Supplemental Materials

Supplemental Figs. S1-27

Supplemental Fig. S1. Schematic representation of the reciprocal cross pig model and summary of multi-omics data generated in this study.

Supplemental Fig. S2. Summary of whole-genome sequencing, *in situ* Hi-C, rRNA-depleted RNA-seq and ChIP-seq data.

Supplemental Fig. S3. Construction of chromosome-span haplotypes for hybrid pigs using genome-wide heterozygous SNVs.

Supplemental Fig. S4. Construction of haplotype-resolved Hi-C maps for hybrid pigs using SNV phasing and a local imputation method.

Supplemental Fig. S5. Summary of phased Hi-C and RNA-seq data.

Supplemental Fig. S6. Assessment of BNBC quantile normalization.

Supplemental Fig. S7. Assessment of Hi-C normalization using multiHiCcompare.

Supplemental Fig. S8. Initial characteristics of 3D nuclear organization in the diploid pig genome.

Supplemental Fig. S9. Genomic features of compartments and TADs in the haplotype-resolved pig genome.

Supplemental Fig. S10. Allelically compartmental rearrangement between tissues.

Supplemental Fig. S11. Allelically compartment rearrangement and shifted TAD boundaries.

Supplemental Fig. S12. Collection of known imprinted genes.

Supplemental Fig. S13. Allelically shifted TAD boundaries between tissues.

Supplemental Fig. S14. Haplotype-resolved interrogation of PEIs in gene expression control.

Supplemental Fig. S15. Reorganized haplotype-resolved PEIs between tissues.

Supplemental Fig. S16. Representative examples of tissue-specific PEIs.

Supplemental Fig. S17. Representative examples of specific PEIs for germ layer markers.

Supplemental Fig. S18. Imprinting-specific PEI organizations.

Supplemental Fig. S19. Characteristics of allelic structural variants (SVs).

Supplemental Fig. S20. Effects of sequence variation and histone modification on allelic PEI rewiring in F1 hybrids.

Supplemental Fig. S21. Breed-specific mutations induced allele-specific abolition of two CTCF-mediated loops.

Supplemental Fig. S22. Representative examples of genes showing differential RPS between Berkshire and Tibetan alleles across the different tissues.

Supplemental Fig. S23. Differential RPS between Berkshire and Tibetan alleles in the skeletal muscle.

Supplemental Fig. S24. Characteristics of homolog pairing in the pig genome at 20-kb resolution.

Supplemental Fig. S25. Characteristics of homolog pairing in the pig genome at 100-kb resolution, which recapitulates the findings obtained at 20-kb resolution.

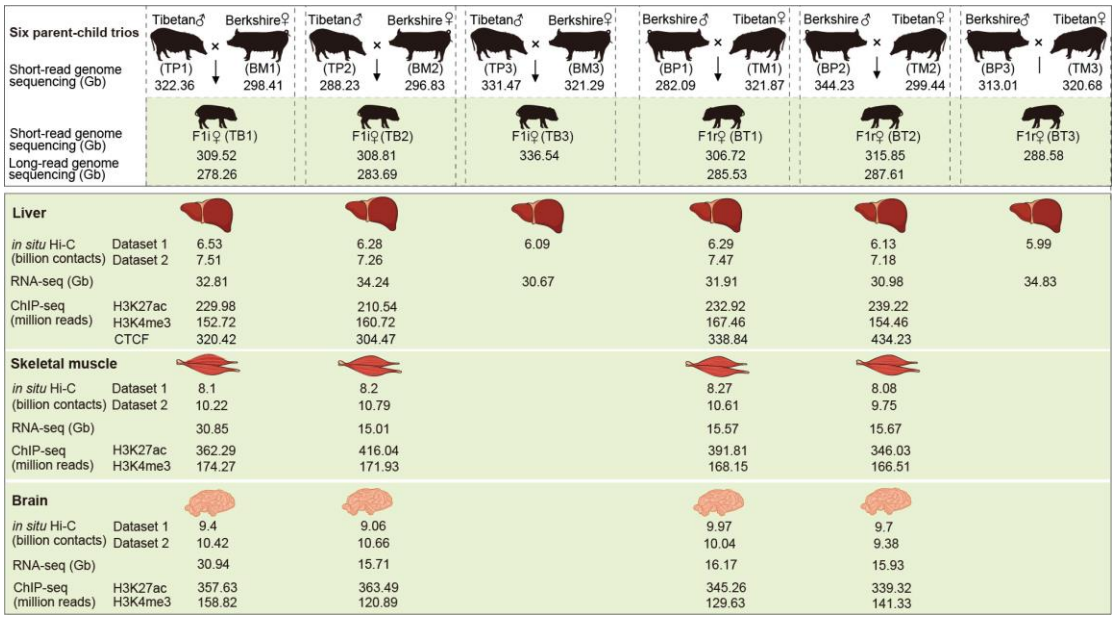
Supplemental Fig. S26. Detection of homolog pairing with heterozygous SNVs at different confidence levels to assign parental origins of chromatin interactions.

Supplemental Fig. S27. Human GWAS and pig QTLs associated with specific traits or diseases were enriched in enhancers.

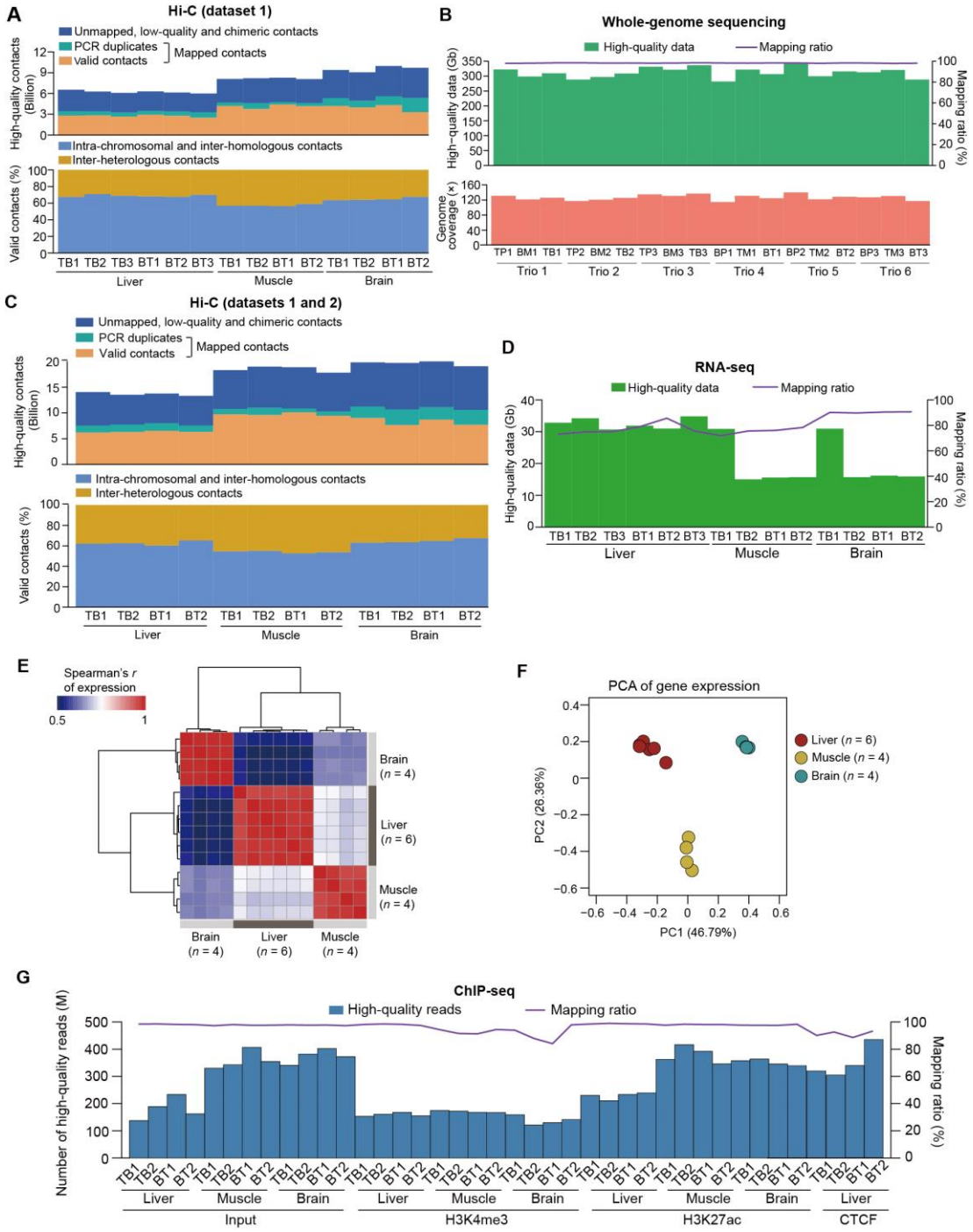
Supplemental Methods

Supplemental References

Supplemental Figures

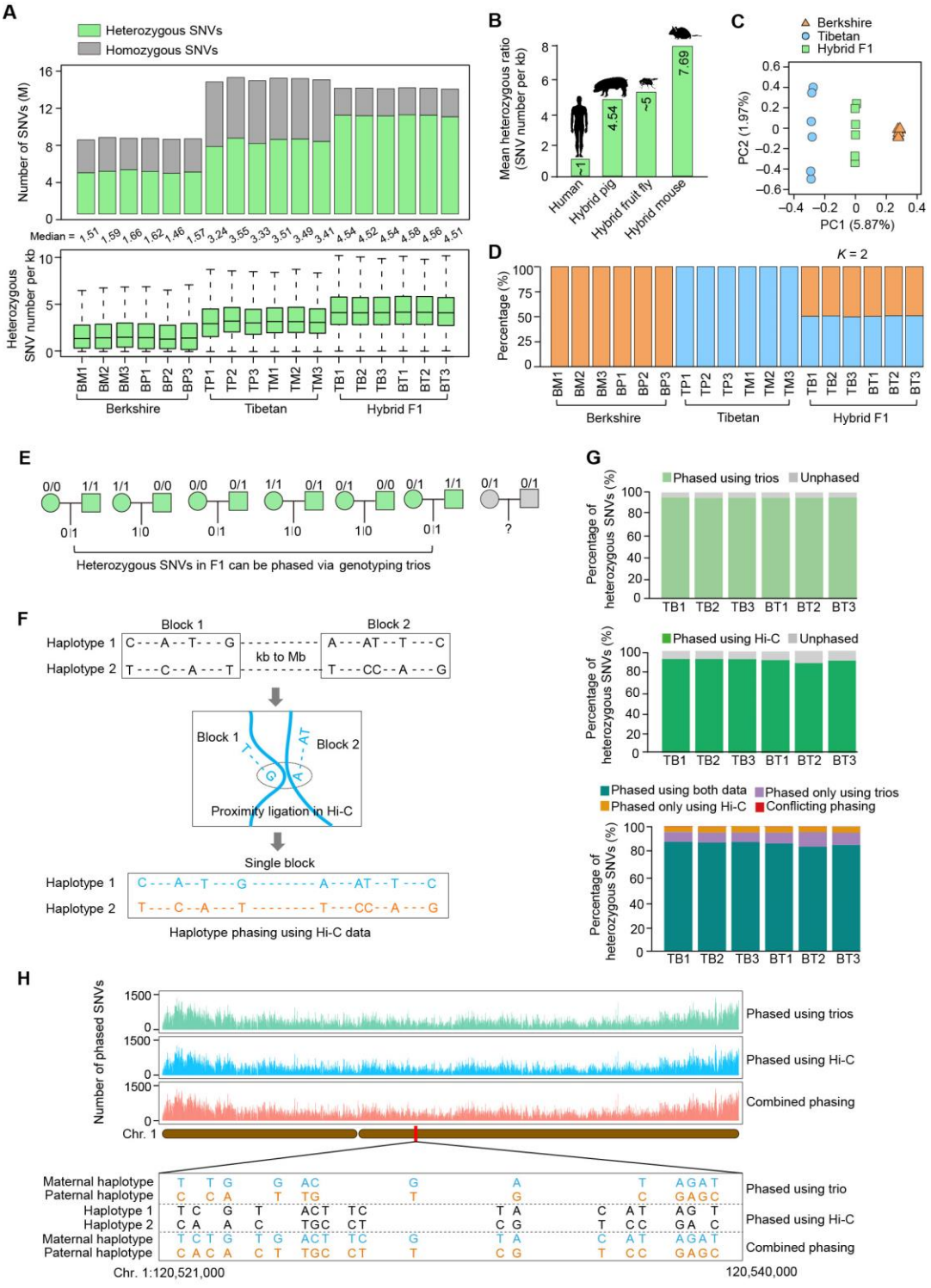


Supplemental Fig. S1. Schematic representation of the reciprocal cross pig model and summary of multi-omics data generated in this study.



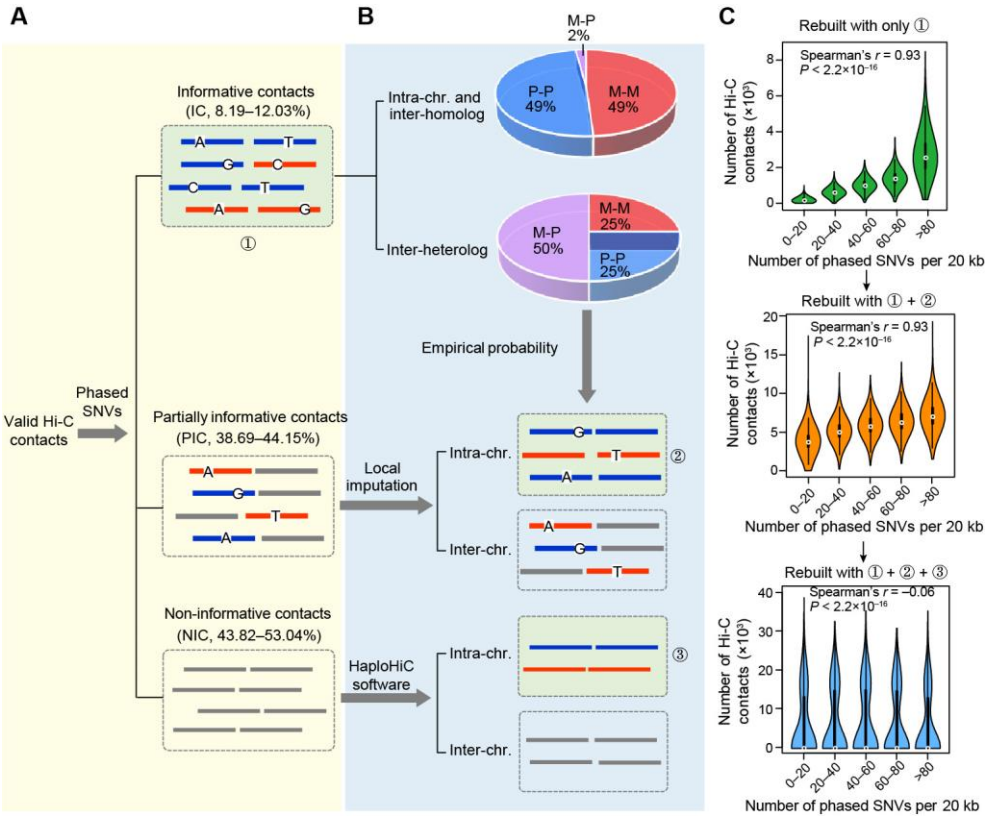
Supplemental Fig. S2. Summary of whole-genome sequencing, *in situ* Hi-C, rRNA-depleted RNA-seq and ChIP-seq data. **(A)** Composition of high-quality reads and valid contacts obtained for the *in situ* Hi-C dataset 1. The calculated average proportion of unphased intra-chromosomal (*i.e.*, occurring within chromosomes and between homologs) and inter-chromosomal (*i.e.*, between heterologs) contacts were 64.54% and 35.46% for the fourteen samples, respectively. **(B)** Summary of the whole-genome resequencing data for six parent-child trios. The high-quality data (upper, green bars), mapping ratio (purple lines) and coverage depth (~122x of the reference genome for each of the 18 individuals) (lower, pink bars) are shown. **(C)** Summary of the combined *in situ* Hi-C data (datasets 1

and 2) of twelve tissue samples. These data are described in detail in **Supplemental Fig. S1** and used for PEI identification. Valid contacts and proportions of unphased intra- and inter-chromosomal contacts are shown for each sample. **(D)** High-quality reads and mapping ratios of the RNA-seq data. **(E** and **F)** Similarity of transcriptional profiles between samples from hybrid pigs using Spearman's *r* correlation (**E**) and PCA (**F**), showing a tissue-dominant pattern. **(G)** Mapping ratios of high-quality ChIP-seq data for H3K27ac, H3K4me3 and CTCF.

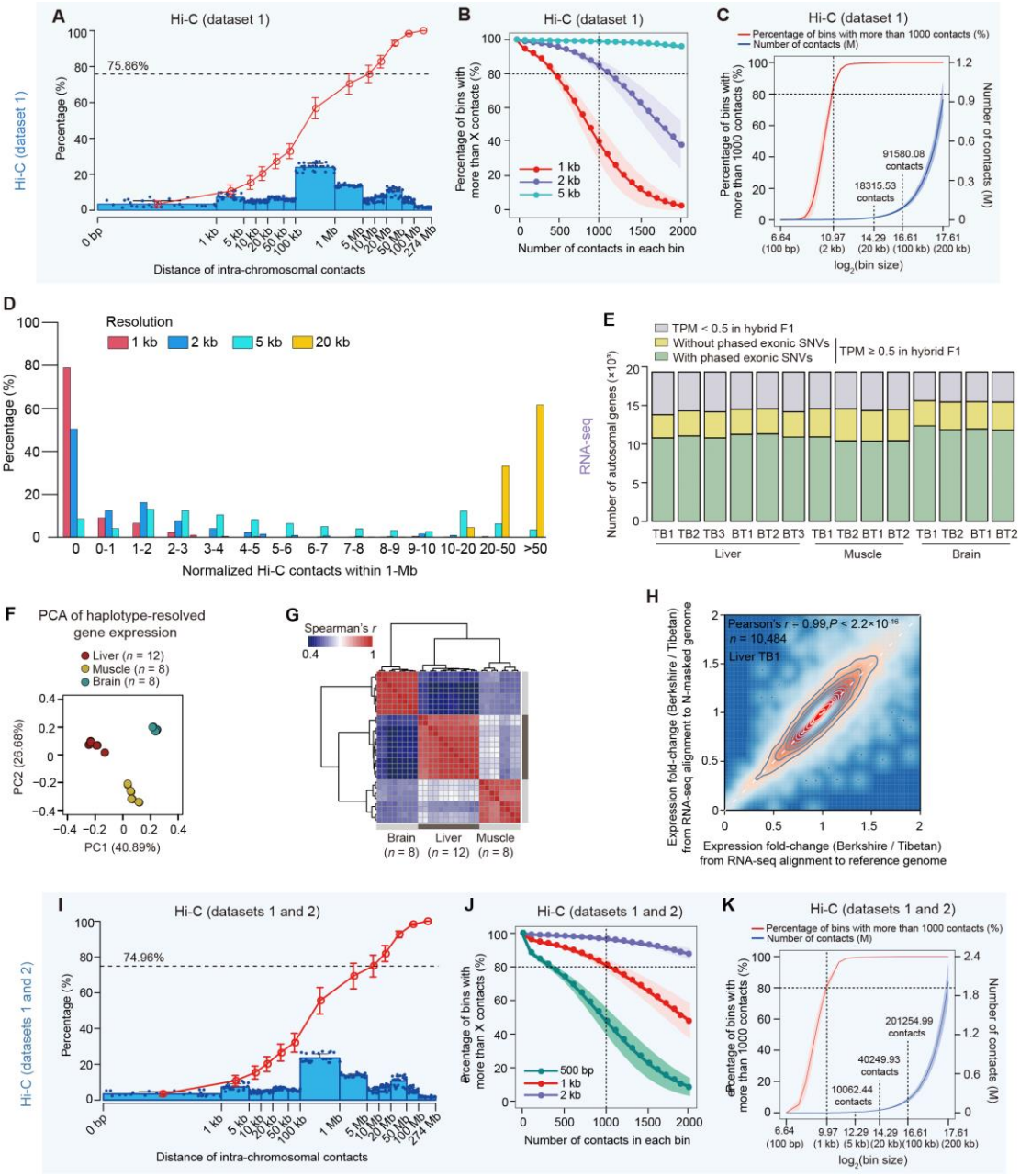


Supplemental Fig. S3. Construction of chromosome-span haplotypes for hybrid pigs using genome-wide heterozygous single nucleotide variants (SNVs). **(A)** Summary of SNVs called based on whole-genome resequencing data. The higher number of SNVs identified in Asian Tibetan pigs compared to European Berkshire pigs after mapping to the reference European Duroc pig genome (heterozygous SNVs: 8.12 M versus 4.72 M, *P* = 0.002; homozygous SNVs: 6.92 M versus 3.69 M, *P* = 0.002; heterozygous SNV ratio: 3.42 versus 1.57, *P* = 0.002; Mann-Whitney *U* test), reveals the extensive genomic divergence

