Supplementary Figures

A telomere-to-telomere pangenome reveals structural variations as key drivers of complex traits in chickens

Xinting Yang[1#], Yapeng Zhang[1,4#], Lu Bai[1#], Xiaochun Ma[1#], Zhengxiao He[1], Kui Li[3], Qiao Wang[1], Luohao Xu[5], Maiqing Zheng[1], Zhangyuan Pan[1,4], Ying Qiu[3], Chaowu Yang[6], Jinyan Teng[7], Ranran Liu[1*], Lingzhao Fang[2*], Guiping Zhao[1,4*], Jie Wen[1*].

[1]State Key Laboratory of Animal Biotech Breeding, Institute of Animal Sciences, Chinese Academy of Agricultural Sciences, The Broiler Research Center of the Chinese Academy of Agricultural Sciences, Beijing, 100193, China

[2]Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark.

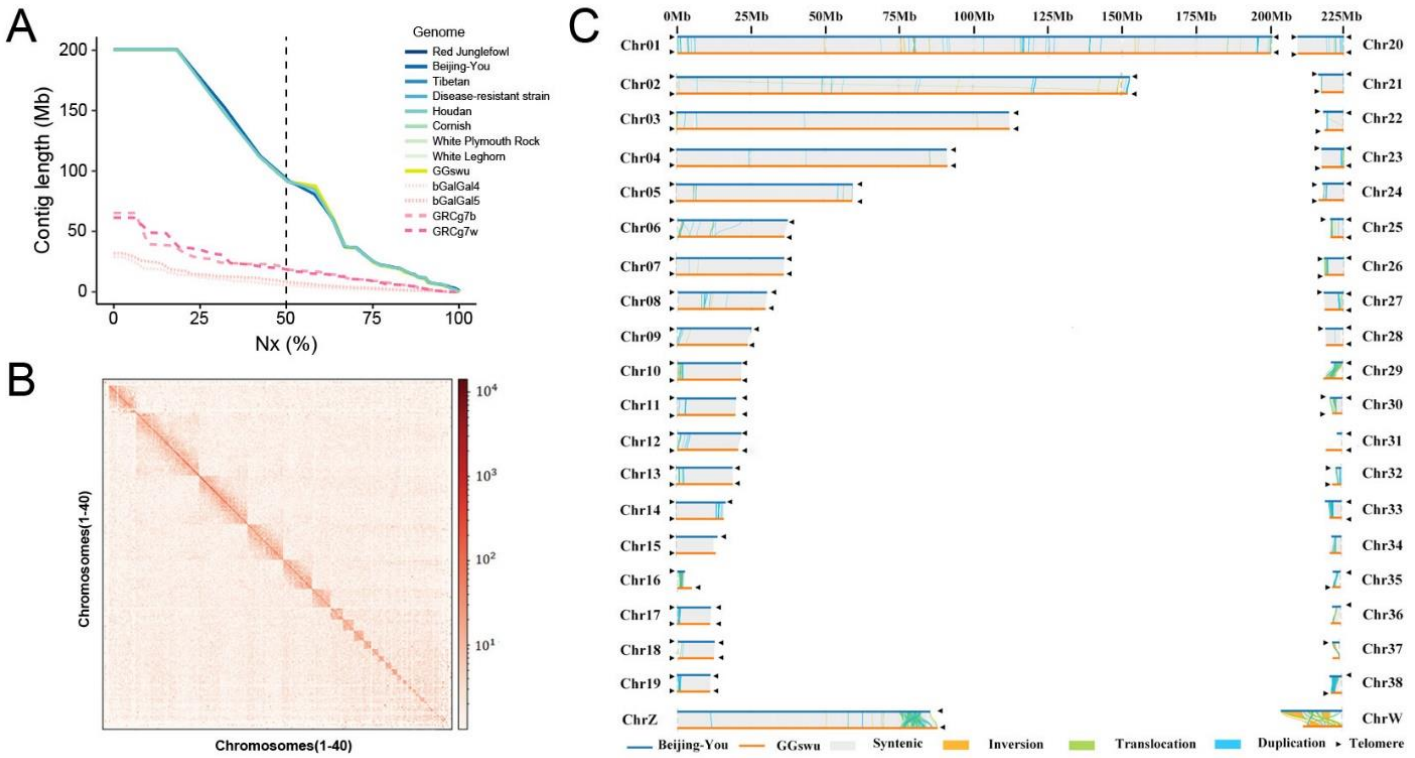
[3]Huazhi Biotechnology Co. Ltd, Changsha, China

[4]National Nanfan Research Institute (Sanya), Chinese Academy of Agricultural Sciences, Sanya,572024, China

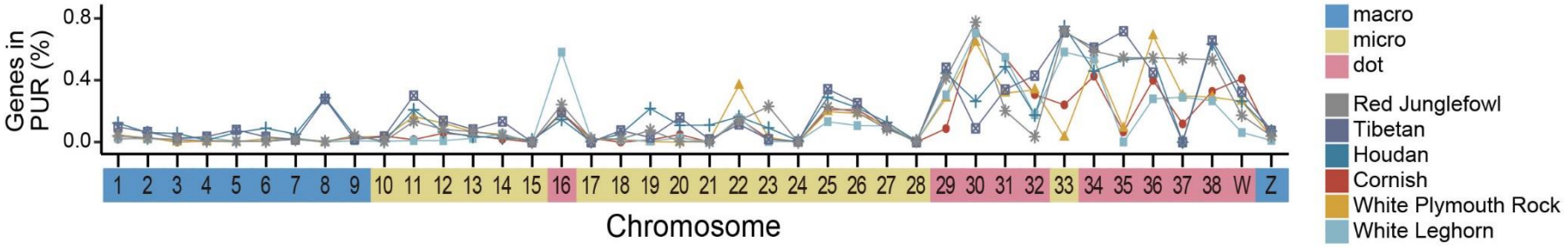
[5]School of Life Sciences, Southwest University, Chongqing, 400715, China

