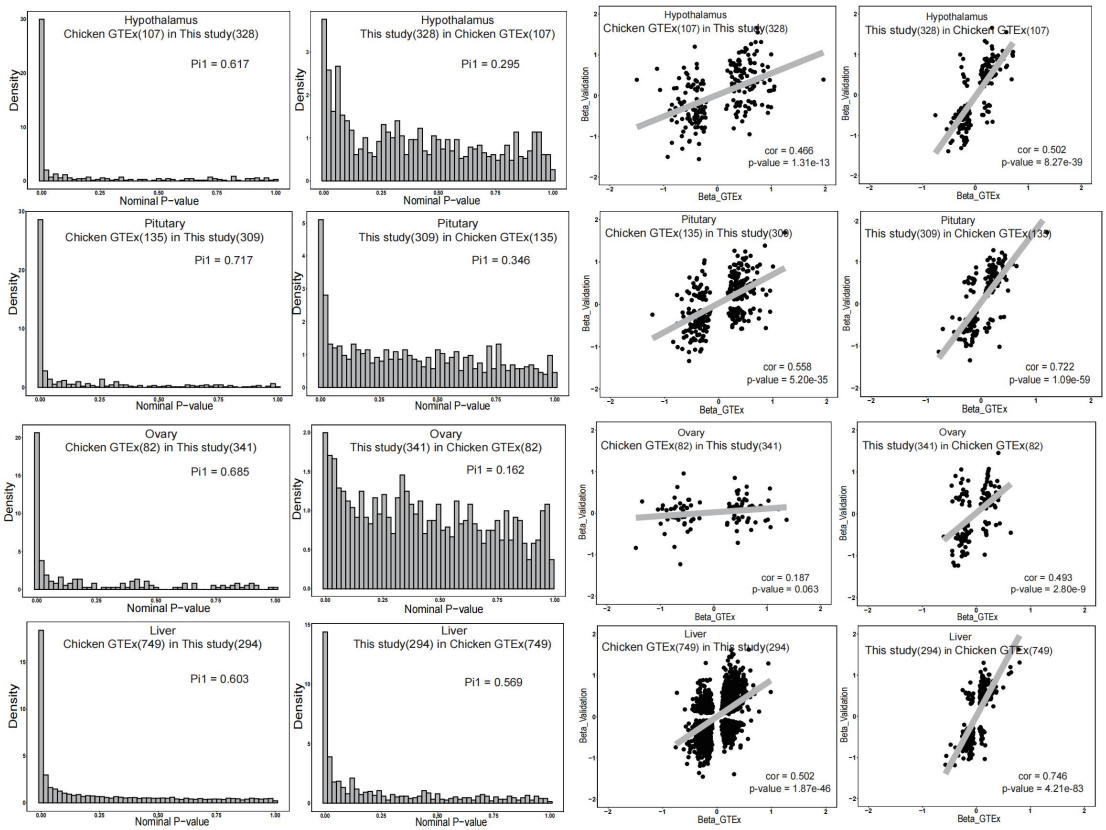

Figure S11 Molecular QTL validation using chicken GTEx.

The left panel shows the repetition rate of eQTL in each tissue, represented by Pi 1, while the right panel illustrates the effect correlation of eQTL between the two projects. The results from both studies are compared, with one serving as the discovery set and the other as the validation set.