between Eurasian pig breeds. **(B)** In our study, the heterozygous SNV ratio calculated in hybrid pigs (4.54 per kb) was lower than that of hybrid mice (7.70 per kb; de Wit 2017), comparable to the values obtained in hybrid fruit flies (~5 per kb; AlHaj Abed et al. 2019), and significantly higher than humans (~1 per kb; de Wit 2017). The diploid 3D genome of these three species was successfully reconstructed. **(C)** Principal component analysis (PCA) of the six parent-child trios constructed using genomic SNVs. **(D)** Genetic structure of six parent-child trios using genome-wide SNVs. The length of the colored segments represents the proportion of the genome of each individual that originates from two 'ancestral populations' (*K* = 2, two parental breeds of F1 hybrid pigs, *i.e.*, Berkshire and Tibetan). **(E)** Schematic representation of heterozygous SNV phasing based on the genotypes of trios. Heterozygous SNVs in F1 hybrids can be phased when at least one of the parents is homozygote. **(F)** Schematic representation of haplotype construction based on Hi-C contacts using the HapCUT2 algorithm (Edge et al. 2017). **(G)** Comparisons between the phased heterozygous SNVs for F1 hybrid pigs using three approaches: 1) genome sequencing data of parent-child trios alone (top), 2) Hi-C data of F1 hybrids (middle), and 3) the combination data of both (bottom). **(H)** Illustration of SNV phasing in Chromosome 1 using the three different approaches described in (**G**). Top: tracks showing the number of phased heterozygous SNVs per 100-kb genomic bin. Bottom: zoom-in of the constructed haplotypes. Parental origin can only be determined via genotyping of parent-child trios. Seed haplotypes rebuilt with the Hi-C data of F1 hybrids were arbitrarily designated as '1' and '2', and contained genetic variants that were absent in the trio-phased haplotypes. The integration of these data can generate high-quality haplotypes spanning the entire chromosome, whereby employing a combination of genome sequencing data of parent-child trios and Hi-C data of F1 hybrids improves the efficiency of chromosome-span haplotype reconstruction.

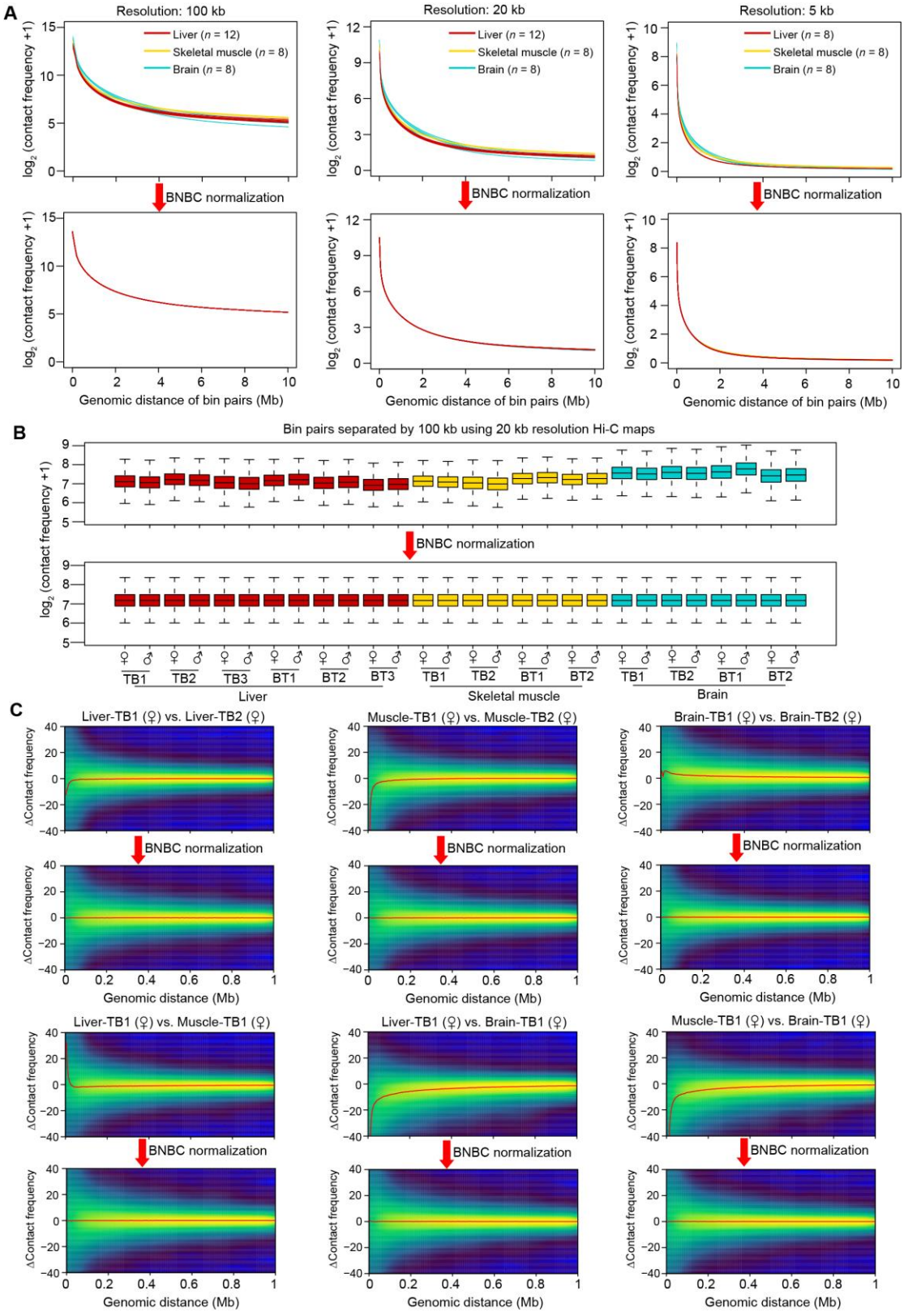


Supplemental Fig. S4. Construction of haplotype-resolved Hi-C maps for hybrid pigs using SNV phasing and a local imputation method. **(A)** Phasing of Hi-C contacts. Paired-end Hi-C reads can be classified into three categories based on haplotype-resolved SNVs: informative contacts (IC, reads containing phased SNVs at both ends, top), partially informative contacts (PIC, reads containing phased SNVs at one end, middle), and non-informative contacts (NIC, reads containing no phased SNVs, bottom). **(B)** According to the proportion of contacts occurring within maternal (intra-maternal, 'M-M') or paternal (intra-paternal, 'P-P') alleles, and between maternal and parental alleles (inter-parental, 'M-P') (top), we assigned unphased intra-chromosomal PIC and NIC to their parental origins using a local imputation strategy and the HaploHiC software (Lindsly et al. 2021), respectively (bottom). **(C)** Violin plots showing the correlation between the number of phased SNVs and the phased contact counts for phased IC (denoted as ①), IC+PIC (①+②), and IC+PIC+NIC ①+②+③), respectively, which indicates that allele assignment of all categories of Hi-C contacts can efficiently eliminate the negative effects caused by differences in SNV density across the genome.



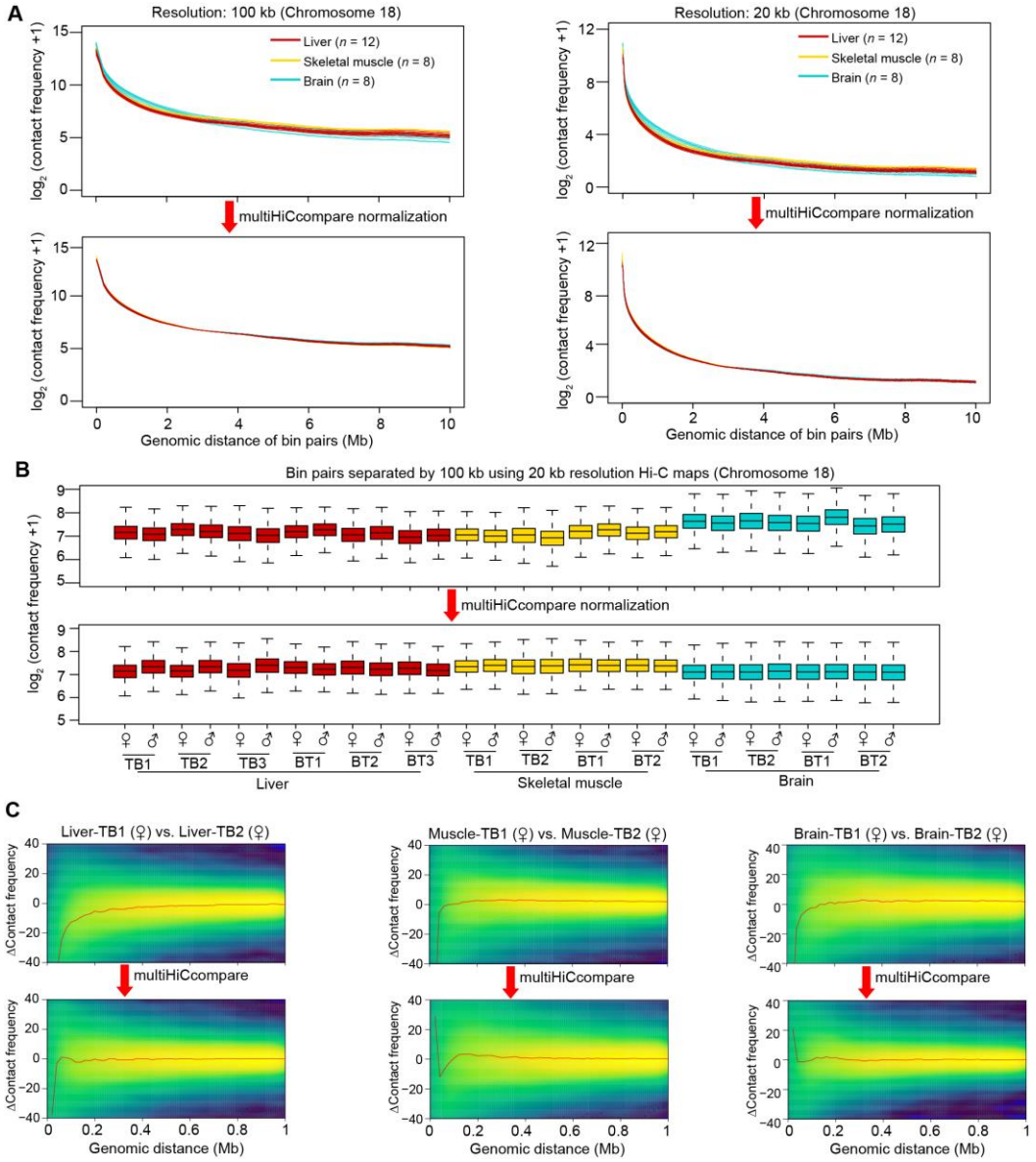
Supplemental Fig. S5. Summary of phased Hi-C and RNA-seq data. **(A)** Insert size distribution of haplotype-resolved intra-chromosomal Hi-C contacts for dataset 1. The column represents the proportion of contacts within each distance interval. The circle on the column indicates the proportion of each sample. The red curve shows the cumulative percentage. Data are represented as mean ± SD (*n* = 28). The intra-chromosomal contacts (average 75.85%, dashed lines) mainly occurred within 10 Mb. **(B** and **C)** Resolution of haplotype-resolved intra-chromosomal Hi-C maps at different bin sizes. The Hi-C map resolution was defined as the smallest bin size where 80% of bins have at least 1000 reads to allow for reliable discerning of local features (Rao et al. 2014). Data are shown as mean ± SD (*n* = 28). The mean number of contacts at resolution of 20 kb and 100 kb are shown in (**C**). **(D)** The different percentages of Hi-C contacts within 1 Mb genomic distance for the

resolution of 1 kb, 2 kb, 5 kb and 20 kb. **(E)** The number of autosomal genes with phased transcription in haplotypes (*n* = 28). **(F)** PCA of allelic transcription for haplotypes (*n* = 28), showing a tissue dominant pattern. **(G)** Similarity of allelic transcriptional profiles between haplotypes (*n* = 28) from hybrid pigs using Spearman's *r* correlation, showing a tissue-dominant pattern. **(H)** Evaluation of reference bias using RNA-seq from a liver sample of TB1. The x-axis represents the fold-change between breeds (Berkshire / Tibetan) obtained from the alignment against reference genome using Allelome.PRO (Andergassen et al. 2015) and the y-axis represents the fold-change obtained from the alignment against the N-masked genome using SNPsplit (Krueger and Andrews 2016). (**I**–**K**) Insert size distribution (**I**) and resolution estimation (**J** and **K**) of phased intra-chromosomal Hi-C contacts for datasets 1 and 2. Data are shown as mean ± SD (*n* = 24). The mean number of contacts at resolution of 5 kb, 20 kb and 100 kb are shown in (**K**).

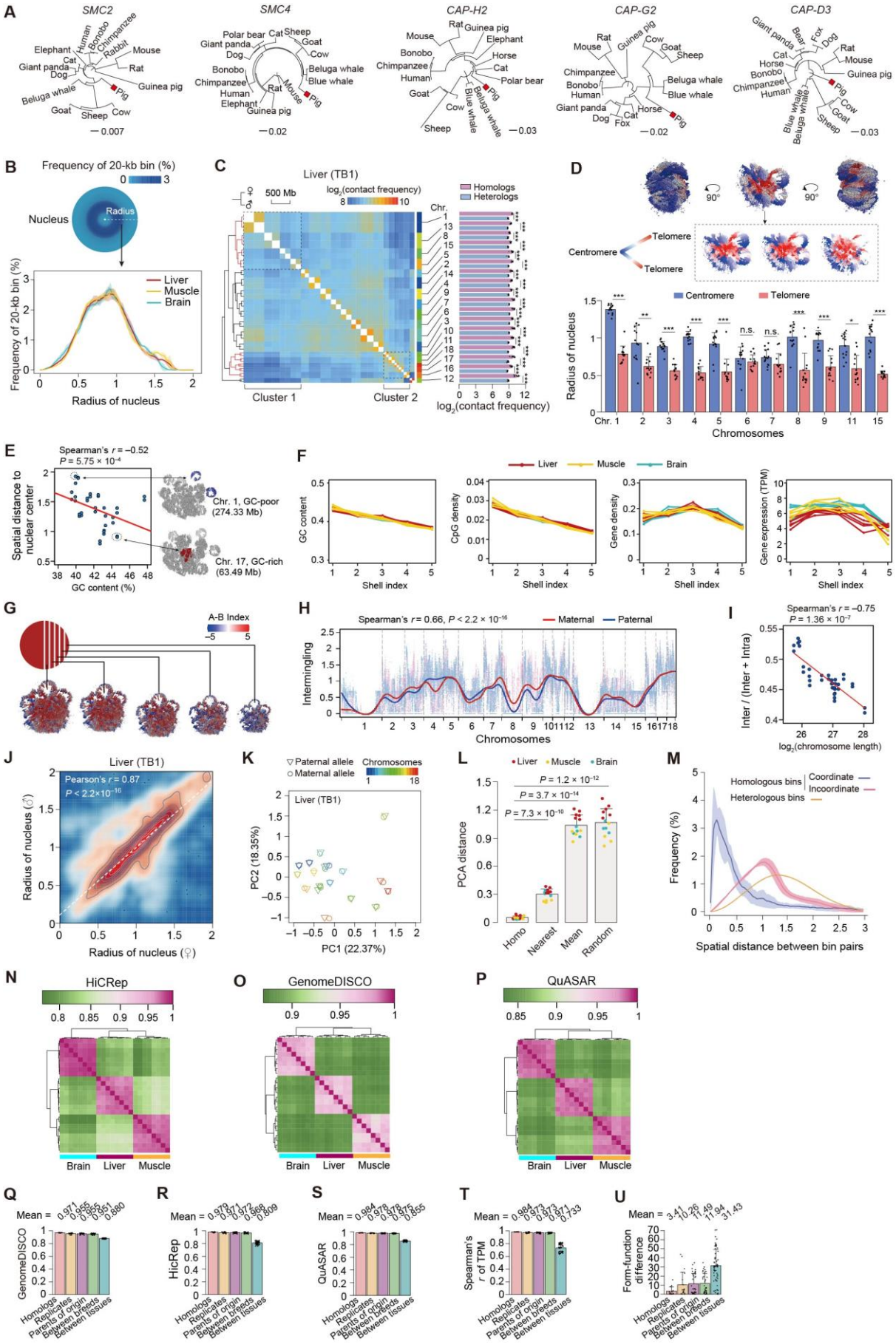


Supplemental Fig. S6. Assessment of BNBC quantile normalization. **(A)** Distance-dependent decay of all Hi-C maps before (top) and after (bottom) BNBC normalization. From left to right: 100-kb, 20-kb and 5-kb resolution. **(B)** Boxplots showing the distribution of contact frequency at 100 kb bridging distance using 20-kb resolution Hi-C maps before

(top) and after (bottom) BNBC normalization. **(C)** Difference between haplotypes in contact frequency across different genomic ranges within each tissue and between tissues before (top) and after (bottom) BNBC normalization. The red lines represent the median values at variable bridging distances.