[6]Sichuan Animal Science Academy, Chengdu, 610066, China

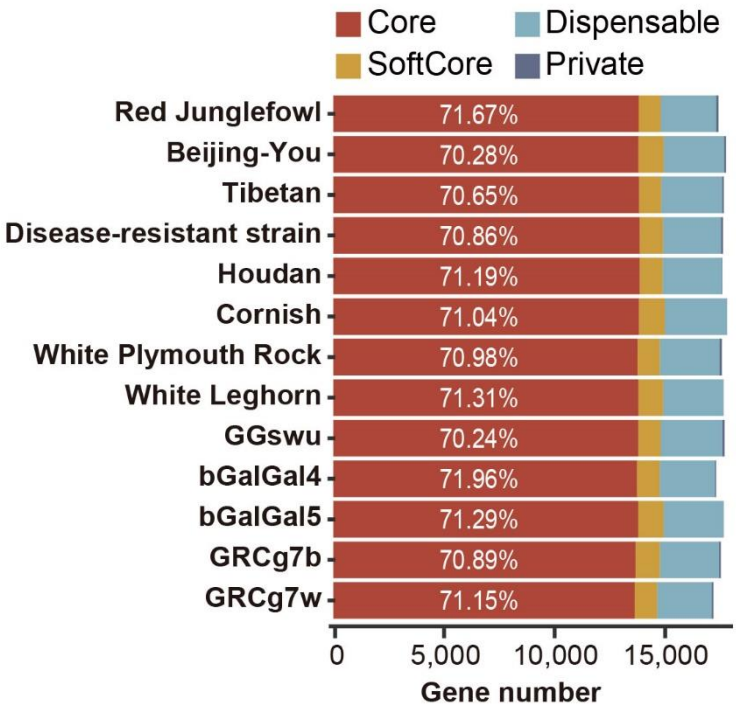
[7]College of Animal Science, South China Agricultural University, Guangzhou, 510642, China



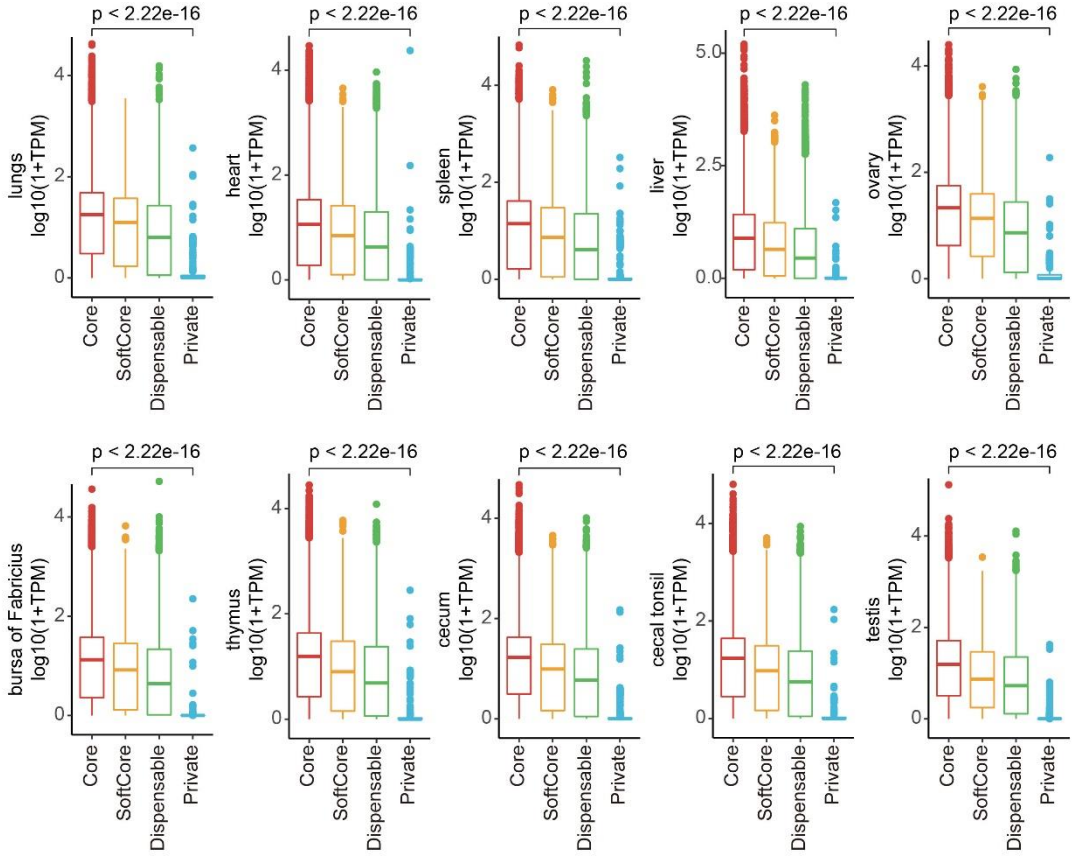
Supplementary Fig. 1. Assembly characteristics of the eight new T2T genomes. (A) Contig Nx sizes of all the genome assemblies. We accumulated the contigs from long to short when the accumulated length of the contigs reached x% of the total genome length and the length of the corresponding contig was n (x). The vertical dashed line in black corresponds to the contig N50 values of each genome assembly. **(B)** Hi–C heatmap of chromosome interactions in Beijing-You chicken. The chromosome numbers from top to bottom are in sequence as follows: W, Z, 1-38. Each heatmap is shown at a resolution of 40 kb. The dark red dots represent a high probability of interaction, and the light dots represent a low probability of interaction. **(C)** Identification of syntenic and rearranged regions between Beijing-You and GGswu using Syri software.