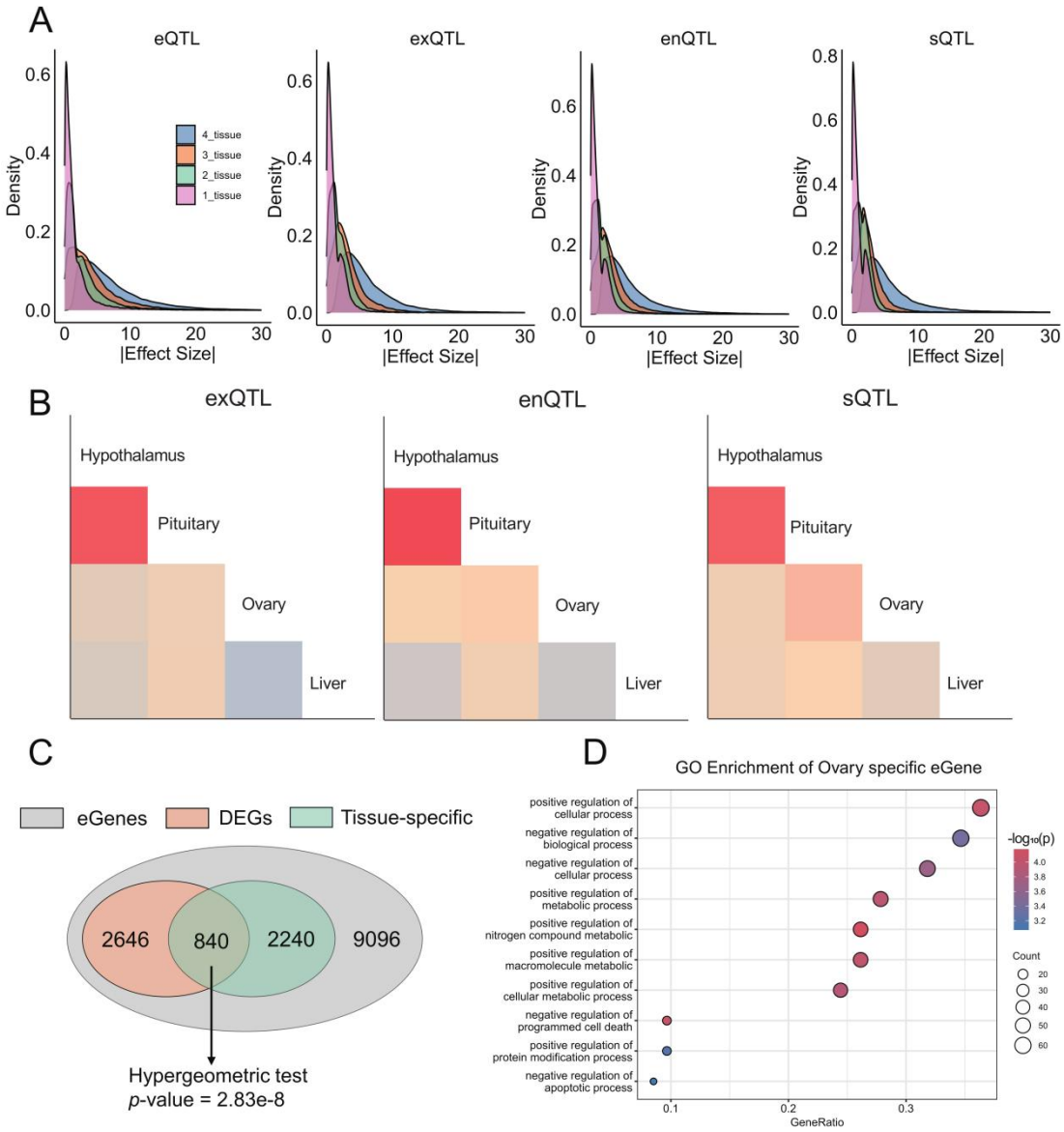

Figure S12 Regulatory patterns of molecular QTL across tissues.

(A) Distribution of molecular QTL effects shared across different numbers of tissues, with effect sizes obtained using MashR. (B) Heatmap showing the Spearman correlation of exQTL, enQTL and sQTL effect sizes between tissues, where effect sizes were estimated using MashR. (C) The relationships between all eGenes, tissue-differentially expressed genes, and tissue-specific eGenes. Overlaps among these gene sets were evaluated using hypergeometric tests. (D) GO enrichment of ovary specific eGenes.