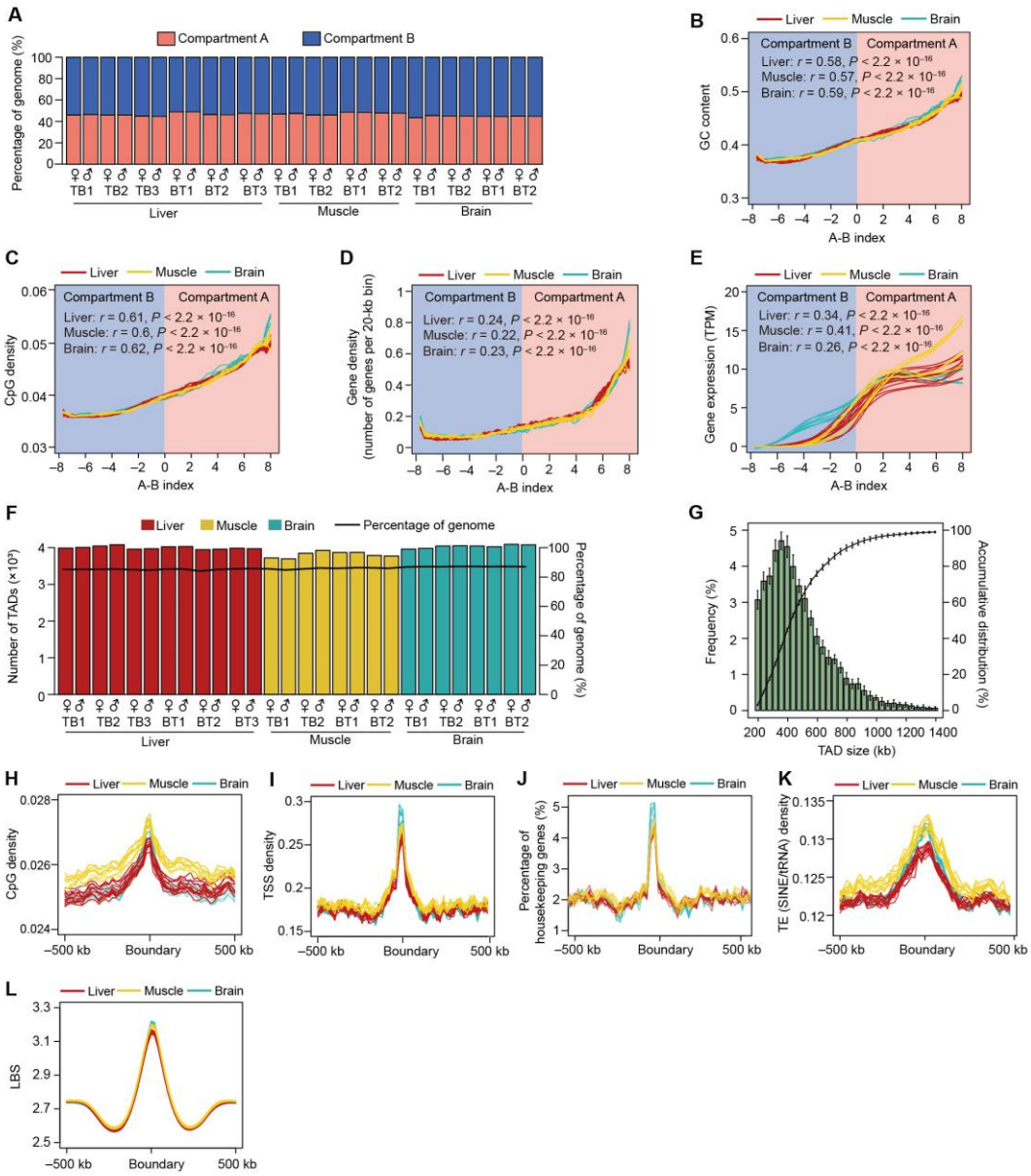


Supplemental Fig. S7. Assessment of Hi-C normalization using multiHiCcompare. **(A)** Distance-dependent decay of Hi-C maps before (top) and after (bottom) normalization. From left to right: 100-kb and 20-kb resolution for Chromosome 18. **(B)** Boxplots showing the distribution of contact frequencies at 100 kb bridging distance using 20-kb resolution Hi-C maps before (top) and after (bottom) normalization of Chromosome 18. **(C)** Difference in contact frequencies across different genomic ranges between haplotypes within each tissue before (top) and after (bottom) normalization of Chromosome 18. Red lines represent the median value for each genomic distance.


A
SMC2
SMC4
CAP-H2
CAP-G2
CAP-D3
B
Frequency of 20-kb bin (%)
Nucleus
Radius
Liver
Muscle
Brain
Frequency of 20-kb bin (%)
Radius of nucleus
C
Liver (TB1)
500 Mb
log2(contact frequency)
Homologs
Heterologs
Cluster 1
Cluster 2
log2(contact frequency)
D
Centromere
Telomere
Telomere
Centromere
Telomere
Radius of nucleus
Chromosomes
E
Spearman's r = −0.52
P = 5.75 × 10−4
Spatial distance to nuclear center
GC content (%)
Chr. 1, GC-poor (274.33 Mb)
Chr. 17, GC-rich (63.49 Mb)
F
Liver
Muscle
Brain
GC content
CpG density
Gene density
Gene expression (TPM)
Shell index
G
A-B Index
H
Spearman's r = 0.66, P < 2.2 × 10−16
Maternal
Paternal
Intermingling
Chromosomes
I
Spearman's r = −0.75
P = 1.36 × 10−7
Inter / (Inter + Intra)
log2(chromosome length)
J
Liver (TB1)
Pearson's r = 0.87
P < 2.2×10−16
Radius of nucleus (♂)
Radius of nucleus (♀)
K
Paternal allele
Maternal allele
Chromosomes
Liver (TB1)
PC2 (18.35%)
PC1 (22.37%)
L
Liver
Muscle
Brain
P = 1.2 × 10−12
P = 3.7 × 10−14
P = 7.3 × 10−10
PCA distance
Homo
Nearest
Mean
Random
M
Homologous bins
Heterologous bins
Coordinate
Incoordinate
Frequency (%)
Spatial distance between bin pairs
N
HiCRep
Brain
Liver
Muscle
O
GenomeDISCO
Brain
Liver
Muscle
P
QuASAR
Brain
Liver
Muscle
Q
Mean = 0.971 0.965 0.955 0.951 0.880
GenomeDISCO
R
Mean = 0.979 0.971 0.972 0.968 0.909
HicRep
S
Mean = 0.984 0.978 0.978 0.975 0.956
QuASAR
T
Mean = 0.984 0.973 0.973 0.971 0.733
Spearman's r of TPM
U
Mean = 3.41 10.26 11.49 11.94 31.43
Form-function difference
Homologs
Replicates
Parents of origin
Between breeds
Between tissues

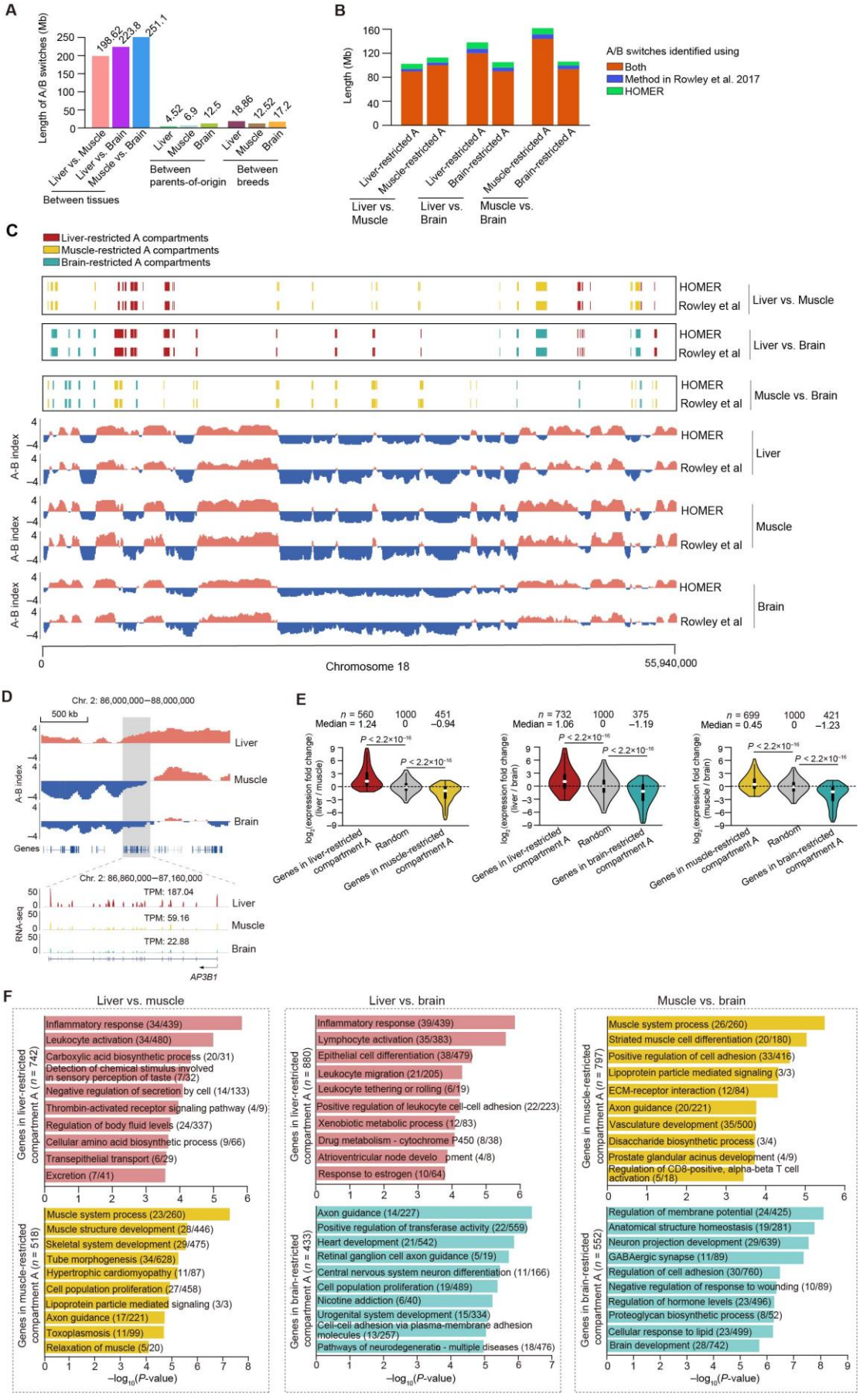
Supplemental Fig. S8. Initial characteristics of 3D nuclear organization in the diploid pig genome. **(A)** The pig genome contains full sets of functional condensin II subunits, including *SMC2* (ENSSSCG00000005403), *SMC4* (ENSSSCG00000011731), *CAP-H2* (ENSSSCG00 000000965), *CAP-G2* (ENSSSCG00000028169), and *CAP-D3* (ENSSSCG0 0000022192). The topological relationships of each of the five condensin II subunits (Hoencamp et al. 2021) across mammals were constructed based on protein sequences using the multiple sequence alignment program Clustal Omega tool (v.1.2.1). The phylogenetic relationships across species of emphasized condensin II subunits follow the evolutionary distance to the pig lineage (red). (**B**) Uneven distribution of mass in the 3D nucleus. The nucleus is divided into 100 shells of equal radius from the center to the periphery in order to determine the distribution of 20-kb genomic bins. Colored lines represent mean values across tissues, and shading around the mean shows dispersion calculated using the standard deviation divided by the cumulative sum of all means. (**C**) Inter-chromosomal interaction profiles (18×18 matrix consisting of the homologs of each autosome) show chromosome territories (dotted boxes indicate the two clusters of chromosomes) and preferential interactions between homologs compared to that between heterologs (Wilcoxon rank-sum test, n.s., $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$). (**D**) Quantification of centromere and telomere organization. 3D genome structure and its intersecting regions in a representative liver sample. Most centromeres had a larger radius in the nucleus than telomeres (i.e., centromeres were localized to outer nuclear regions relative to telomeres). Only autosomes with recognizable centromeres (Chromosomes 1-9, 11 and 15) were compared. Data are presented as mean values ± SD ($n = 14$). Wilcoxon rank-sum test, n.s., $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$. **(E)** The radial distances of chromosomes to the nuclear center (calculated as the median radial distance of all bins in a chromosome) negatively correlates with their GC content (mean Spearman's $r = –0.52$ for 14 diploid genomes, $P < 0.01$; left). Examples of preferential localization in the nucleus for two chromosomes (right). GC-rich Chromosome 17 (GC content = 44.56%) preferentially localized to the nuclear interior, whereas GC-poor Chromosome 1 (GC content = 40.20%) was consistently observed on the nuclear periphery. The 3D genome structure of the representative liver sample is shown. **(F)** Sequence-dependent spatial organization of the chromatin. The nucleus is divided into five shells with equal nucleotides from the center to the periphery (indexed as 1 to 5), respectively. Generally, GC-rich, gene rich, transcript-abundant chromatins were more likely to inhabit the interior of the nucleus, whereas GC-poor, transcript-sparse chromatins, were preferentially located in the nuclear periphery. **(G)** Compartments A/B were aggregated in the 3D space of the pig genome of the representative liver. The plot was visualized in quintuplicate, with five intersecting sections plotted from the interior regions of the nucleus (left) to the periphery (right) based on the distance (schematically depicted in the upper-left inset). Color bars indicate the

value of the A-B index. **(H)** Highly similar multi-chromosome intermingling (mean Spearman's r = 0.66 for 14 diploid genomes, $P < 2.2\times10^{-16}$) between maternal (red) and paternal (blue) alleles at 20-kb resolution. **(I)** Ratio of inter-chromosomal contacts are negatively correlated with chromosome length (mean Spearman's r = –0.67 for 14 diploid genomes, $P < 0.004$). **(J)** Homologs exhibited highly similar distances to the center of 3D nucleus (Pearson's r = 0.87, $P < 2.2\times10^{-16}$). The center of the 3D nucleus was calculated as the mean coordinate of all autosomes. **(K)** PCA analysis of autosomal interactions in a representative liver sample in which homologs that are closer in the plot have more similar interaction patterns. **(L)** Distance between homologs is almost always shortest between potential pairs in the 2D PCA projection. Homo, homologs; Nearest, an autosome and its nearest heterolog; Mean, a chromosome and all the other autosomes; Random, two randomly picked autosomes. Data are presented as means ± SD (n = 14). P values are from paired Student's t-test. **(M)** Spatial distance between bin pairs of homologs and heterologs. Coordinated bin pairs of homologs (purple) were spatially closer than non-coordinated bin pairs of homologs (pink) or heterologs (yellow). Lines indicate means and shadows show ± SD. (**N**–**P**) The correlation between haplotype-resolved Hi-C maps using HiCRep (**N**), GenomeDISCO (**O**) and QuASAR (**P**). **(Q**–**U)** Haplotype-resolved interrogation of similarities in chromatin architecture, gene expression and histone modifications among 14 diploid genomes. The correlations of chromatin architecture were separately determined using GenomeDISCO (**Q**), HiCRep (**R**), and QuASAR (**S**) for the Hi-C maps; The correlations of gene expression (**T**) and the combined differences in chromatin architecture (*i.e.*, form) and gene expression (*i.e.*, function) (**U**) among haplotypes are also shown. These correlation rates were grouped into the following five categories: between homologs, among biological replicates, between parent-of-origins, between parental breeds, among tissues. Data are shown as mean ± SD.

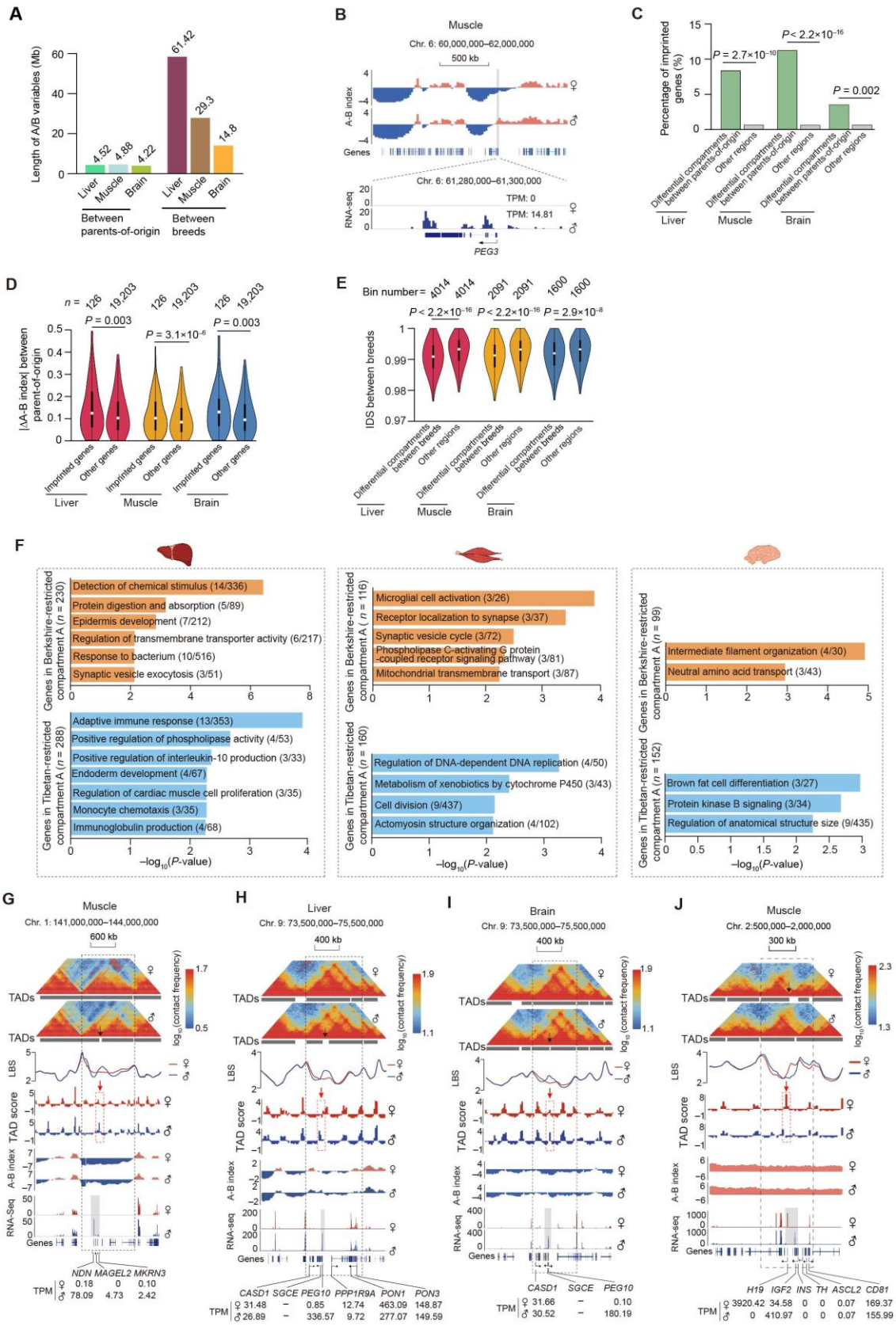


Supplemental Fig. S9. Genomic features of compartments and TADs in the haplotype-resolved pig genome. **(A)** Length proportions of A/B compartments among 28 haploid genomes. On average, 45.71% of the genome (~1035.7 Mb) could be recognized as accessible A compartments, while the remaining of the genome was categorized as less accessible B compartments (52.59% of the genome or ~1191.56 Mb). **(B–E)** GC content (**B**), CpG density (**C**), gene density (**D**), and gene expression (**E**) of A/B compartments with different activities. According to the A-B index, A and B compartments were separately divided into 50 equal parts across each of the three tissues (x-axis). Compared to compartment B, compartment A regions contain a higher GC content (45.19% vs. 39.08%), are more gene-rich (13.62 vs. 4.38 genes per Mb), and have relatively higher transcriptional activity (median TPM 8.18 vs. 0.20). Spearman's correlation between A-B index and the genomic features in each tissue are shown. **(F)** Numbers and genome

coverage of TADs across 28 haploid genomes. TADs occupy ~85.91% of the genome. **(G)** Size distribution of TADs. TAD median size was estimated at ~440 kb. **(H–L)** Genomic features around TAD boundaries (1 Mb centered on the TAD). Compared to other regions, TAD boundaries exhibited relatively higher densities of CpG (**H**) and transcription start site (TSS) (**I**), an increased ratio of house-keeping genes (**J**), larger proportion of short interspersed element-transposable RNA (SINE/tRNA) (**K**), and higher local boundary scores (LBS; low LBS means high density of local interactions) (**L**). These features were consistent with previous reports in mammals (Dixon et al. 2012), confirming the robustness of the TADs in the pig genome.