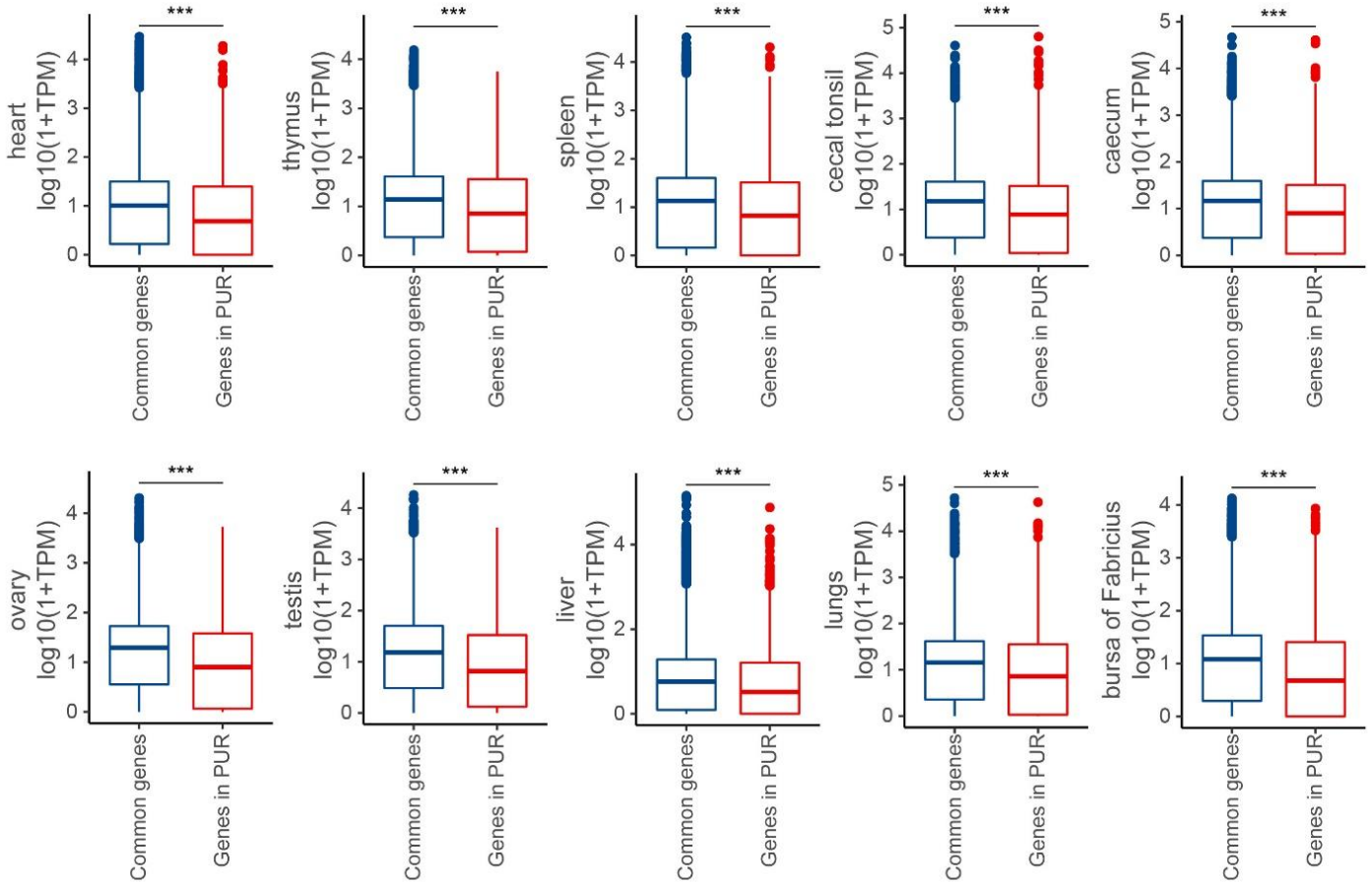
Supplementary Fig. 2. The proportion of PUR-located genes on each chromosome. The chromosomes are grouped into macro-, micro- and dot chromosomes.