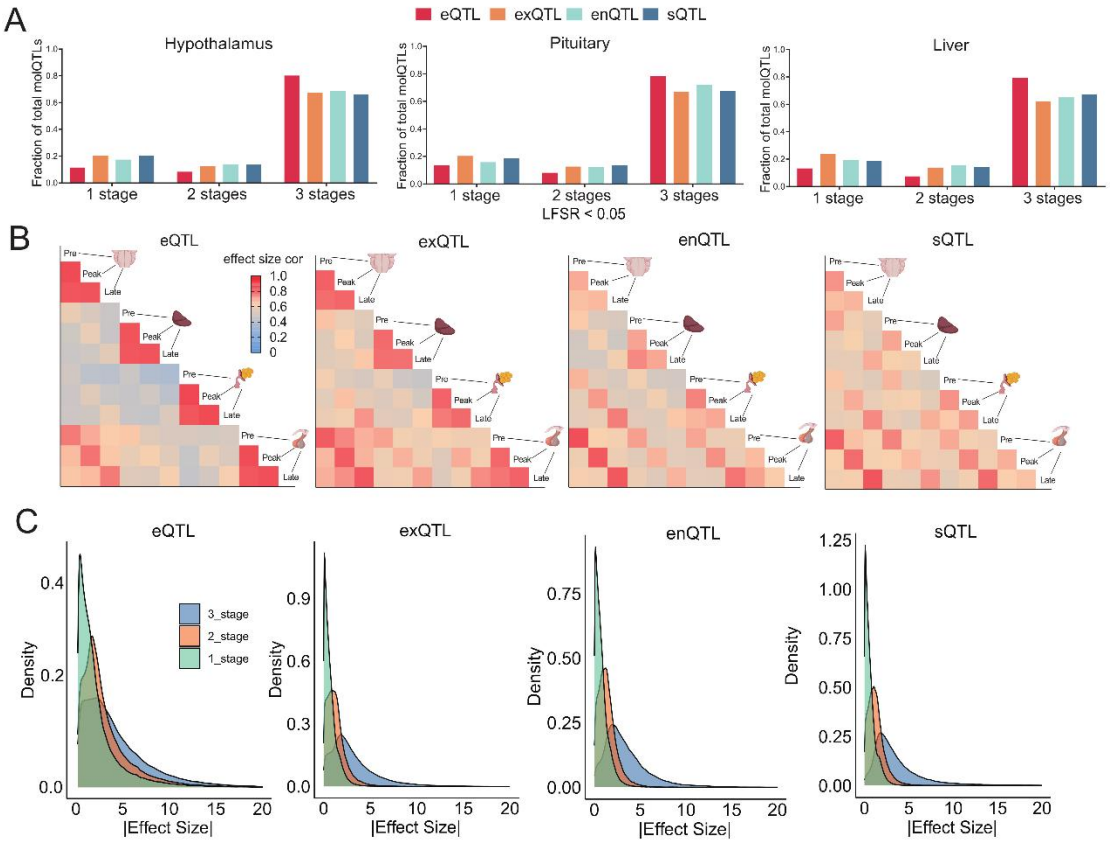

Figure S13 Regulatory patterns of molecular QTL across egg-laying stages.

(A) The bar plot shows the proportion of four types of molecular QTL shared across different laying stages, as determined by an LFSR < 5% threshold obtained using MashR. (B) Correlation of molecular QTL effects across laying stages. Illustrative images were obtained from BioRender (https://www.biorender.com/). (C) Distribution of molecular QTL effects shared across different numbers of laying stages.