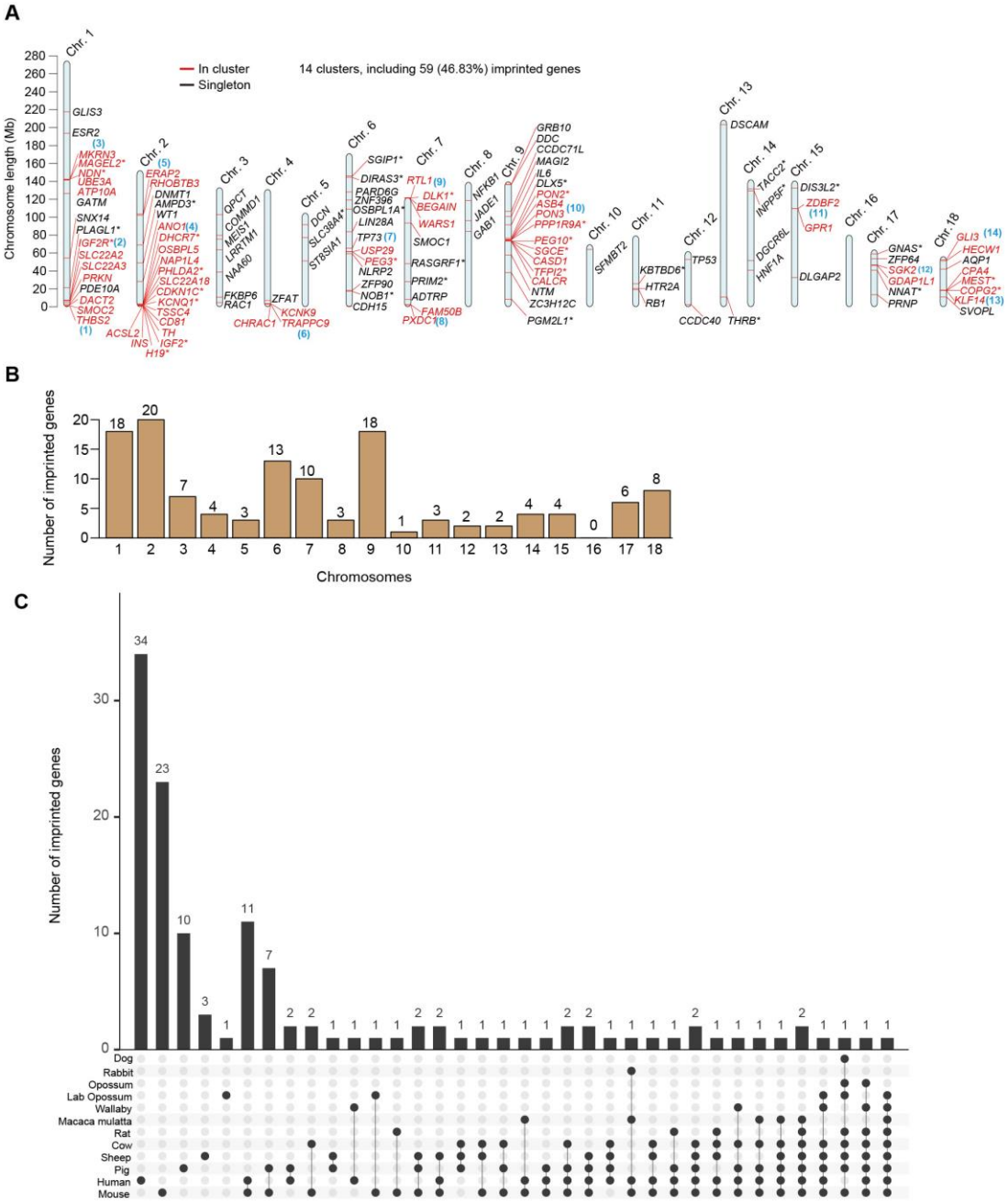

A
Length of A/B switches (Mb)
198.62
223.8
251.1
4.52
6.9
12.5
18.86
12.52
17.2
Liver vs. Muscle
Liver vs. Brain
Muscle vs. Brain
Between tissues
Liver
Muscle
Brain
Between parents-of-origin
Between breeds
B
Length (Mb)
A/B switches identified using
Both
Method in Rowley et al. 2017
HOMER
Liver-restricted A
Muscle-restricted A
Liver-restricted A
Brain-restricted A
Muscle-restricted A
Brain-restricted A
Liver vs. Muscle
Liver vs. Brain
Muscle vs. Brain
C
Liver-restricted A compartments
Muscle-restricted A compartments
Brain-restricted A compartments
HOMER
Rowley et al
Liver vs. Muscle
Liver vs. Brain
Muscle vs. Brain
A-B index
Liver
Muscle
Brain
Chromosome 18
0
55,940,000
D
Chr. 2: 86,000,000–88,000,000
500 kb
Liver
Muscle
Brain
Genes
Chr. 2: 86,860,000–87,160,000
RNA-seq
TPM: 187.04
TPM: 59.16
TPM: 22.88
AP3B1
E
n = 560
Median = 1.24
1000
0
451
−0.94
n = 732
Median = 1.06
1000
0
375
−1.19
n = 699
Median = 0.45
1000
0
421
−1.23
P < 2.2×10−16
log2(expression fold change) (liver / muscle)
log2(expression fold change) (liver / brain)
log2(expression fold change) (muscle / brain)
Genes in liver-restricted compartment A
Random
Genes in muscle-restricted compartment A
Genes in brain-restricted compartment A
F
Liver vs. muscle
Liver vs. brain
Muscle vs. brain
Genes in liver-restricted compartment A (n = 742)
Inflammatory response (34/439)
Leukocyte activation (34/480)
Carboxylic acid biosynthetic process (20/31)
Detection of chemical stimulus involved in sensory perception of taste (7/32)
Negative regulation of secretion by cell (14/133)
Thrombin-activated receptor signaling pathway (4/9)
Regulation of body fluid levels (24/337)
Cellular amino acid biosynthetic process (9/66)
Transepithelial transport (6/29)
Excretion (7/41)
Genes in muscle-restricted compartment A (n = 518)
Muscle system process (23/260)
Muscle structure development (28/446)
Skeletal system development (29/475)
Tube morphogenesis (34/628)
Hypertrophic cardiomyopathy (11/87)
Cell population proliferation (27/458)
Lipoprotein particle mediated signaling (3/3)
Axon guidance (17/221)
Toxoplasmosis (11/99)
Relaxation of muscle (5/20)
−log10(P-value)
Genes in liver-restricted compartment A (n = 880)
Inflammatory response (39/439)
Lymphocyte activation (35/383)
Epithelial cell differentiation (38/479)
Leukocyte migration (21/205)
Leukocyte tethering or rolling (6/19)
Positive regulation of leukocyte cell-cell adhesion (22/223)
Xenobiotic metabolic process (12/83)
Drug metabolism - cytochrome P450 (8/38)
Atrioventricular node development (4/8)
Response to estrogen (10/64)
Genes in brain-restricted compartment A (n = 433)
Axon guidance (14/227)
Positive regulation of transferase activity (22/559)
Heart development (21/542)
Retinal ganglion cell axon guidance (5/19)
Central nervous system neuron differentiation (11/166)
Cell population proliferation (19/489)
Nicotine addiction (6/40)
Urogenital system development (15/334)
Cell-cell adhesion via plasma-membrane adhesion molecules (13/257)
Pathways of neurodegeneration - multiple diseases (18/476)
−log10(P-value)
Genes in muscle-restricted compartment A (n = 797)
Muscle system process (26/260)
Striated muscle cell differentiation (20/180)
Positive regulation of cell adhesion (33/416)
Lipoprotein particle mediated signaling (3/3)
ECM-receptor interaction (12/84)
Axon guidance (20/221)
Vasculature development (35/500)
Disaccharide biosynthetic process (3/4)
Prostate glandular acinus development (4/9)
Regulation of CD8-positive, alpha-beta T cell activation (5/18)
Genes in brain-restricted compartment A (n = 552)
Regulation of membrane potential (24/425)
Anatomical structure homeostasis (19/281)
Neuron projection development (29/639)
GABAergic synapse (11/89)
Regulation of cell adhesion (30/760)
Negative regulation of response to wounding (10/89)
Regulation of hormone levels (23/496)
Proteoglycan biosynthetic process (8/52)
Cellular response to lipid (23/499)
Brain development (28/742)
−log10(P-value)

Supplemental Fig. S10. Allelically compartmental rearrangement between tissues. **(A)** Summary of A/B switches. **(B** and **C)** Comparision of A/B switches identified between tissues using the method described in (Rowley et al. 2017) and HOMER (Lin et al. 2012) for the whole genome (**B**) and Chromosome 18 (**C**). **(D)** A typical example of A/B switches between tissues containing *AP3B1*, a functional gene participating in inflammatory response. **(E)** Expression changes in genes that were located in compartment-switching regions between tissues. Compared to randomly selected genes, the genes located in tissue-specific A compartments showed significantly increased expression. The *P*-values shown are from the Wilcoxon rank-sum test. **(F)** Top ten significantly enriched GO-BP terms or KEGG pathways for genes located in tissue-restricted compartment A regions from pairwise comparisons of the three tissues. Gene enrichment analysis was performed using the software Metascape (Zhou et al. 2019).

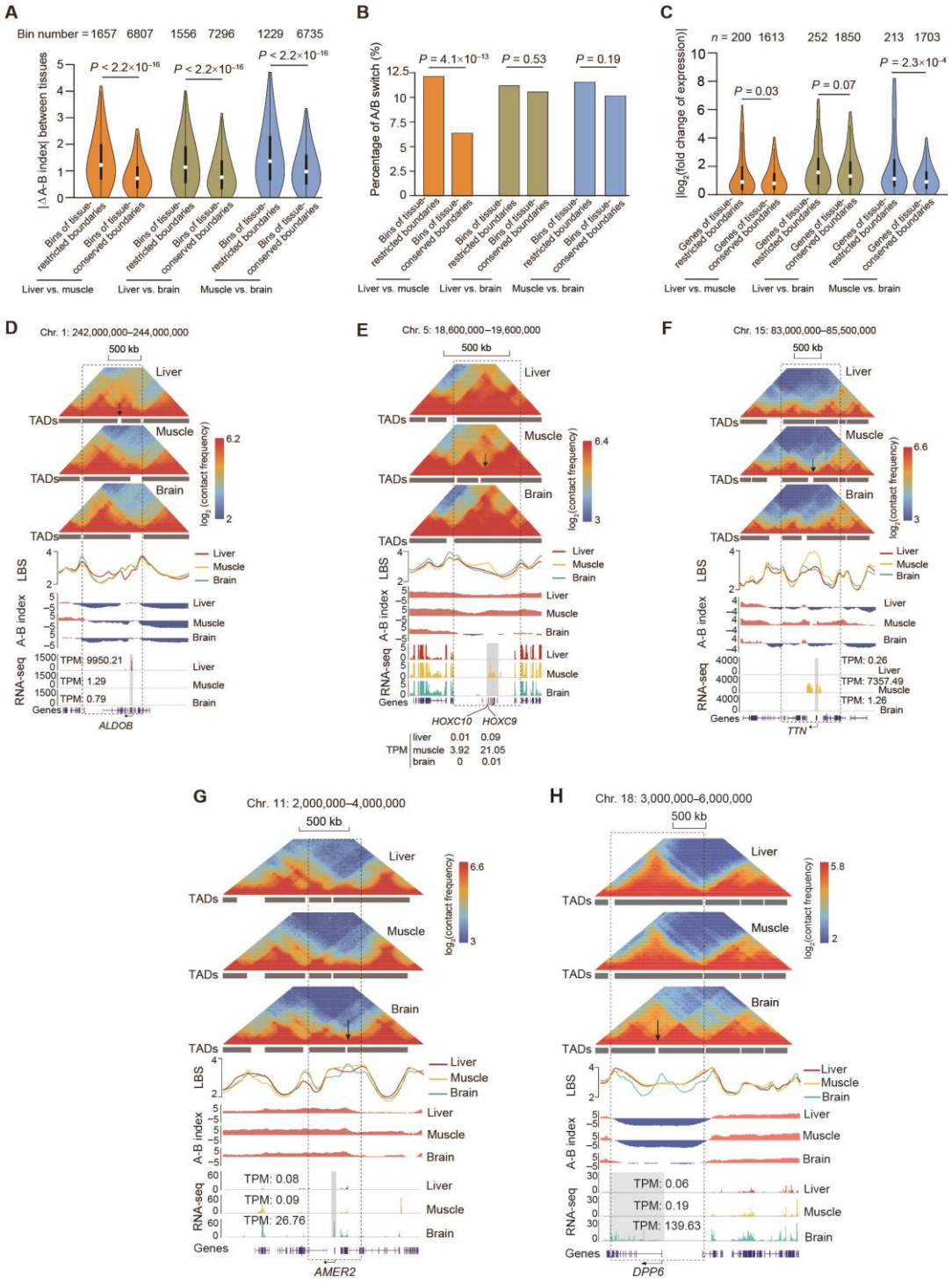


Supplemental Fig. S11. Allelically compartment rearrangement and shifted TAD boundaries. **(A)** Summary of A/B variables between parents-of-origin and between breeds. **(B)** An example of A/B switches between parent-of-origins containing a paternally imprinted

gene *PEG3*. **(C)** Known imprinted genes were statistically enriched in regions that show allelically differential compartmentalization (including A/B switch and A/B variable) between parent-of-origins across the three tissues. The *P* values were calculated using a χ^2 test. **(D)** Regions with known imprinted genes have more variable compartment scores between alleles. Median |ΔA-B index| = 0.13, $P < 0.003$, Wilcoxon rank-sum test. **(E)** Regions showing allelically differential compartmentalization (including A/B switch and A/B variable) exhibited the lowest pairwise haplotype similarities between the two parental breeds measured using the identity score (IDS; see Supplemental Methods). The *P* values were calculated using a Wilcoxon rank sum test. **(F)** Significantly enriched GO-BP terms or KEGG pathways (performed using Metascape) for genes located in regions with differential compartmentalization (*i.e.*, A/B switches and A/B variables) between the two parental alleles across the three tissues. **(G–J)** Parental-restricted TAD boundaries are adjacent to imprinted genes. A paternal-specific boundary in the muscle (**G**) influencing three imprinted genes (*NDN*, *MAGEL2* and *MKRN3)*; a paternal-specific boundary in the liver (**H**) and brain (**I**) influencing two imprinted genes (*CASD1* and *SGCE)*; a maternal-specific boundary in the muscle (**J**) influencing seven imprinted genes (*H19*, *IGF2*, *INS*, *TH*, *ASCL2*, *CD81* and *TSSC4*). Changes in TAD scores which were calculated using the TADCompare (https://github.com/dozmorovlab/TADCompare) further validated these TAD boundary shifts between parents-of-origin.

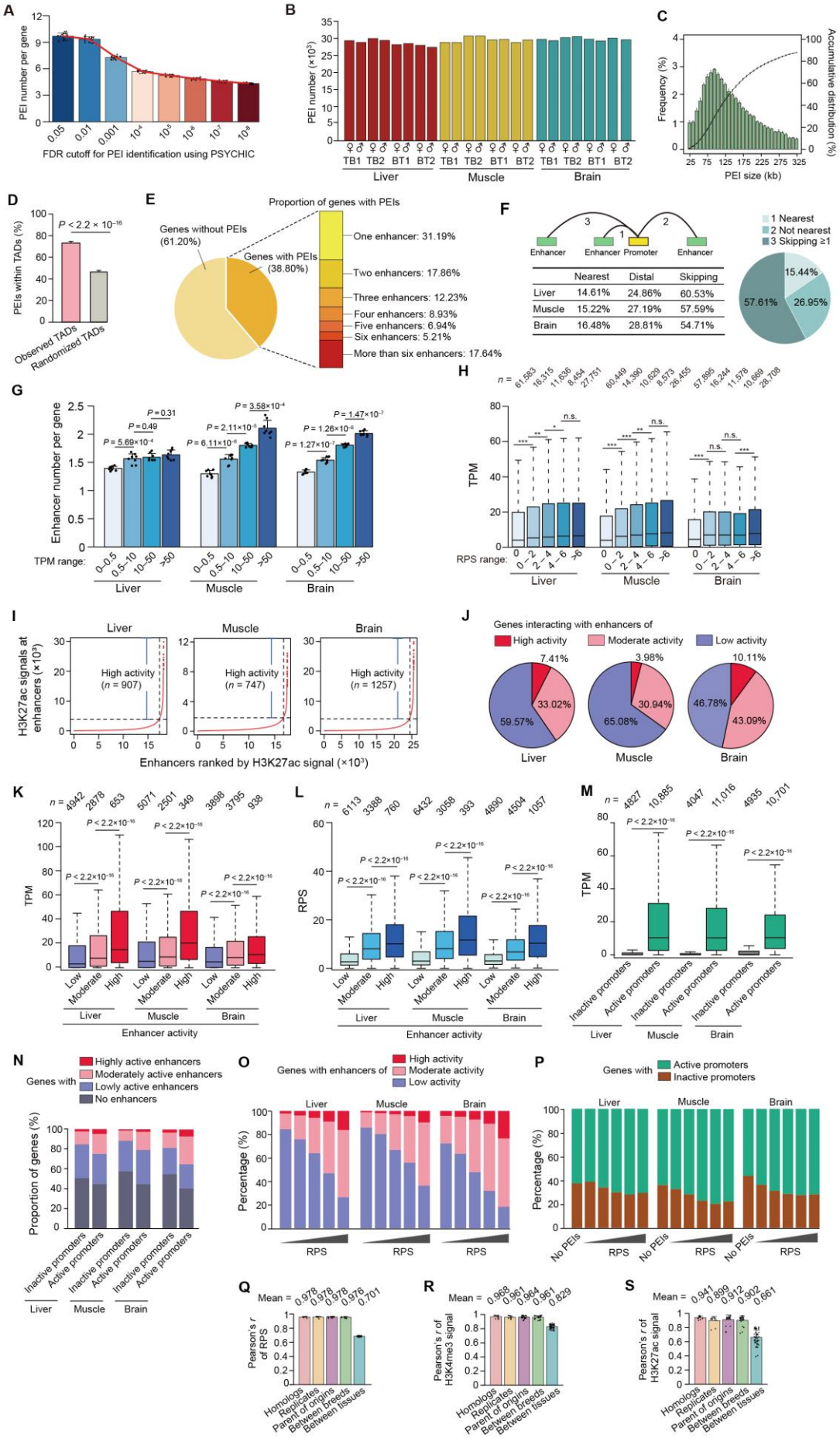


Supplemental Fig. S12. Collection of known imprinted genes. (**A**) Chromosomal organization of the 126 known imprinted genes in the reference pig genome, among which, 36 were previously identified in pigs (labeled with '*') and 90 were retrieved from the orthologs in another 11 mammal species. Members in each imprinted gene cluster are labelled in red, and numbers of genes in each cluster are indicated in blue in the parentheses. (**B**) Number of known imprinted genes located in each chromosome of the pig genome. (**C**) Number of collected imprinted genes shared by different mammal species.