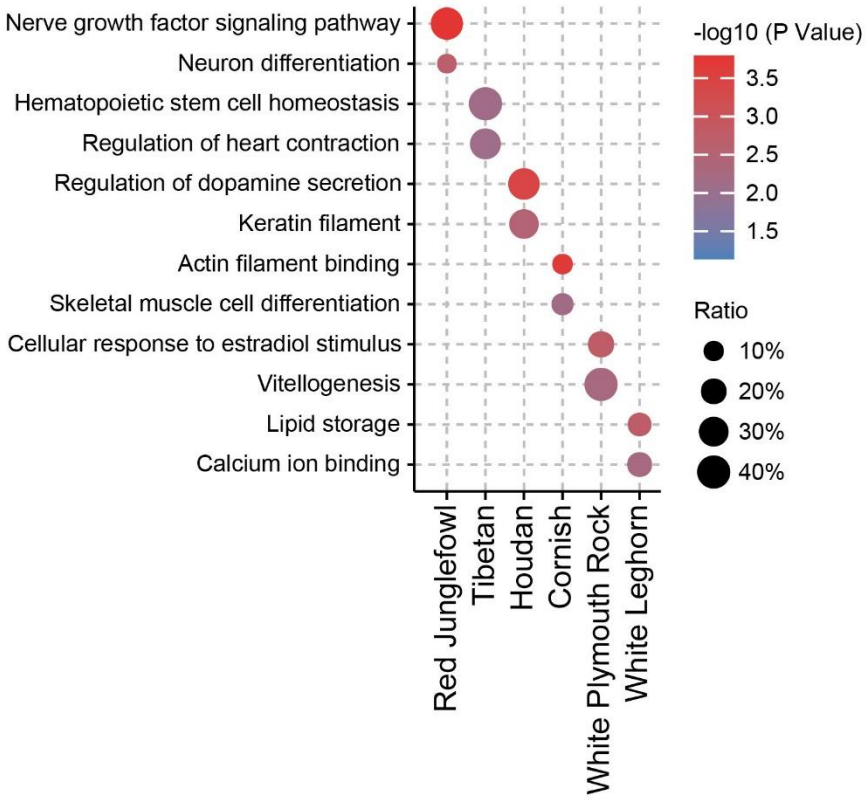
Supplementary Fig. 3. Stacked bar chart of gene family compositions across 13 chicken genome assemblies. For each assembly, the bars represent both the absolute number and the proportion of gene families classified as core (present in all 13 assemblies), soft-core (present in 12 assemblies), dispensable (2-11 assemblies), or private (present in a single assembly).



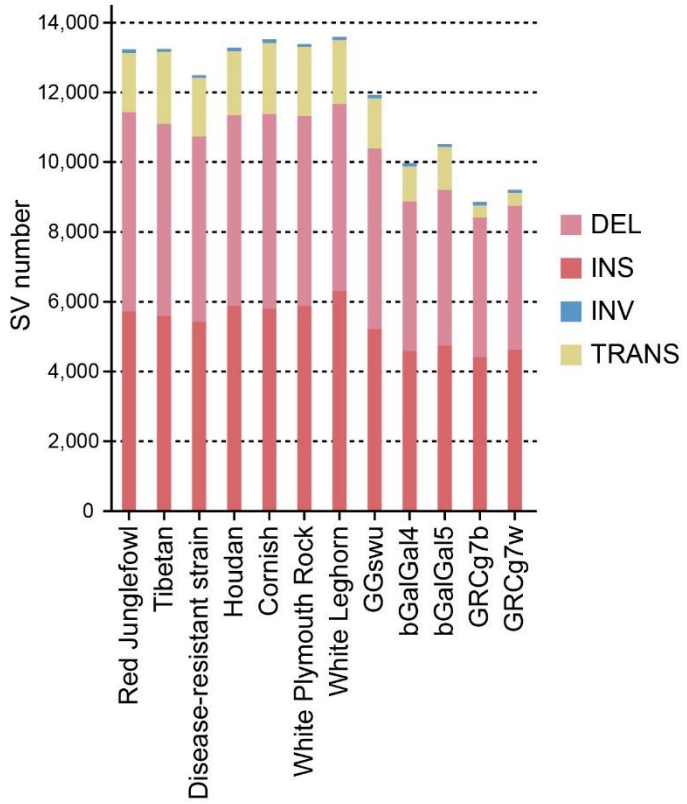
Supplementary Fig. 4. Box plots of gene expression (log_{10}[TPM+1]) for four gene family categories—core, soft-core, dispensable, and private—across 10 tissues (lungs, heart, spleen, liver, ovary, bursa of Fabricius, thymus, cecum, cecal tonsil, and testis). Statistical comparisons among gene families within each tissue were performed using two-sided t-test.