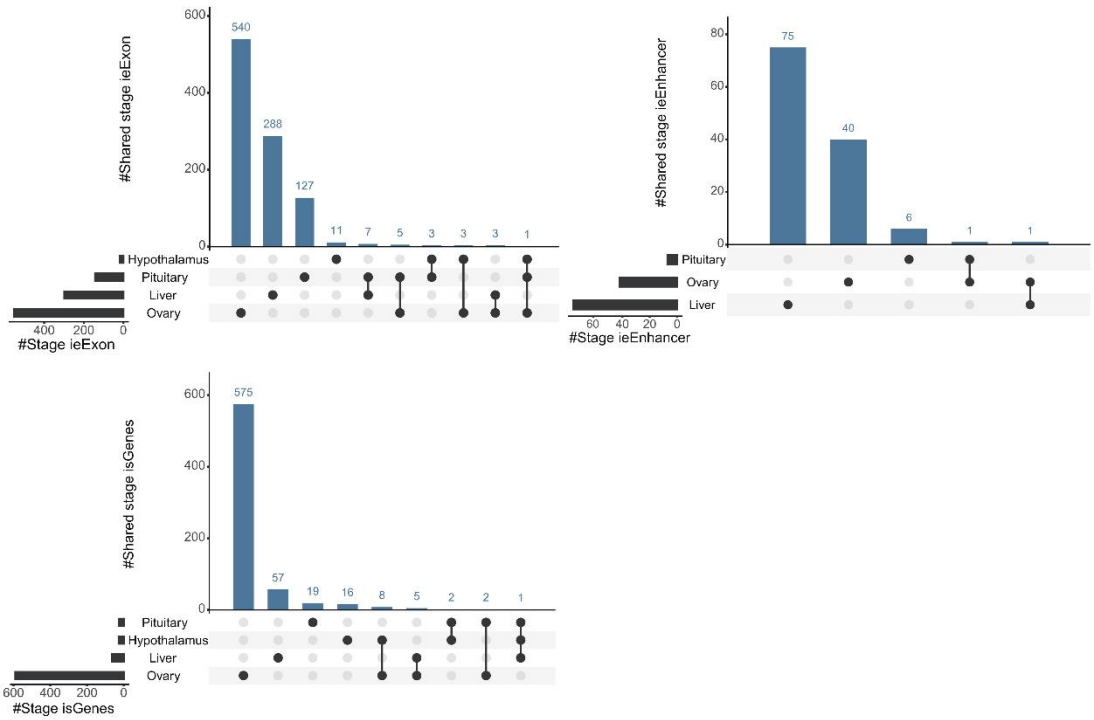

Figure S14 UpSetR plot depicting the sharing patterns of stage-interaction ieExons, ieEnhancers, and isGenes across tissues. ieExon: Exons with significant stage interaction exQTL; ieEnhancer: Enhancers with significant stage interaction enQTL; isGenes: Genes with significant stage interaction sQTL.

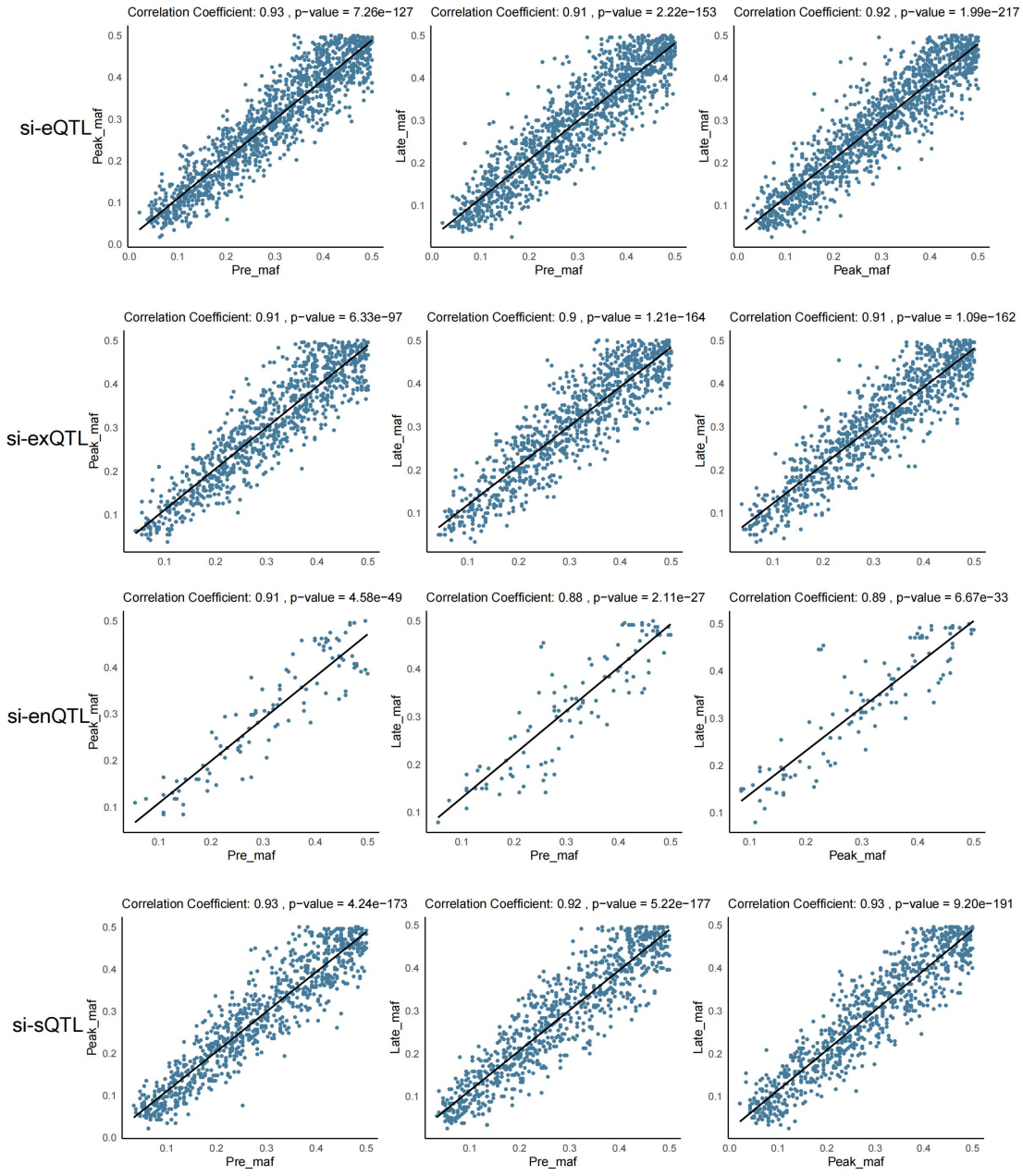

Figure S15 Correlation of minor allele frequency for top variants of stage-interaction molQTL across laying stages. Pearson correlation coefficients and corresponding P-values were calculated to assess the strength and significance of the associations.