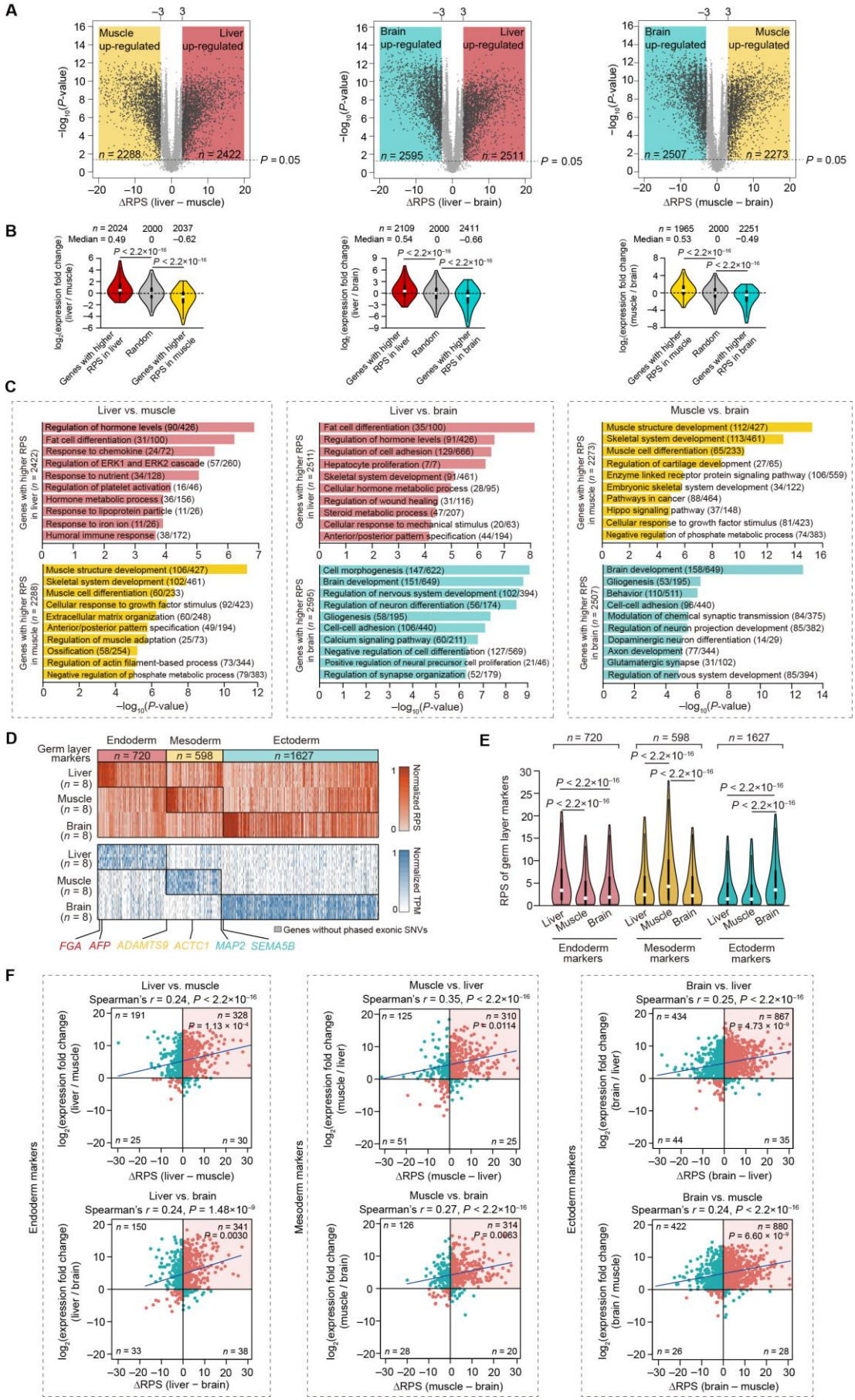


Supplemental Fig. S13. Allelically shifted TAD boundaries between tissues. **(A)** Tissue-restricted TAD boundaries have significantly higher differences in compartmentalization compared to tissue-shared TAD boundaries. *P* values were calculated using a Wilcoxon rank-sum test. **(B)** Tissue-restricted TAD boundaries are enriched in A/B switched regions compared to tissue-shared TAD boundaries. *P* values were calculated using a χ^2 test. **(C)** Compared to genes located close to tissue-shared TAD bounaries (± 60 kb from the anchor bin of TAD boundary), genes nearby tissue-restricted TAD boundaries exhibit a subtle but statistically significant increase in expression. *P* values were calculated using a Wilcoxon

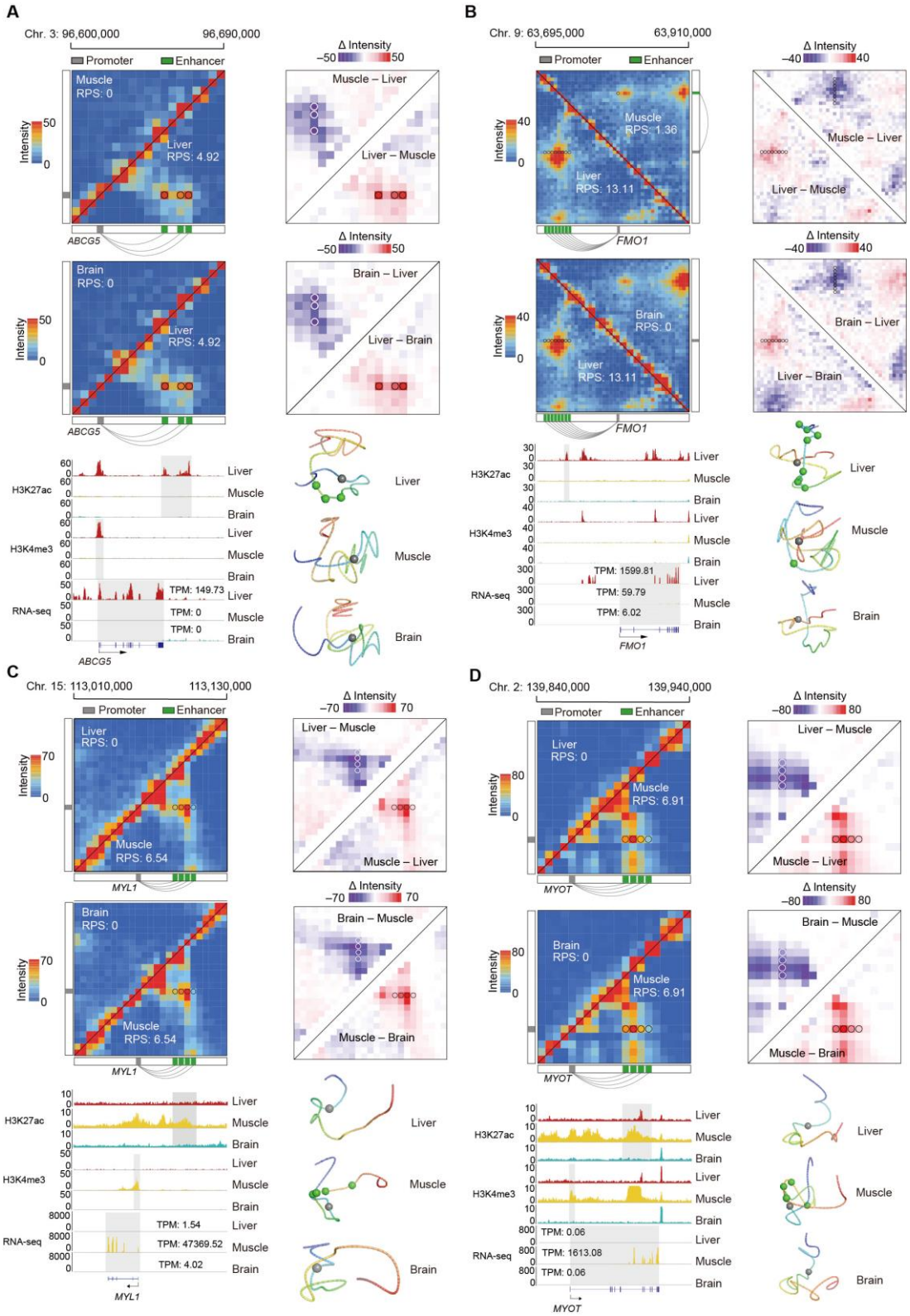
rank-sum test. **(D**–**H)** Examples of tissue-restricted TAD boundaries being adjacent to typical marker genes of relative tissues, including *ALDOB* (**D**) in liver, *HOXC9* (**E**), *-10* (**E**) and *TTN* (**F**) in muscle, and *AMER2* (**G**) and *DPP6* (**H**) in brain tissues. From top to bottom are shown the Hi-C maps, TAD boundaries, local boundary scores (LBS), the A-B index, structures, and expression (TPM) of the different genes. Black dashed boxes indicate the shifted TAD boundaries.

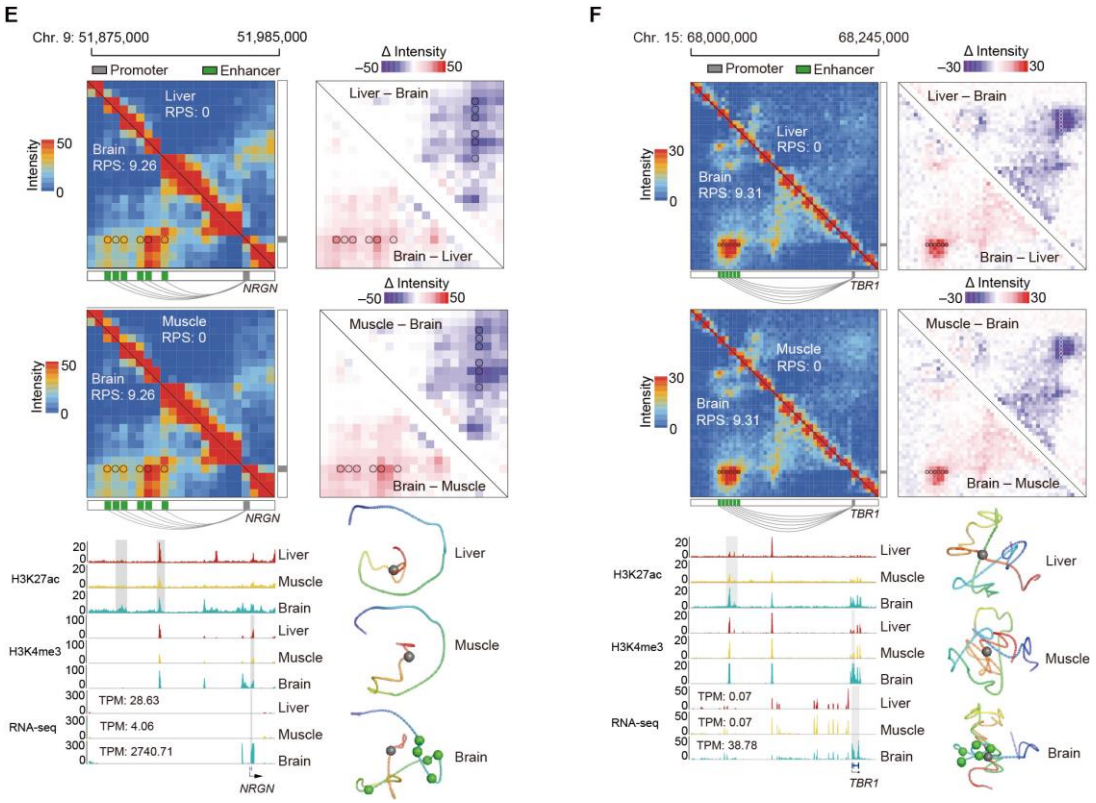

A
PEI number per gene
FDR cutoff for PEI identification using PSYCHIC
B
PEI number (×10³)
TB1 TB2 BT1 BT2
Liver
Muscle
Brain
C
Frequency (%)
Accumulative distribution (%)
PEI size (kb)
D
PEIs within TADs (%)
P < 2.2 × 10⁻¹⁶
Observed TADs
Randomized TADs
E
Proportion of genes with PEIs
Genes without PEIs (61.20%)
Genes with PEIs (38.80%)
One enhancer: 31.19%
Two enhancers: 17.86%
Three enhancers: 12.23%
Four enhancers: 8.93%
Five enhancers: 6.94%
Six enhancers: 5.21%
More than six enhancers: 17.64%
F
Enhancer
Enhancer
Promoter
Enhancer
1 Nearest
2 Not nearest
3 Skipping ≥1
15.44%
26.95%
57.61%
Nearest Distal Skipping
Liver 14.61% 24.86% 60.53%
Muscle 15.22% 27.19% 57.59%
Brain 16.48% 28.81% 54.71%
G
Enhancer number per gene
TPM range:
Liver
Muscle
Brain
H
TPM
RPS range:
Liver
Muscle
Brain
I
Liver
Muscle
Brain
H3K27ac signals at enhancers (×10³)
High activity (n = 907)
High activity (n = 747)
High activity (n = 1257)
Enhancers ranked by H3K27ac signal (×10³)
J
Genes interacting with enhancers of
High activity
Moderate activity
Low activity
7.41%
33.02%
59.57%
3.98%
30.94%
65.08%
10.11%
43.09%
46.78%
Liver
Muscle
Brain
K
TPM
Low
Moderate
High
Liver
Muscle
Brain
Enhancer activity
L
RPS
Enhancer activity
M
TPM
Inactive promoters
Active promoters
N
Genes with
Highly active enhancers
Moderately active enhancers
Lowly active enhancers
No enhancers
Proportion of genes (%)
O
Genes with enhancers of
High activity
Moderate activity
Low activity
Percentage (%)
RPS
P
Genes with
Active promoters
Inactive promoters
Percentage (%)
No PEIs
RPS
Q
Mean = 0.978 0.978 0.978 0.976 0.701
Pearson's r of RPS
R
Mean = 0.968 0.961 0.964 0.961 0.829
Pearson's r of H3K4me3 signal
S
Mean = 0.941 0.899 0.912 0.902 0.661
Pearson's r of H3K27ac signal
Homologs
Replicates
Parent of origins
Between breeds
Between tissues

Supplemental Fig. S14. Haplotype-resolved interrogation of PEIs in gene expression control. **(A)** The number of average PEIs per gene using different FDR threshold. **(B** and **C)** Number (**B**) and bridging size distribution (**C**) of PEIs in 24 haploid genomes. **(D)** Summary of genes (*n* = 19,328 in autosomes) interacting with different numbers of enhancers. **(E)** Proportion of PEIs within or across TADs. Expected frequencies were calculated using shuffling TAD linear positions for 1,000 times. Data are presented as mean ± SD. The *P* values were calculated using a paired Student's *t*-test. **(F)** Proportion of promoters interacting with the nearest enhancer or skipping at least one enhancer (scenarios are illustrated in the *upper left*). **(G)** Genes with higher expression generally contact with more enhancers. The dots on the bar represent the mean number of enhancers of a given gene across eight haplotypes for each of the three tissues. The *P* values were obtained using a paired Student's *t*-test. **(H)** Genes with higher expression generally had larger RPS. The *P* values were calculated using the Wilcoxon rank-sum test. **(I)** Saturation curves of H3K27ac density (see Supplemental Methods). The number of ranked moderately active enhancers and super enhancers of high activity by H3K27ac density (x-axis) and their respective densities (y-axis) are plotted. Horizontal dotted lines represent density cutoffs used for the classification of super enhancers and vertical dotted lines demarcate super enhancers from moderately active enhancers. **(J)** The proportion of genes interacting with enhancers have high activity (*i.e.*, super-enhancer), moderate activity, and lesser activity. **(K** and **L)** Expression (**K**) and RPS (**L**) values of genes interacting with enhancers have different activities. The *P* values were calculated using the Wilcoxon rank-sum test. **(M)** Expression of genes with active (enriched by H3K4me3 peaks) or inactive (depleted in H3K4me3 peaks) promoters. The *P* values were calculated using the Wilcoxon rank-sum test. **(N)** Active promoters containing a higher proportion of genes interacted with more active enhancers. **(O** and **P)** Genes with an active promoter interacted more often with enhancers with increased activity. Genes with higher RPS tended to interact with enhancers (**O**) and promoters (**P**) and have increased activities. **(Q**–**S)** Haplotype-resolved interrogation of similarities in the regulatory potential score (RPS) for genes (**Q**), and genome-wide distribution of ChIP-seq signals of H3K27ac (**R**) and H3K4me3 (**S**). The correlation rates were grouped into the following five categories: between homologs, among biological replicates, between parent-of-origins, between parental breeds, among tissues. Data are shown as mean ± SD.