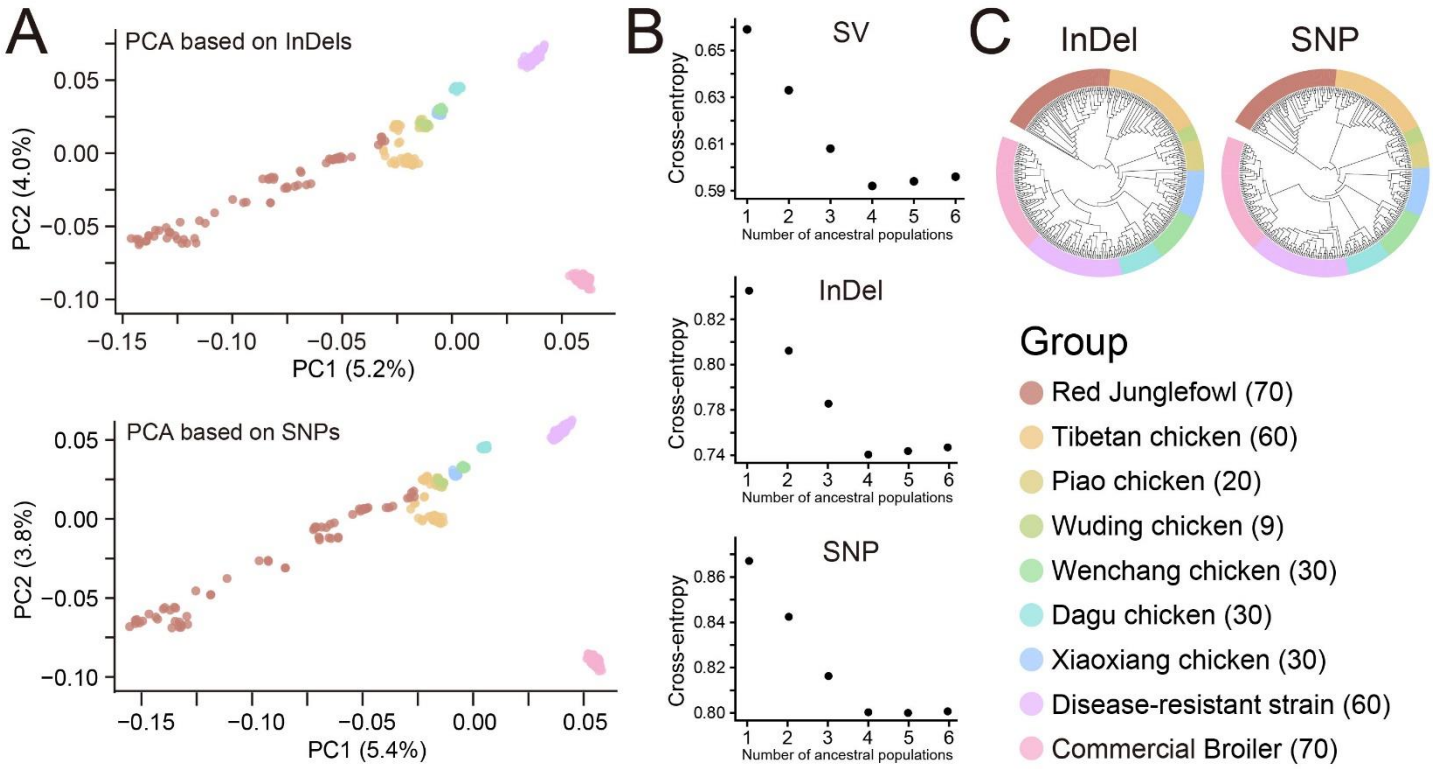
Supplementary Fig. 5. Box plots of gene expression (log_{10}[TPM+1]) for common and PUR-located genes across 10 tissues (lungs, heart, spleen, liver, ovary, bursa of Fabricius, thymus, cecum, cecal tonsil, and testis). Statistical comparisons among gene families within each tissue were performed using two-sided t-test.



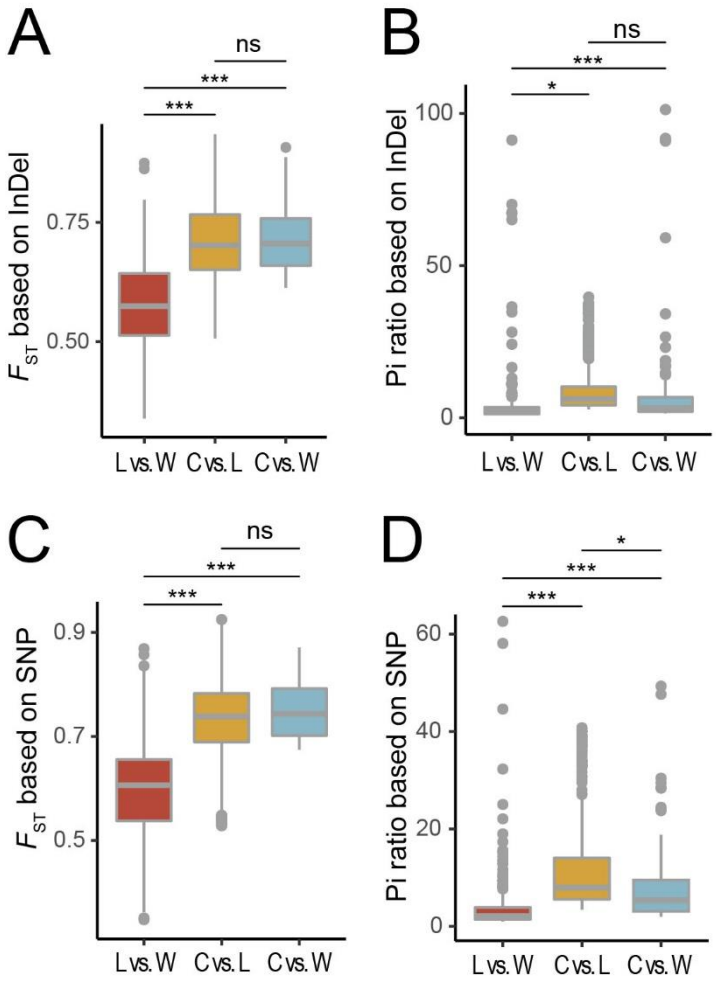
Supplementary Fig. 6. Gene Ontology enrichment analysis of PUR-located genes in each chicken breed. A dot plot shows significantly enriched biological processes for PUR-located genes in each breed.