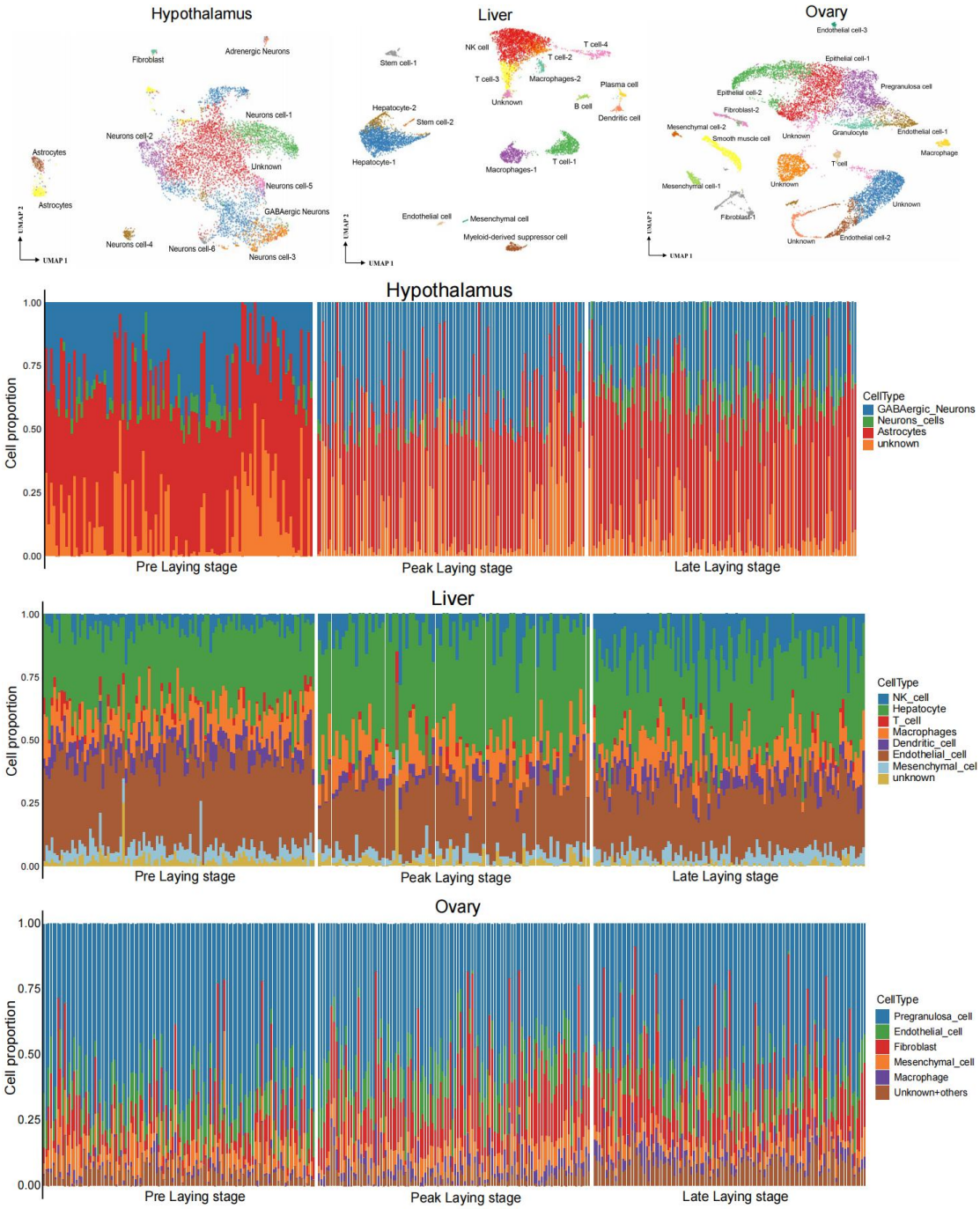

Figure S16 Single-cell deconvolution results across the three tissues.

The upper UMAP plot shows the clustering of different cell types identified from the single-cell RNA-seq data. The lower heatmap displays the proportion of each cell type in bulk RNA-seq samples after deconvolution.