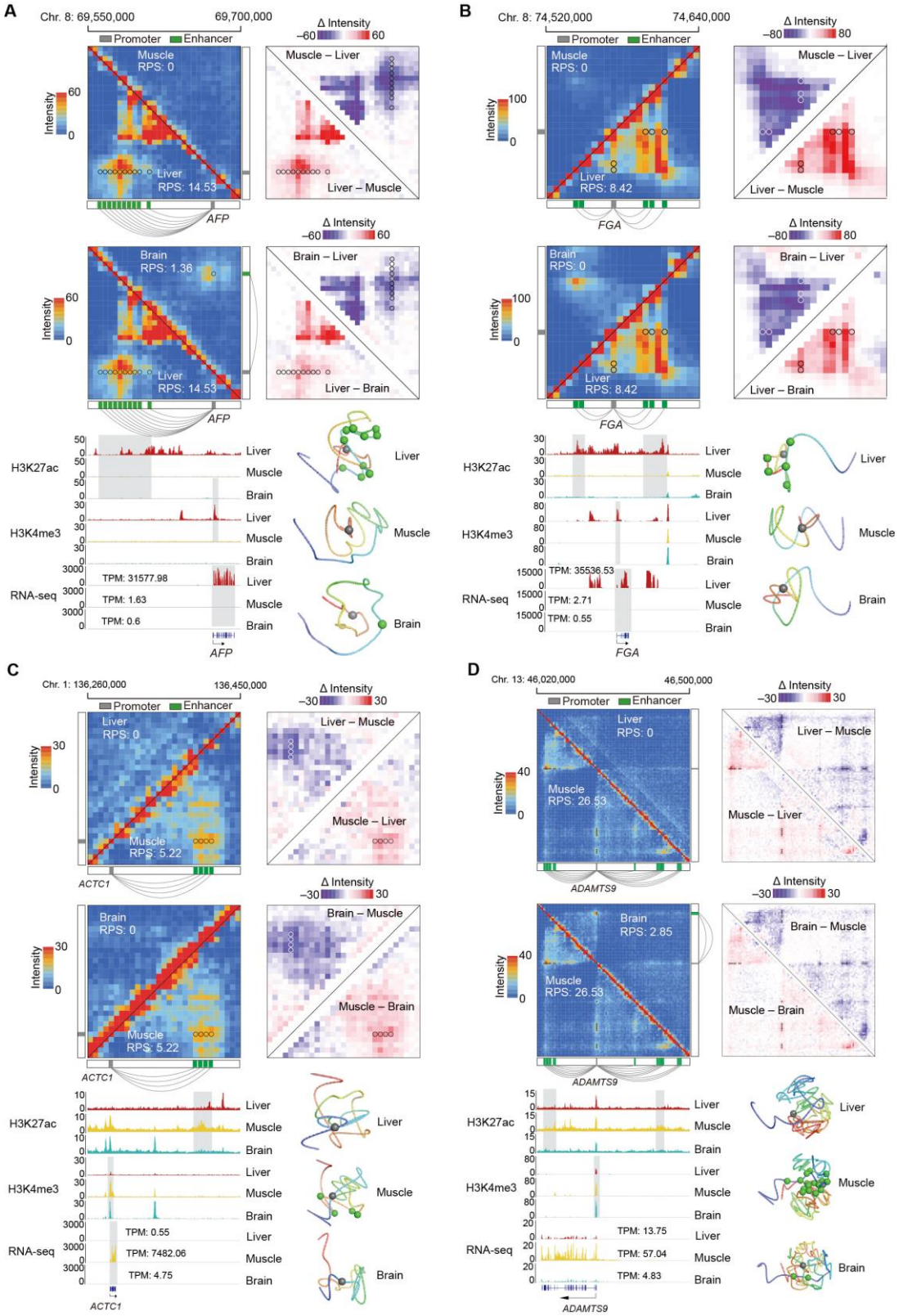

A
Muscle up-regulated
Liver up-regulated
n = 2288
n = 2422
P = 0.05
ΔRPS (liver – muscle)
Brain up-regulated
Liver up-regulated
n = 2595
n = 2511
ΔRPS (liver – brain)
Brain up-regulated
Muscle up-regulated
n = 2507
n = 2273
ΔRPS (muscle – brain)
−log10(P-value)
B
Median = 0.49
Median = 0.54
Median = 0.53
P < 2.2×10−16
log2(expression fold change) (liver / muscle)
log2(expression fold change) (liver / brain)
log2(expression fold change) (muscle / brain)
Genes with higher RPS in liver
Random
Genes with higher RPS in muscle
Genes with higher RPS in brain
C
Liver vs. muscle
Liver vs. brain
Muscle vs. brain
Genes with higher RPS in liver (n = 2422)
Regulation of hormone levels (90/426)
Fat cell differentiation (31/100)
Response to chemokine (24/72)
Regulation of ERK1 and ERK2 cascade (57/260)
Response to nutrient (34/128)
Regulation of platelet activation (16/46)
Hormone metabolic process (36/156)
Response to lipoprotein particle (11/26)
Response to iron ion (11/26)
Humoral immune response (38/172)
Genes with higher RPS in muscle (n = 2288)
Muscle structure development (106/427)
Skeletal system development (102/461)
Muscle cell differentiation (60/233)
Cellular response to growth factor stimulus (92/423)
Extracellular matrix organization (60/248)
Anterior/posterior pattern specification (49/194)
Regulation of muscle adaptation (25/73)
Ossification (58/254)
Regulation of actin filament-based process (73/344)
Negative regulation of phosphate metabolic process (79/383)
−log10(P-value)
Genes with higher RPS in liver (n = 2511)
Fat cell differentiation (35/100)
Regulation of hormone levels (91/426)
Regulation of cell adhesion (129/666)
Hepatocyte proliferation (7/7)
Skeletal system development (91/461)
Cellular hormone metabolic process (28/95)
Regulation of wound healing (31/116)
Steroid metabolic process (47/207)
Cellular response to mechanical stimulus (20/63)
Anterior/posterior pattern specification (44/194)
Genes with higher RPS in brain (n = 2595)
Cell morphogenesis (147/622)
Brain development (151/649)
Regulation of nervous system development (102/394)
Regulation of neuron differentiation (56/174)
Gliogenesis (58/195)
Cell-cell adhesion (106/440)
Calcium signaling pathway (60/211)
Negative regulation of cell differentiation (127/569)
Positive regulation of neural precursor cell proliferation (21/46)
Regulation of synapse organization (52/179)
Genes with higher RPS in muscle (n = 2273)
Muscle structure development (112/427)
Skeletal system development (113/461)
Muscle cell differentiation (65/233)
Regulation of cartilage development (27/65)
Enzyme linked receptor protein signaling pathway (106/559)
Embryonic skeletal system development (34/122)
Pathways in cancer (88/464)
Hippo signaling pathway (37/148)
Cellular response to growth factor stimulus (81/423)
Negative regulation of phosphate metabolic process (74/383)
Genes with higher RPS in brain (n = 2507)
Brain development (158/649)
Gliogenesis (53/195)
Behavior (110/511)
Cell-cell adhesion (96/440)
Modulation of chemical synaptic transmission (84/375)
Regulation of neuron projection development (85/382)
Dopaminergic neuron differentiation (14/29)
Axon development (77/344)
Glutamatergic synapse (31/102)
Regulation of nervous system development (85/394)
D
Germ layer markers
Endoderm
Mesoderm
Ectoderm
n = 720
n = 598
n =1627
Liver (n = 8)
Muscle (n = 8)
Brain (n = 8)
Normalized RPS
Normalized TPM
Genes without phased exonic SNVs
FGA
AFP
ADAMTS9
ACTC1
MAP2
SEMA5B
E
RPS of germ layer markers
Liver
Muscle
Brain
Endoderm markers
Mesoderm markers
Ectoderm markers
F
Endoderm markers
Mesoderm markers
Ectoderm markers
Liver vs. muscle
Spearman's r = 0.24, P < 2.2×10−16
Liver vs. brain
Spearman's r = 0.24, P = 1.48×10−9
Muscle vs. liver
Spearman's r = 0.35, P < 2.2×10−16
Muscle vs. brain
Spearman's r = 0.27, P < 2.2×10−16
Brain vs. liver
Spearman's r = 0.25, P < 2.2×10−16
Brain vs. muscle
Spearman's r = 0.24, P < 2.2×10−16
log2(expression fold change) (liver / muscle)
log2(expression fold change) (liver / brain)
log2(expression fold change) (muscle / liver)
log2(expression fold change) (muscle / brain)
log2(expression fold change) (brain / liver)
log2(expression fold change) (brain / muscle)
ΔRPS (liver – muscle)
ΔRPS (liver – brain)
ΔRPS (muscle – liver)
ΔRPS (muscle – brain)
ΔRPS (brain – liver)
ΔRPS (brain – muscle)

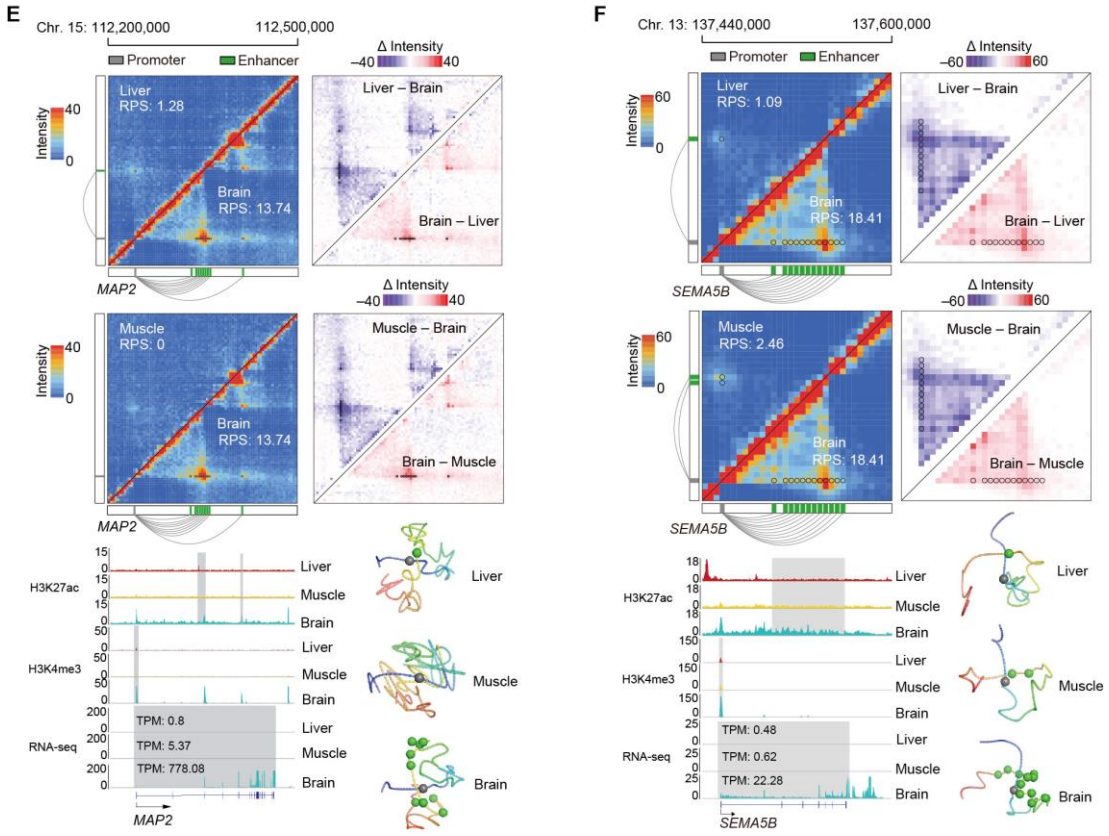
Supplemental Fig. S15. Reorganized haplotype-resolved PEIs between tissues. **(A)** Volcano plots of genes that showed statistically different RPS ($|\Delta RPS| > 3$ and $P < 0.05$, paired Student's *t*-test) between liver and brain (left), liver and muscle (middle), and muscle and brain (right). **(B)** Co-variation between RPS and gene expression. Genes having higher RPS values ($|\Delta RPS| > 3$ and $P < 0.05$, paired Student's *t*-test) between tissues are generally up-regulated compared to randomly selected genes. *P*-values were calculated using the Wilcoxon rank-sum test. **(C)** Top ten significantly enriched GO-BP terms and KEGG pathways (performed using the software Metascape) for genes with increased RPS in pairwise tissue comparisons. **(D)** Heatmaps showing RPS (top) and expression (bottom) of marker genes for the definitive germ layers. Out of 4428 germ layer markers in the pig genome (Jin et al. 2021), 2945 marker genes (720 for endoderm, 598 for mesoderm, and 1627 for ectoderm) are in contact with enhancers and thus have RPS values in at least one tissue. Each gene's RPS values and their respective expression levels are standardized in order to range between 0 and 1. Typical genes are indicated below. **(E)** Comparison between RPS values of germ layer markers across tissues. *P* values were calculated using a Wilcoxon rank-sum test. **(F)** A scatter plot and trend line (Pearson's *r*) showed correlation between changes in RPS and gene expression for germ layer makers of endoderm (512 and 574 markers with RPS and evident expression, *i.e.*, TPM > 0.5, in at least one tissue for liver vs. muscle and liver vs. brain, respectively; left), mesoderm (511 and 488 markers with RPS and evident expression in at least one tissue for muscle vs. liver and muscle vs. brain, respectively; middle), and ectoderm (1380 and 1356 markers with RPS and evident expression in at least one tissue for brain vs. liver and brain vs. muscle, respectively; right) across the three tissues. The blue line represents the linear regression. Red and teal dots represent genes with coordinate and incoordinate changes in RPS and expression, respectively. The number of genes in each quadrant showing a specific relationship between RPS and expression differences is provided (upper left: genes with lower PRS that are upregulated; bottom left: genes with lower RPS that are downregulated; upper right: genes with higher RPS that are upregulated; bottom right: genes with higher RPS that are downregulated). As expected, the makers of the three germ layers generally exhibited higher RPS and upregulated expression in the tissue that originates from the respective germ layer (the endoderm-derived liver, mesoderm-derived muscle, and ectoderm-derived brain), as indicated by the significant enrichment ($P < 0.05$, χ^2 test) of genes in the upper right quadrant (in colored shadow) of the scatter plots.


A
Chr. 3: 96,600,000
96,690,000
Promoter
Enhancer
Muscle
RPS: 0
Liver
RPS: 4.92
Intensity
ABCG5
Δ Intensity
Muscle – Liver
Liver – Muscle
Brain
RPS: 0
Brain – Liver
Liver – Brain
H3K27ac
H3K4me3
RNA-seq
TPM: 149.73
TPM: 0
TPM: 0
Liver
Muscle
Brain
B
Chr. 9: 63,695,000
63,910,000
Muscle
RPS: 1.36
Liver
RPS: 13.11
FMO1
Brain
RPS: 0
TPM: 1599.81
TPM: 59.79
TPM: 6.02
C
Chr. 15: 113,010,000
113,130,000
Liver
RPS: 0
Muscle
RPS: 6.54
MYL1
Brain – Muscle
Muscle – Brain
TPM: 1.54
TPM: 47369.52
TPM: 4.02
D
Chr. 2: 139,840,000
139,940,000
Liver
RPS: 0
Muscle
RPS: 6.91
MYOT
Brain
RPS: 0
TPM: 0.06
TPM: 1613.08
TPM: 0.06

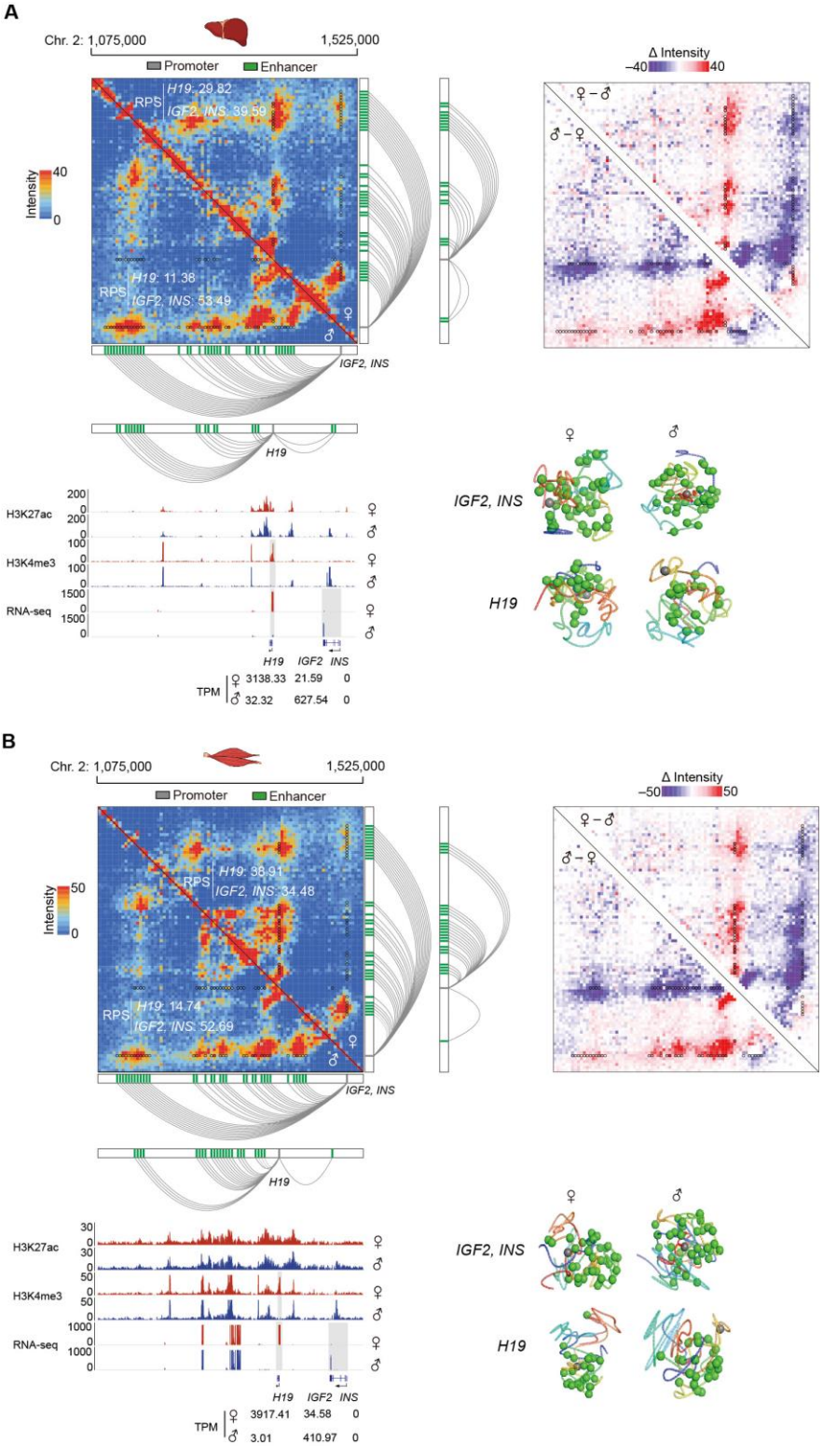


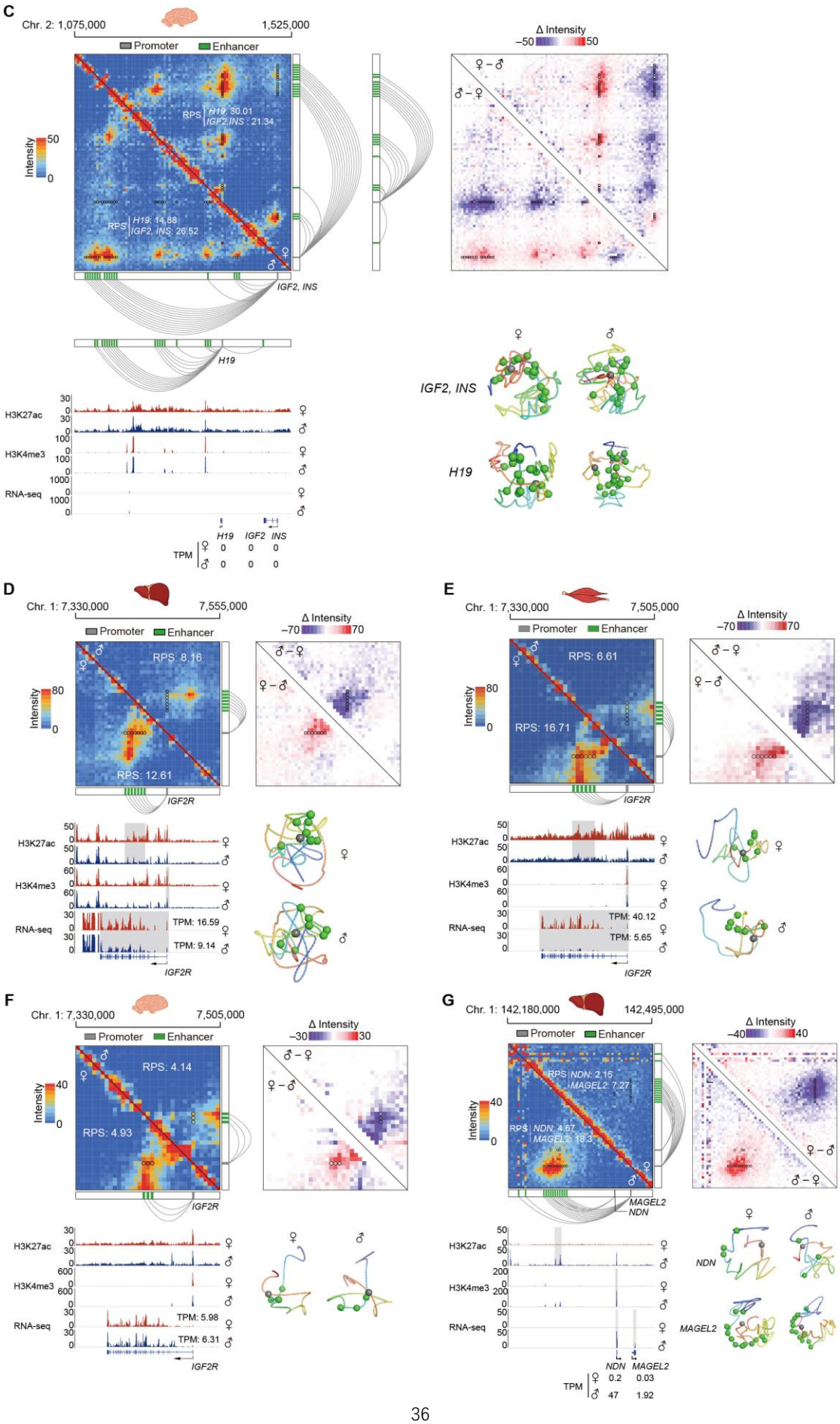
Supplemental Fig. S16. Representative examples of tissue-specific PEIs for *ABCG5* (**A**) and *FMO1* (**B**) in the liver, *MYL1* (**C**) and *MYOT* (**D**) in the muscle, and *NRGN* (**E**) and *TBR1* (**F**) in the brain. Top to bottom: Hi-C maps (upper left) and their allelic differences (upper right), ChIP-seq signals of H3K27ac and H3K4me3, and gene expression levels (lower left) and 3D structural models (lower right) of the corresponding genomic regions. Promoter (grey square), enhancers (green squares), and PEIs (connecting lines) are displayed beside the Hi-C maps.