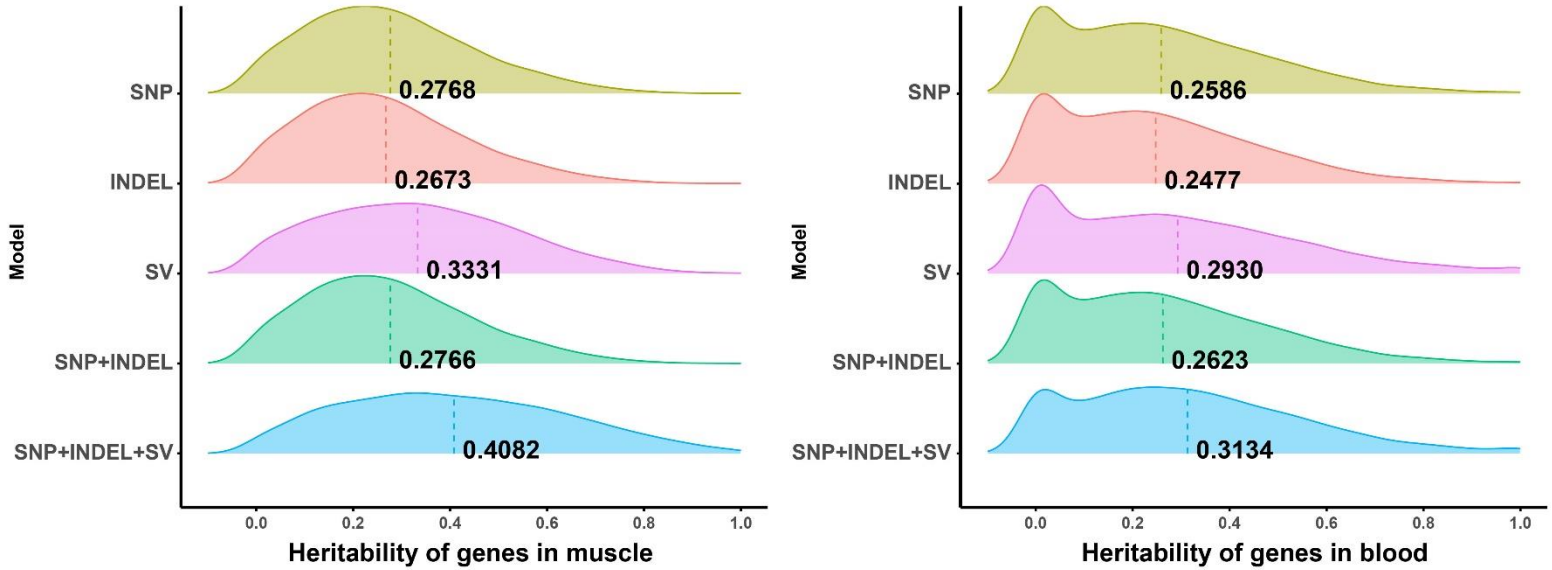
Supplementary Fig. 7. The number of different types of SVs from the nonredundant set of SVs in the 13 chicken genomes.



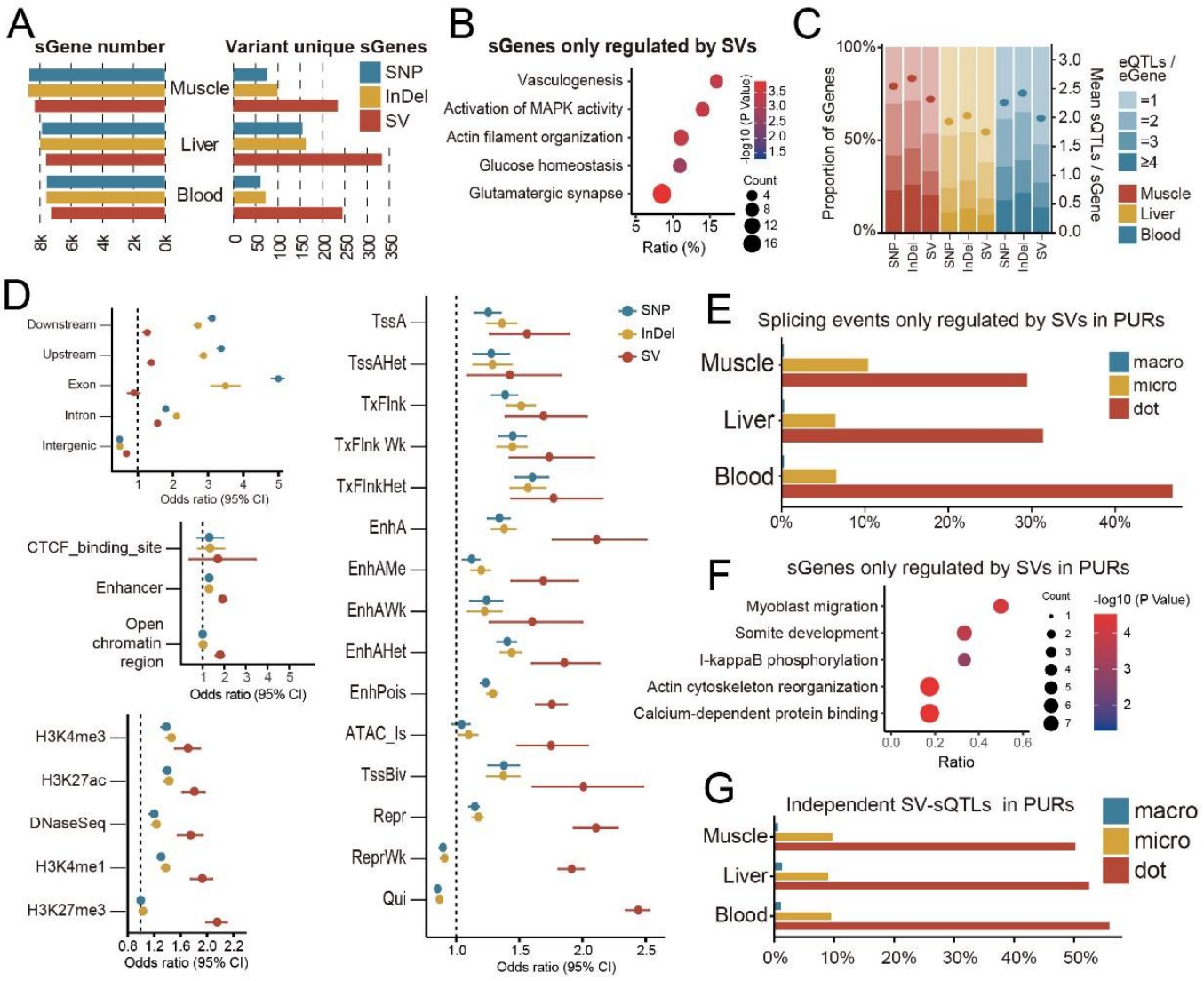
Supplementary Fig. 8. Population structure of wild, local, and commercial chicken breeds. (A) Principal component analysis (PCA) of 379 individuals based on InDels (top) and SNPs (bottom). Each dot represents one sample, colored by breed as indicated in the legend. **(B)** Cross-validation error curves from ADMIXTURE (v1.3.0) for K = 1–6 using SVs, InDels, and SNPs. **(C)** Phylogenetic trees constructed from InDel (left) and SNP (right) genotypes across all samples.