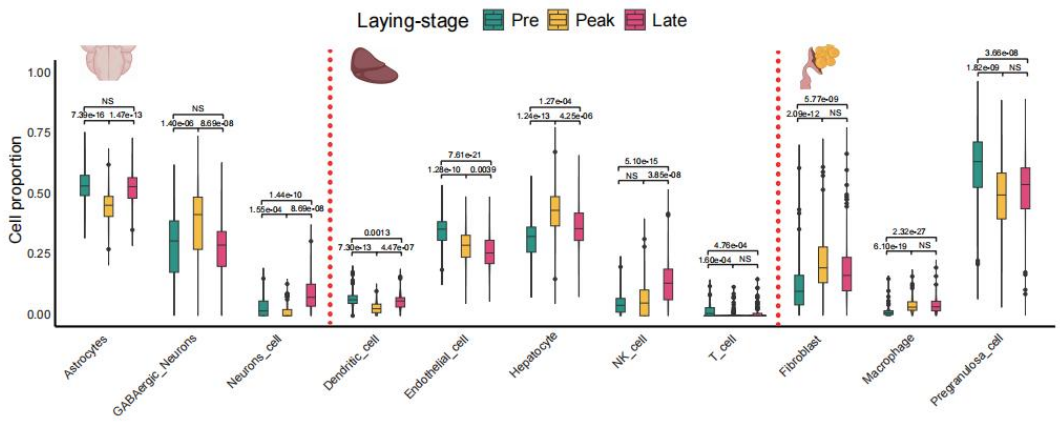

Figure S17 Cell types with significant differences across stages identified through deconvolution. Cell types with significant differences across laying stages were identified based on deconvolution results. P-values were calculated using two-sided, unpaired t-tests. Illustrative images of tissues were obtained from BioRender (https://www.biorender.com/).

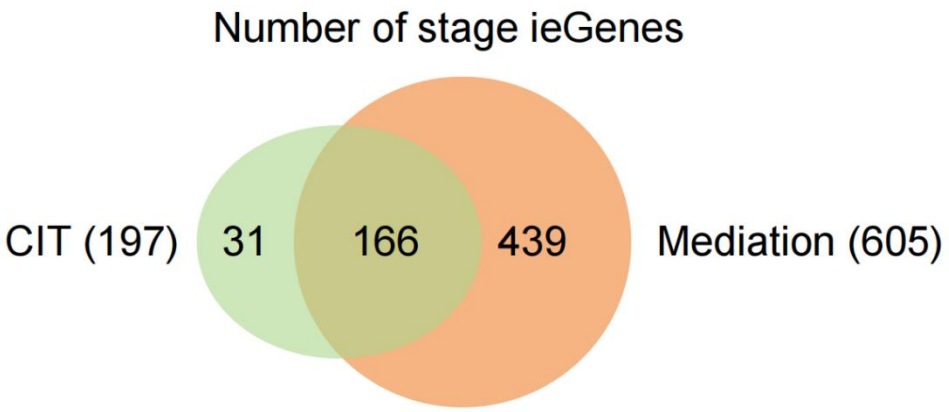

Figure S18 Comparison of CIT and mediation results: number of stage-specific ieGenes explained and their overlap.