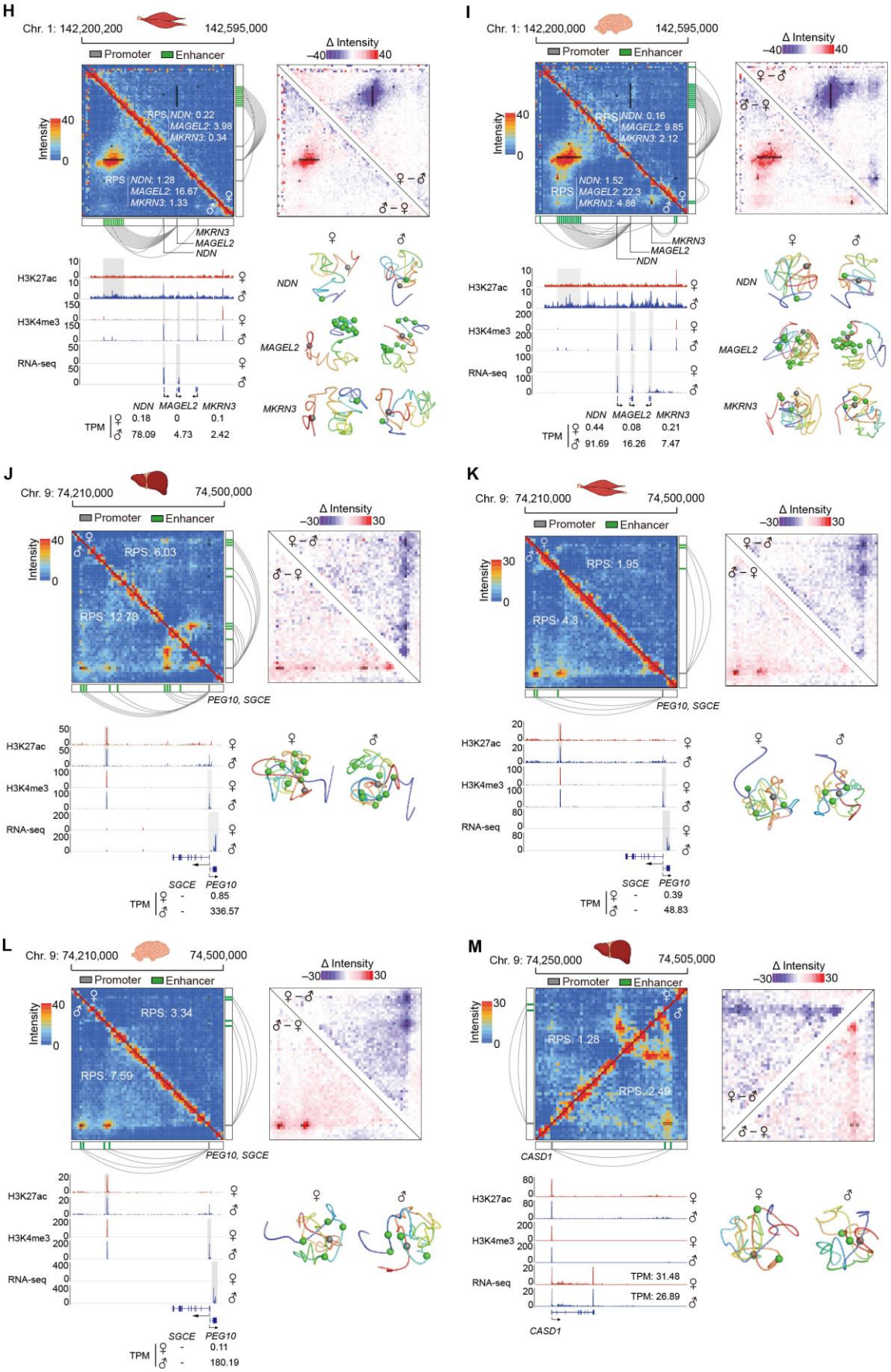

A
Chr. 8: 69,550,000
69,700,000
Promoter
Enhancer
Muscle
RPS: 0
Liver
RPS: 14.55
Intensity
Δ Intensity
Muscle – Liver
Liver – Muscle
AFP
Brain
RPS: 1.36
Liver
RPS: 14.53
Brain – Liver
Liver – Brain
H3K27ac
H3K4me3
RNA-seq
TPM: 31577.98
TPM: 1.63
TPM: 0.6
Liver
Muscle
Brain
B
Chr. 8: 74,520,000
74,640,000
Muscle
RPS: 0
Liver
RPS: 8.42
FGA
Brain
RPS: 0
TPM: 35536.53
TPM: 2.71
TPM: 0.55
C
Chr. 1: 136,260,000
136,450,000
Liver
RPS: 0
Muscle
RPS: 5.22
ACTC1
Brain
RPS: 0
Brain – Muscle
Muscle – Brain
TPM: 0.55
TPM: 7482.06
TPM: 4.75
D
Chr. 13: 46,020,000
46,500,000
Liver
RPS: 0
Muscle
RPS: 26.53
ADAMTS9
Brain
RPS: 2.85
TPM: 13.75
TPM: 57.04
TPM: 4.83

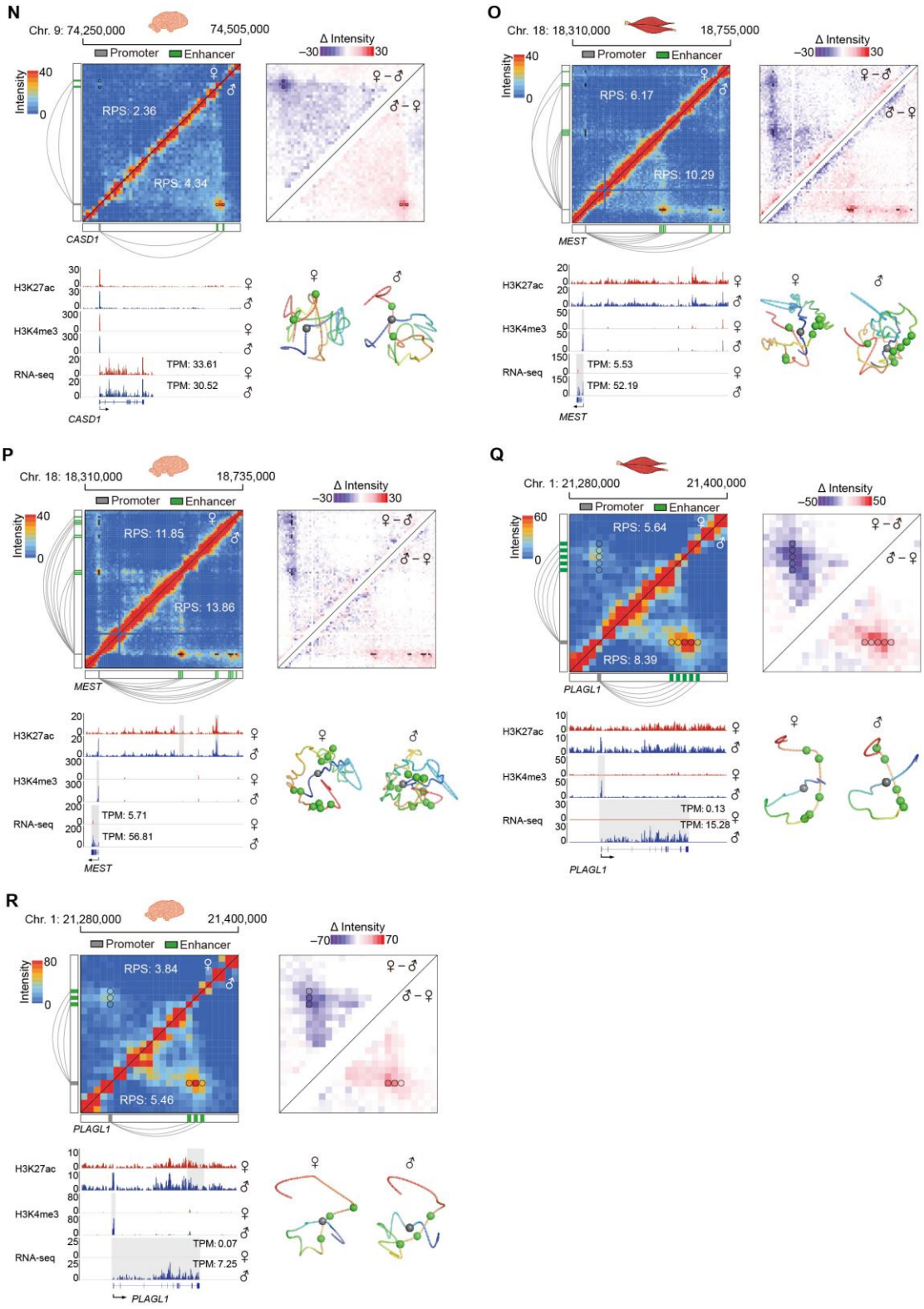


Supplemental Fig. S17. Representative examples of specific PEIs for germ layer markers in the tissues originated from their respective germ layers. Specifiers of (**A** and **B**) endoderm (*AFP* and *FGA*) in endoderm-derived liver; (**C** and **D**) mesoderm (*ACTC1* and *ADAMTS9*) in mesoderm-derived muscle; and (**E** and **F**) ectoderm (*MAP2* and *SEMA5B*) in ectoderm-derived brain. The schematics are similar to those shown in **Supplemental Fig. S16**.


A
Chr. 2: 1,075,000
1,525,000
Promoter
Enhancer
H19: 29.82
RPS
IGF2, INS: 39.59
Intensity
40
0
H19: 11.38
RPS
IGF2, INS: 53.49
IGF2, INS
H19
Δ Intensity
−40
40
♀ − ♂
♂ − ♀
H3K27ac
H3K4me3
RNA-seq
200
100
1500
H19
IGF2
INS
TPM
♀ 3138.33 21.59 0
♂ 32.32 627.54 0
IGF2, INS
H19
B
Chr. 2: 1,075,000
1,525,000
Promoter
Enhancer
H19: 38.91
RPS
IGF2, INS: 34.48
Intensity
50
0
H19: 14.74
RPS
IGF2, INS: 52.69
IGF2, INS
H19
Δ Intensity
−50
50
♀ − ♂
♂ − ♀
H3K27ac
H3K4me3
RNA-seq
30
50
1000
H19
IGF2
INS
TPM
♀ 3917.41 34.58 0
♂ 3.01 410.97 0
IGF2, INS
H19

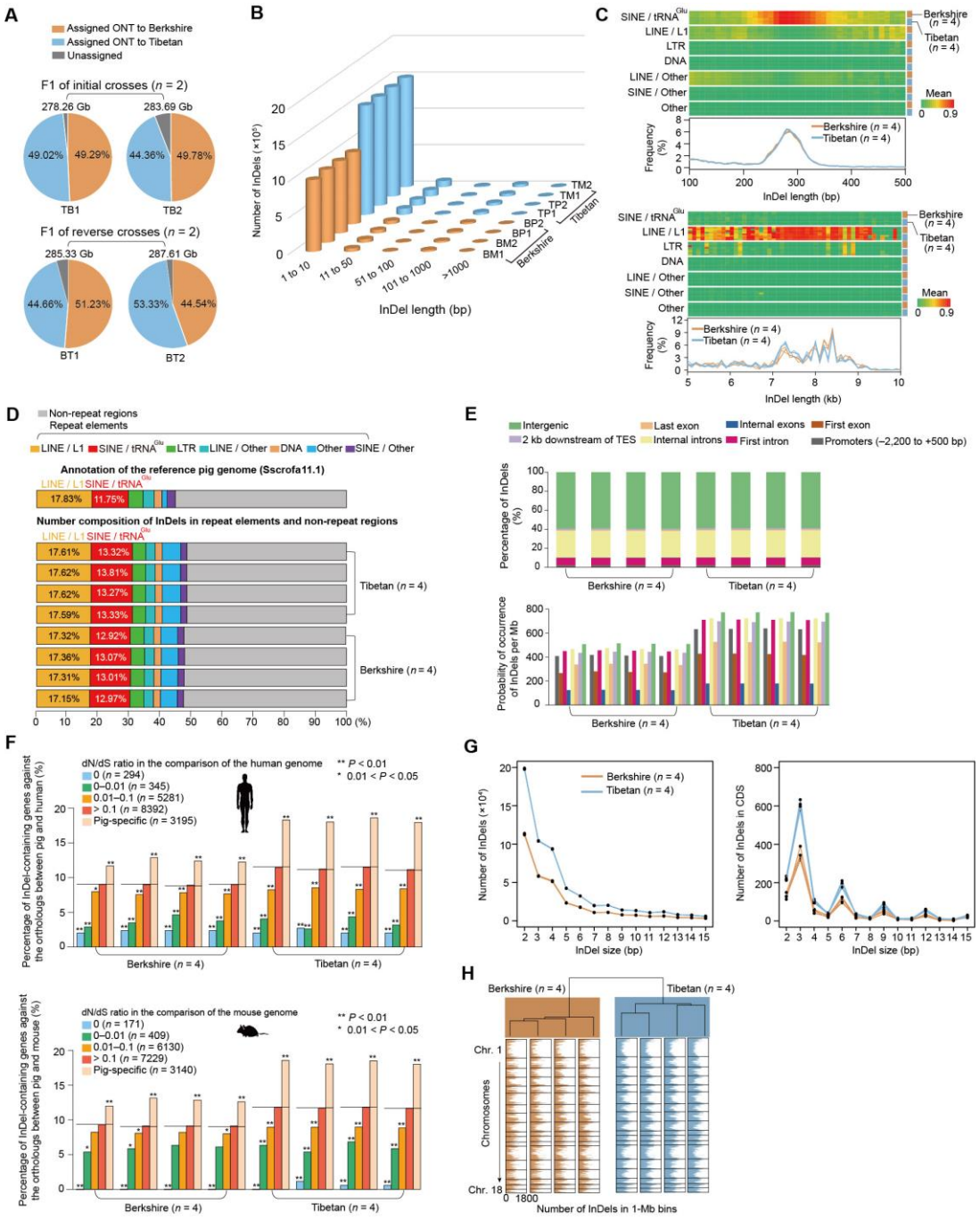

C
Chr. 2: 1,075,000
1,525,000
Promoter
Enhancer
Δ Intensity
−50
50
RPS
H19: 30.01
IGF2.INS: 21.34
RPS
H19: 14.88
IGF2, INS: 26.52
Intensity
50
0
IGF2, INS
H19
H3K27ac
H3K4me3
RNA-seq
TPM
H19
IGF2
INS
IGF2, INS
H19
D
Chr. 1: 7,330,000
7,555,000
Promoter
Enhancer
Δ Intensity
−70
70
RPS: 8.16
RPS: 12.61
Intensity
80
0
IGF2R
H3K27ac
H3K4me3
RNA-seq
TPM: 16.59
TPM: 9.14
IGF2R
E
Chr. 1: 7,330,000
7,505,000
Promoter
Enhancer
Δ Intensity
−70
70
RPS: 6.61
RPS: 16.71
Intensity
80
0
IGF2R
H3K27ac
H3K4me3
RNA-seq
TPM: 40.12
TPM: 5.65
IGF2R
F
Chr. 1: 7,330,000
7,505,000
Promoter
Enhancer
Δ Intensity
−30
30
RPS: 4.14
RPS: 4.93
Intensity
40
0
IGF2R
H3K27ac
H3K4me3
RNA-seq
TPM: 5.98
TPM: 6.31
IGF2R
G
Chr. 1: 142,180,000
142,495,000
Promoter
Enhancer
Δ Intensity
−40
40
RPS
NDN: 2.15
MAGEL2: 7.27
RPS
NDN: 4.67
MAGEL2: 18.3
Intensity
40
0
MAGEL2
NDN
H3K27ac
H3K4me3
RNA-seq
NDN
MAGEL2
TPM
NDN
MAGEL2
0.2
0.03
47
1.92
NDN
MAGEL2


H
Chr. 1: 142,200,200
142,595,000
Promoter
Enhancer
Δ Intensity
−40
40
Intensity
RPS
NDN: 0.22
MAGEL2: 3.98
MKRN3: 0.34
NDN: 1.28
MAGEL2: 16.67
MKRN3: 1.33
MKRN3
MAGEL2
NDN
H3K27ac
H3K4me3
RNA-seq
NDN
MAGEL2
MKRN3
TPM
NDN MAGEL2 MKRN3
0.18 0 0.1
78.09 4.73 2.42
I
Chr. 1: 142,200,000
142,595,000
NDN: 0.16
MAGEL2: 9.85
MKRN3: 2.12
NDN: 1.52
MAGEL2: 22.3
MKRN3: 4.86
0.44 0.08 0.21
91.69 16.26 7.47
J
Chr. 9: 74,210,000
74,500,000
RPS: 6.03
RPS: 12.78
PEG10, SGCE
SGCE PEG10
- 0.85
- 336.57
K
Chr. 9: 74,210,000
74,500,000
RPS: 1.95
RPS: 4.3
PEG10, SGCE
SGCE PEG10
- 0.39
- 48.83
L
Chr. 9: 74,210,000
74,500,000
RPS: 3.34
RPS: 7.59
PEG10, SGCE
SGCE PEG10
- 0.11
- 180.19
M
Chr. 9: 74,250,000
74,505,000
RPS: 1.28
RPS: 2.49
CASD1
TPM: 31.48
TPM: 26.89
CASD1



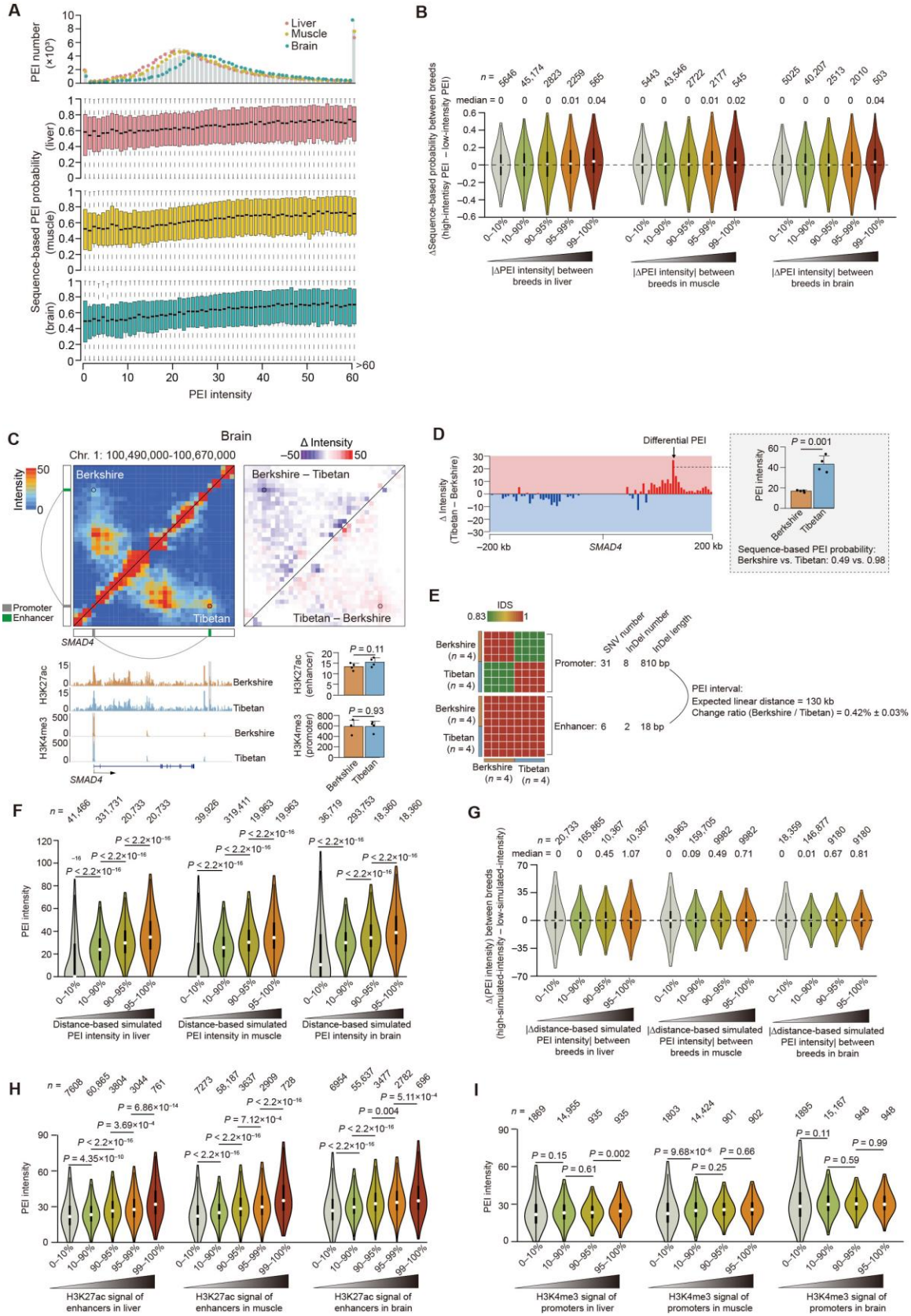
Supplemental Fig. S18. Imprinting-specific PEI organizations. A total of 12 allelic-expression testable imprinted genes (with informative SNVs and interacting with enhancers in at least one tissue) that exhibited strong evidence of parent-of-origin-specific PEI organization in at least two of the three tissues are shown. Paternally imprinted *IGF2/INS* and maternally imprinted *H19* in the liver (**A**), muscle (**B**), and brain (**C**); maternally

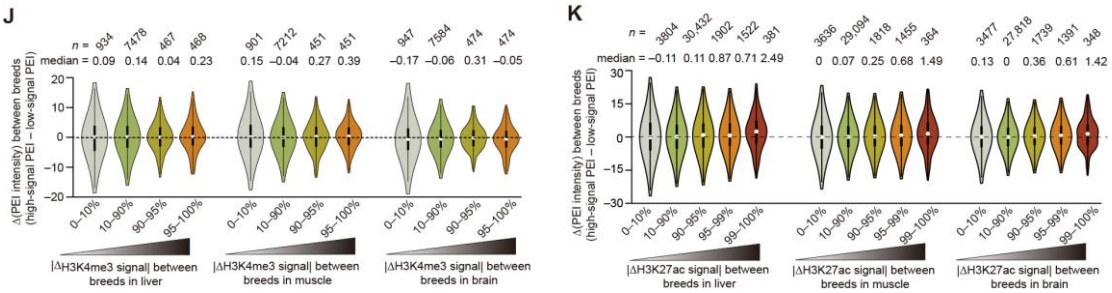
imprinted *IGF2R* in the liver (**D**), muscle (**E**), and brain (**F**); paternally imprinted *NDN, MAGEL2* and *MKRN3* in the liver (**G**), muscle (**H**), and brain (**I**); paternally imprinted *PEG10* and *SGCE* in the liver (**J**), muscle (**K**), and brain (**L**); paternally imprinted gene *CASD1* in the liver (**M**) and brain (**N**); paternally imprinted *MEST* in the muscle (**O**) and the brain (**P**); and paternally imprinted *PLAGL1* in the muscle (**Q**) and brain (**R**). The schematics are similar to those shown in **Supplemental Fig. S16**.