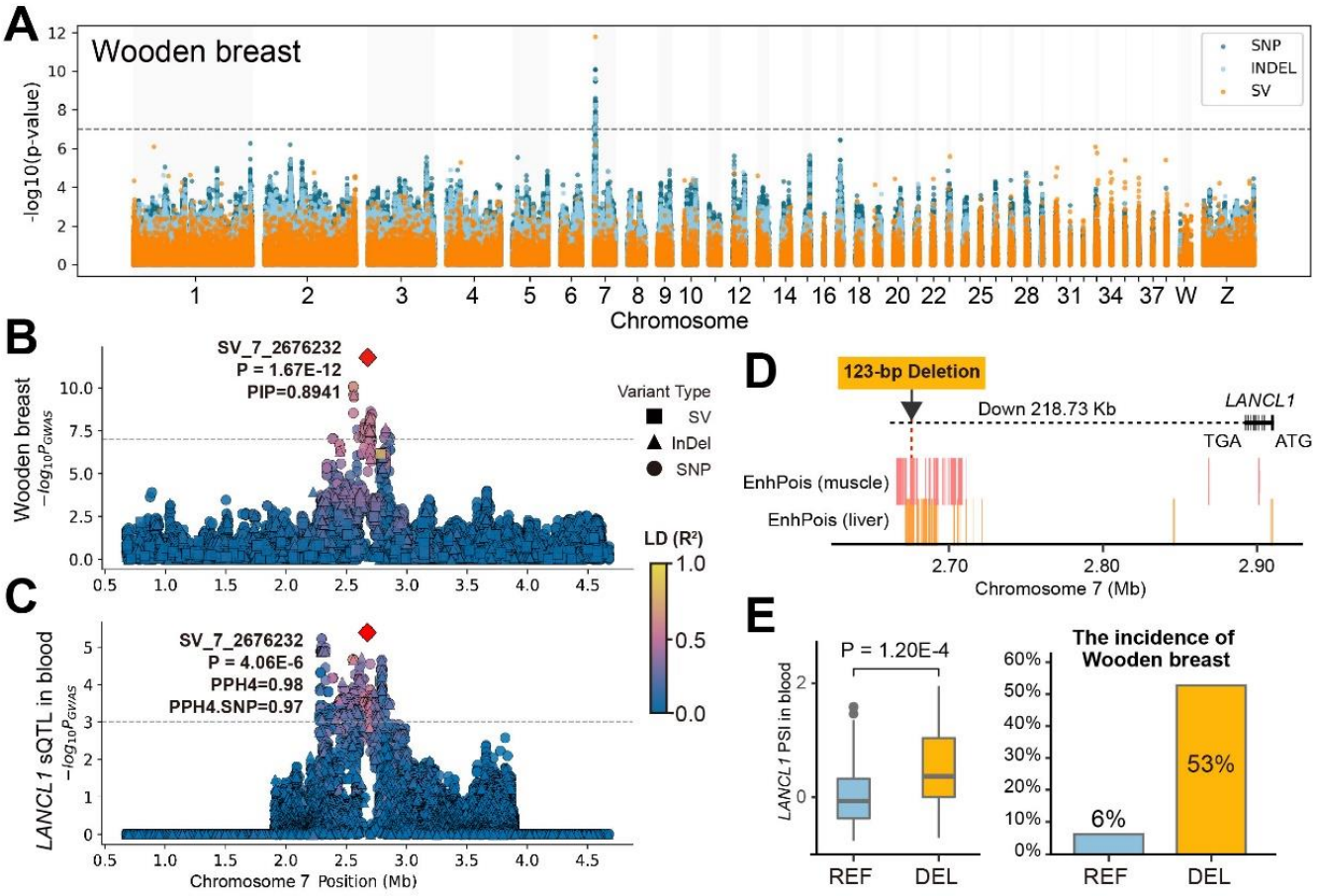
Supplementary Fig. 9. Comparison of selection metrics derived from InDels and SNPs across population contrasts. (A) Distribution of F_{ST} values calculated from InDel variants for local versus wild (L vs. W), commercial versus local (C vs. L) and commercial versus wild (C vs. W) comparisons. **(B)** Distribution of nucleotide diversity ratios for InDels across the same contrasts. **(C)** Distribution of F_{ST} values calculated from SNP variants for the three population pairs. **(D)** Distribution of π ratios for SNPs. Statistical differences between contrasts were assessed by two-sided t-test: $*P < 0.05$; $**P < 0.01$; $***P < 0.0001$; ns, $P > 0.05$.