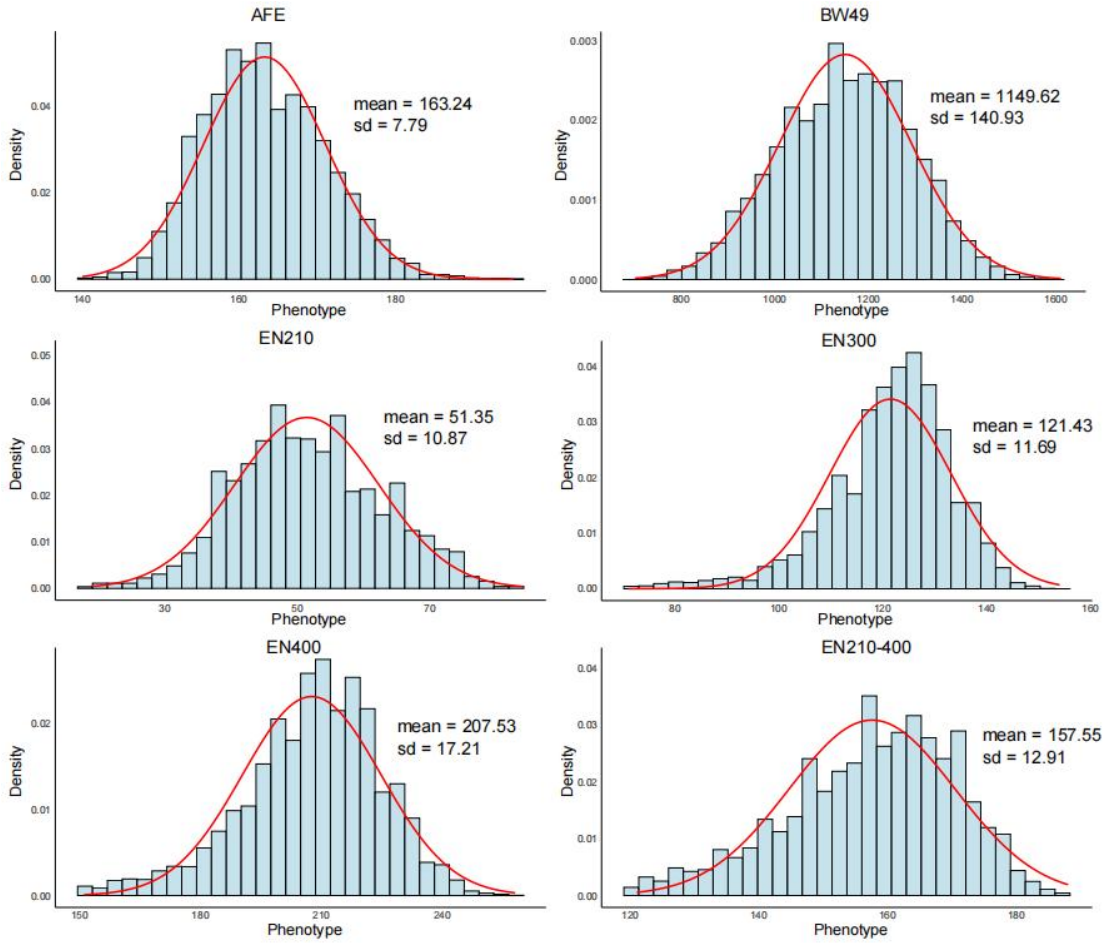

Figure S19 Distribution of six complex traits in the LDS population.

AFE, age at first egg; BW49, body weight at 49 days of age; EN210, egg number at 210 days; EN300, egg number at 300 days; EN400, egg number at 400 days; EN210–400, cumulative egg number from 210 to 400 days. The mean and standard deviation are indicated in the figure.