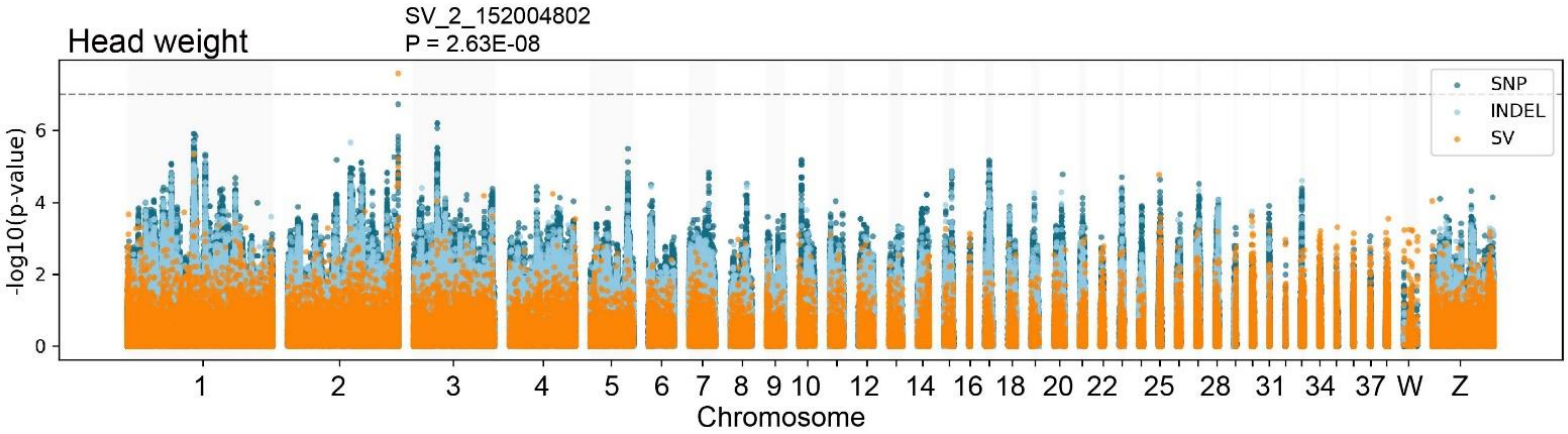
Supplementary Fig. 10. Estimated heritability (h^2) of molecular phenotypes for variant combinations from 946 breast muscle samples and 317 blood samples.



Supplementary Fig. 11. The sGene and sQTL statistics. (A) The total number of sGenes detected for each type of variation (left) and the number of unique sGenes detected for each type of variation (right). **(B)** The significantly enriched pathways of sGenes uniquely regulated by SV. **(C)** Distribution and average number of independent cis-sQTLs per gene. **(D)** Odds ratios (ORs) with 95% confidence intervals for enrichment of cis-sQTLs in genomic features, including genomic region, functional element, histone modification and chromatin state. **(E)** Proportion of sGenes that are specifically regulated by SVs in PUR among the three types of chromosomes. **(F)** The significantly enriched pathways of sGenes that are uniquely regulated by SV in PURs. **(G)** Among the three types of chromosomes, the proportion of independent sQTLs in PURs among all independent sQTLs.



Supplementary Fig. 12. Functional impact of SV in wooden breast muscle. (A) Genome-wide association plots for wooden breast status in 380 broilers using SVs, InDels and SNPs. The significance threshold ($P = 1\times10^{-7}$) is indicated as a dashed horizontal line. **(B)** Regional Manhattan plot in chromosome 7 for the wooden breast GWAS. The lead SV (red diamond) is highlighted. **(C)** Regional Manhattan plot for a cis-sQTL of *LANCL1* in blood (n = 317). **(D)** The diagram shows that a 123-bp deletion located 218.73 kb downstream of *LANCL1* overlapping muscle and liver enhancer. **(E)** Population allelic variation in *LANCL1* percent-spliced-in (PSI) in blood and the incidence of wooden breast between the samples without and with the 123-bp deletion. The *P* value is obtained from a two-sided t-test.



Supplementary Fig. 13. Functional impact of SV on head weight. Genome-wide association plots for head weight in 271 broilers using SV, InDel and SNP markers. The significance threshold ($P = 1\times10^{-7}$) is indicated as a dashed horizontal line.