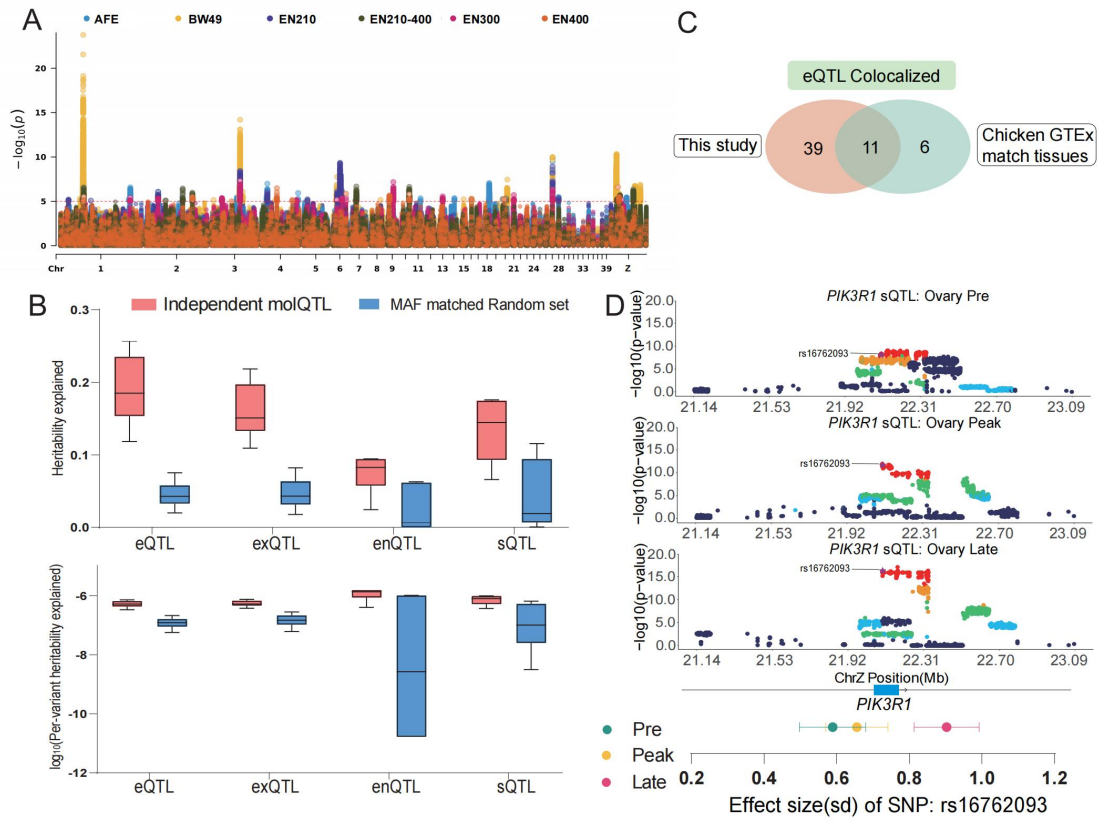

Figure S20 Integrative analysis between GWAS and molQTL.

(A) Manhattan plot of GWAS results for six complex traits in the LDS population. A total of 12,952 samples were included in this analysis. (B) Heritability of 6 complex traits explained by independent molQTL and those MAF-matched random SNPs.(n=6) (C) Comparison of the number of GWAS loci co-localized with eQTL from this study and matched tissues in Chicken GTEx. (D) The stage-interaction sQTL of *PIK3R1*, with the largest effect size observed at the Late stage. The bottom panel shows the effect sizes of rs16762093 on *PIK3R1* expression across the three laying stages, with bars representing the standard errors of the effect estimates.

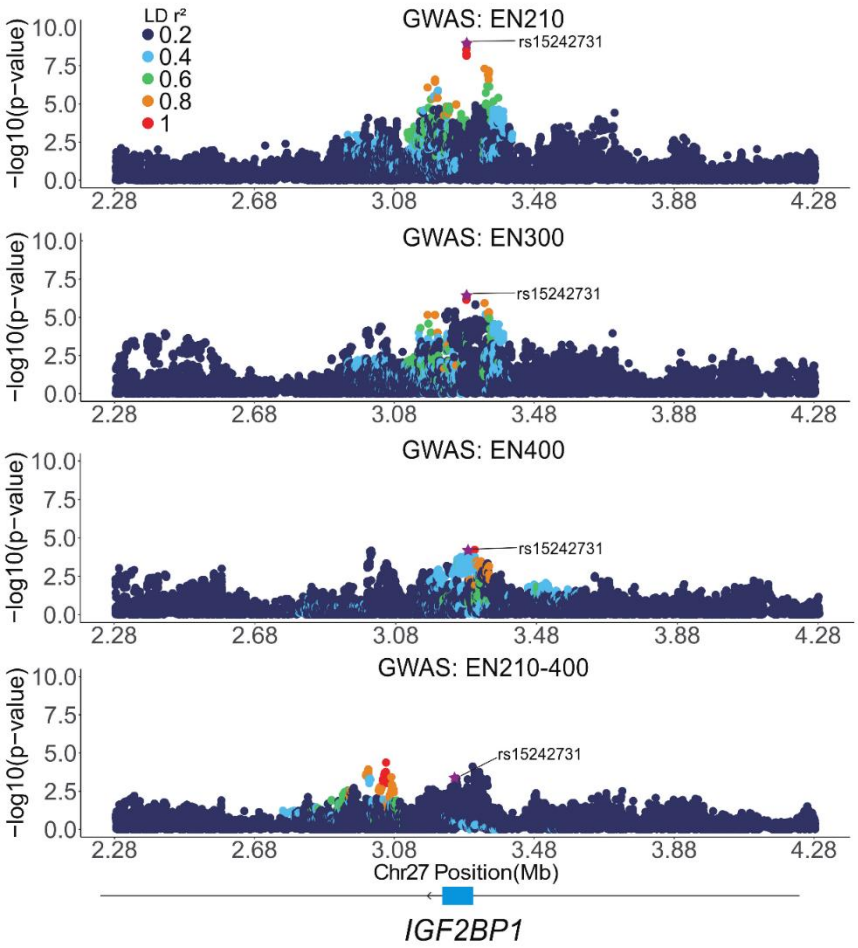

Figure S21 Zoom plot of GWAS results for egg production across different stages in the *IGF2BP1* region. Showing a gradual decrease in significance with the extension of the recording time.