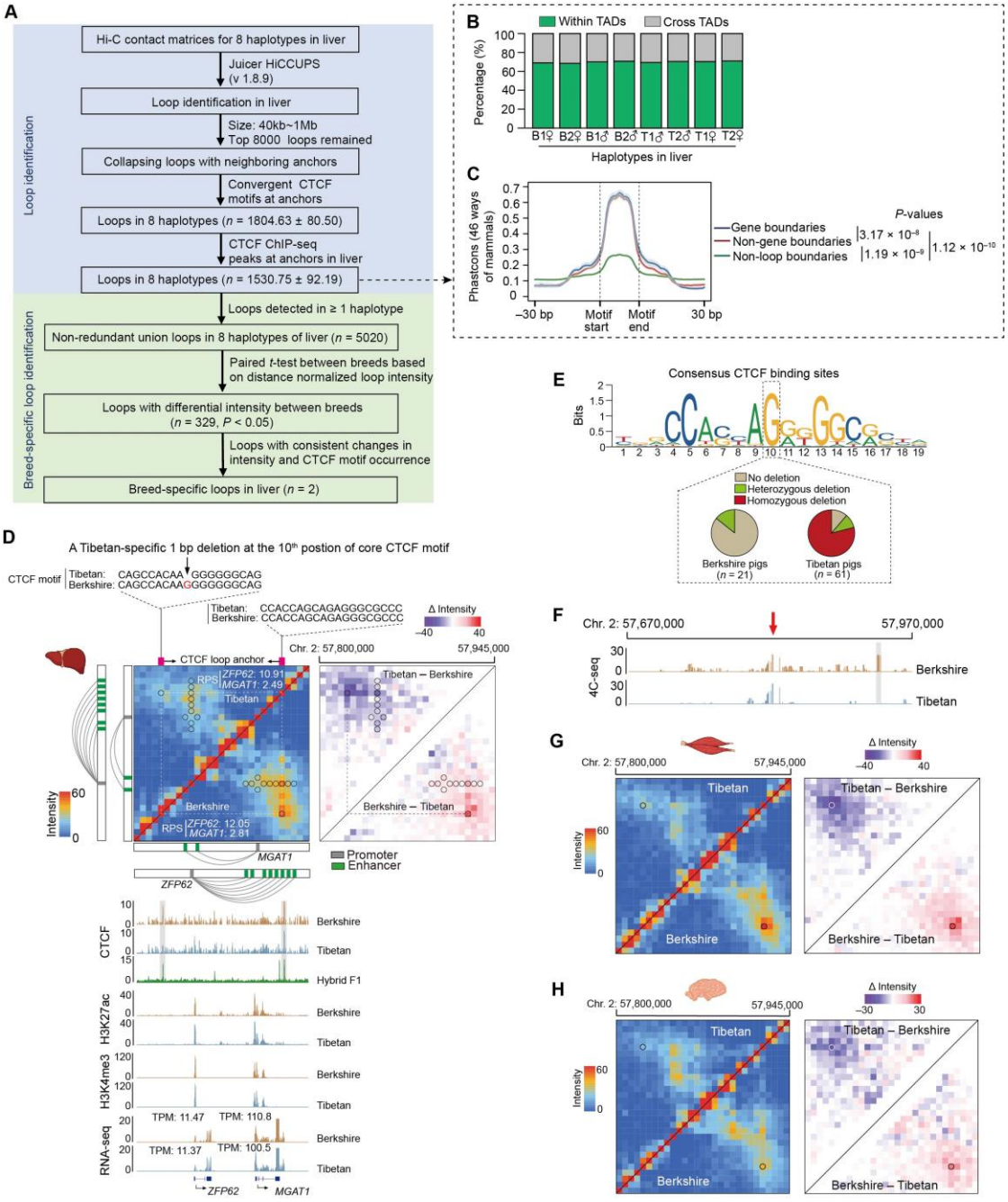
Supplemental Fig. S19. Characteristics of allelic structural variants (SVs). **(A)** Allele assignment of high-quality ONT reads of four F1 hybrids (see Supplemental Methods). **(B)** Length distribution of the indels identified in F1 hybrids. We detected ~734.78 k (~8.42 Mb in length) insertions and ~340.13 k (~9.44 Mb in length) deletions in Berkshire haplotypes mapped against the pig reference genome, and ~1.009 M (~12.51 Mb in length) insertions and ~633.92 k (~15.15 Mb in length) deletions in Tibetan haplotypes mapped against the pig reference genome. More than 91% of the indels spanned 1–10 bp in length. **(C)** Distribution of indels between 100 bp and 500 bp, and 5 kb and 10 kb in length. The peaks at ~300bp and ~8kb in length (top) are probably associated with the enrichment of indels

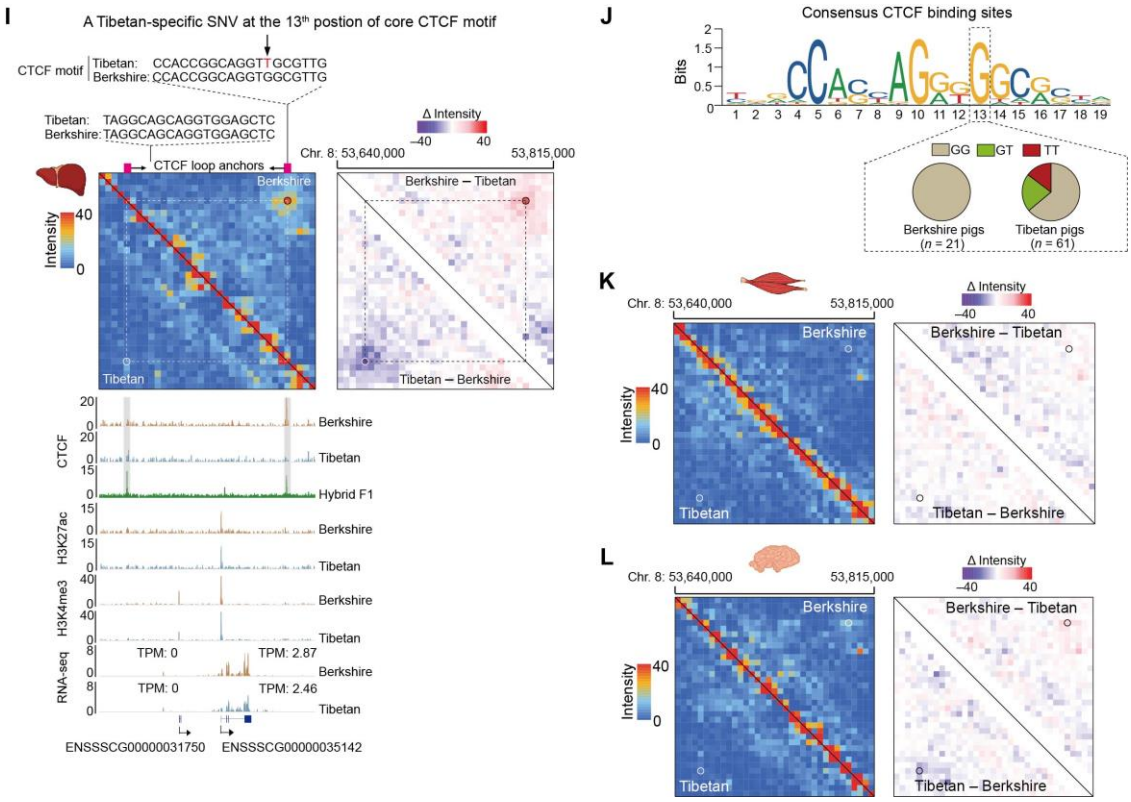
of $tRNA^{Glu}$-derived short interspersed elements (SINE/$tRNA^{Glu}$) and LINE/L1-derived long interspersed elements, respectively (bottom). **(D)** Indels overlapping with repeat elements in pig genomes. Length composition of repeat elements in the pig reference genome is displayed (top). Number composition of indels across 8 haploid genomes in repeat elements and non-repeat regions are shown (bottom). Repetitive elements constitute ~46.21% of the genome, of which ~48.08% were indels, which suggests they are an important source of SVs in the pig genome. Moreover, the SINE/ $tRNA^{Glu}$ (294.47 Mb containing ~13.15% of all indels) showed a higher incidence of indels than the predominant long interspersed elements (LINE/L1) (446.51 Mb containing ~17.45% of all indels). **(E)** Percent distribution (measured by length of indels, top) and probability of occurrence (measured by number of indels, bottom) of indels across different genomic elements. The indels appear to be constrained by selection: more than 50% of indels (~58.95% in length) were located in intergenic regions, and the indel ratio was lower in the coding sequences than in introns. **(F)** Conservation levels of indel-containing genes, which were categorized following dN/dS ratios compared to gene sets of human (top) and mouse (bottom) genomes (downloaded from Ensembl using the biomart online tool). The black horizontal lines represent the overall percentage of indel-containing genes against all the orthologs between the pig and human/mouse genomes for 8 haplotypes. By counting the number of orthologs and indel-containing genes against the genome-wide background, we determined the statistical significance of enrichment of indel-containing genes for each group that was categorized by the dN/dS ratio using a χ^2 test (*$0.01 < P < 0.05$; **$0.001 < P < 0.01$). This analysis showed that more conserved genes contained fewer indels. **(G)** Length distribution of short indels (1–15 bp in length) in the whole genome (left) and coding sequence (CDS) regions (right). We observed an enrichment of short indels (1–15 bp in length) in coding sequences (~28.97%) that were multiples of 3 bp, which is expected to preserve the reading frame. **(H)** Analysis of the phylogenetic relationship among 8 haplotypes using indel numbers in 1 Mb non-overlapping genomic bins. Hierarchical clustering based on Pearson's *r* of indel occurrence was used as distance metrics. As in the case of SNVs (**Supplemental Fig. S3C**), the distribution of short indels across the genome also reflected a deep phylogenetic split between European (*i.e.*, Berkshire) and Asian pigs (Tibetan).


A
PEI number (×10³)
Liver
Muscle
Brain
Sequence-based PEI probability (liver)
Sequence-based PEI probability (muscle)
Sequence-based PEI probability (brain)
PEI intensity
B
ΔSequence-based probability between breeds (high-intentisy PEI – low-intensity PEI)
n =
median =
|ΔPEI intensity| between breeds in liver
|ΔPEI intensity| between breeds in muscle
|ΔPEI intensity| between breeds in brain
C
Brain
Chr. 1: 100,490,000-100,670,000
Δ Intensity
Berkshire
Tibetan
Berkshire – Tibetan
Tibetan – Berkshire
Promoter
Enhancer
SMAD4
H3K27ac
H3K4me3
H3K27ac (enhancer)
H3K4me3 (promoter)
P = 0.11
P = 0.93
D
Differential PEI
Δ Intensity (Tibetan – Berkshire)
SMAD4
PEI intensity
P = 0.001
Sequence-based PEI probability: Berkshire vs. Tibetan: 0.49 vs. 0.98
E
IDS
Berkshire (n = 4)
Tibetan (n = 4)
SNV number
InDel number
InDel length
Promoter: 31 8 810 bp
Enhancer: 6 2 18 bp
PEI interval:
Expected linear distance = 130 kb
Change ratio (Berkshire / Tibetan) = 0.42% ± 0.03%
F
PEI intensity
Distance-based simulated PEI intensity in liver
Distance-based simulated PEI intensity in muscle
Distance-based simulated PEI intensity in brain
G
Δ(PEI intensity) between breeds (high-simulated-intensity – low-simulated-intensity)
|Δdistance-based simulated PEI intensity| between breeds in liver
|Δdistance-based simulated PEI intensity| between breeds in muscle
|Δdistance-based simulated PEI intensity| between breeds in brain
H
PEI intensity
H3K27ac signal of enhancers in liver
H3K27ac signal of enhancers in muscle
H3K27ac signal of enhancers in brain
I
PEI intensity
H3K4me3 signal of promoters in liver
H3K4me3 signal of promoters in muscle
H3K4me3 signal of promoters in brain



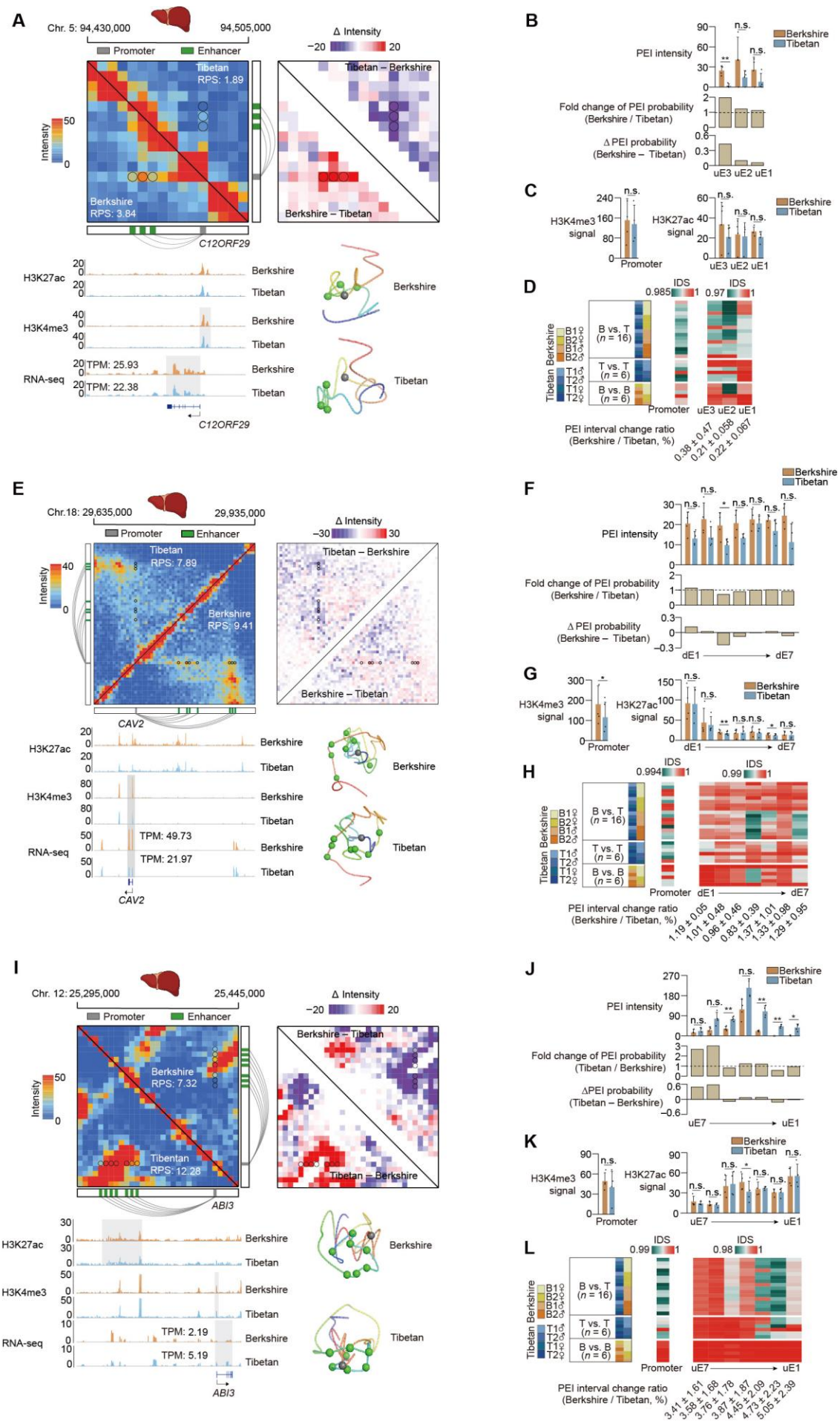
Supplemental Fig. S20. Effects of sequence variation and histone modification on allelic PEI rewiring in F1 hybrids. **(A)** Among distinct PEIs within each of the 24 haploid genomes, we found that those with increased intensity (x-axis) generally have elevated interaction probabilities (y-axis). Data are shown as mean ± SD (eight haploid genomes for each of the three tissues). The probability of long-range interactions for each allelic PEI were estimated by considering their variation using the PEP algorithm that only employs sequence-based features (see Supplemental Methods; Yang et al. 2017b). The intensity distribution of PEIs in each tissue are also shown in the top. **(B)** For the comparison of particular PEIs between two parental alleles in F1 hybrids in a specific tissue, those with greater changes in intensity (especially those in the highest top 1% allelic intensity changes; see Supplemental Methods) generally have elevated allelic differences of estimated interaction probabilities. The *P*-values were calculated using the Wilcoxon rank-sum test. (**C**–**E**) The example PEI in *SMAD4* was potentially disrupted by sequence variations between the two parental alleles. (**C**) Hi-C maps (upper left) and their allelic differences (upper right); ChIP-seq signals of H3K27ac (enhancer) and H3K4me3 (promoter) (bottom left), and a statistical comparison of the example PEI between breeds (bottom right). (**D**) Allelic differences in promoter-centered interaction patterns (200 kb up- and downstream regions from the promoter) (left); confirmation of allelic rewiring of the example PEI by predicted sequence-based PEI probabilities (see Supplemental Methods; right). *P*-values are from paired Student's *t*-test. (**E**) Pairwise comparison of the degree of haplotype similarity (measured by IDS) in promoters and enhancers among eight haplotypes; note that Berkshire pigs appear to carry a single allele. Summary of sequence variants (SNVs and indels) as well as changes in PEI interval length are shown. **(F** and **G)** Among the distinct PEIs within each haploid genome (**F**), or for a given PEI between two parental alleles (**G**), the simulated PEI intensities based on the variable bridging distances of the PEI using Huynh's algorithm (see Supplemental Methods) suggest that promoter and enhancers neighbor each other in the linear genome, and generally have an elevated interaction intensity. The *P*-values were calculated using the Wilcoxon rank-sum test. **(H**–**K)** Among the distinct PEIs within each haploid genome (**H** and **J**), or for a given PEI between two parental alleles (**I** and **K**), the increased activities for H3K4me3-marker promoters (**H** and **I**) and H3K27ac-marked enhancers (**J** and **K**) generally have an elevated PEI intensity. The *P*-values were calculated using the Wilcoxon rank-sum test.

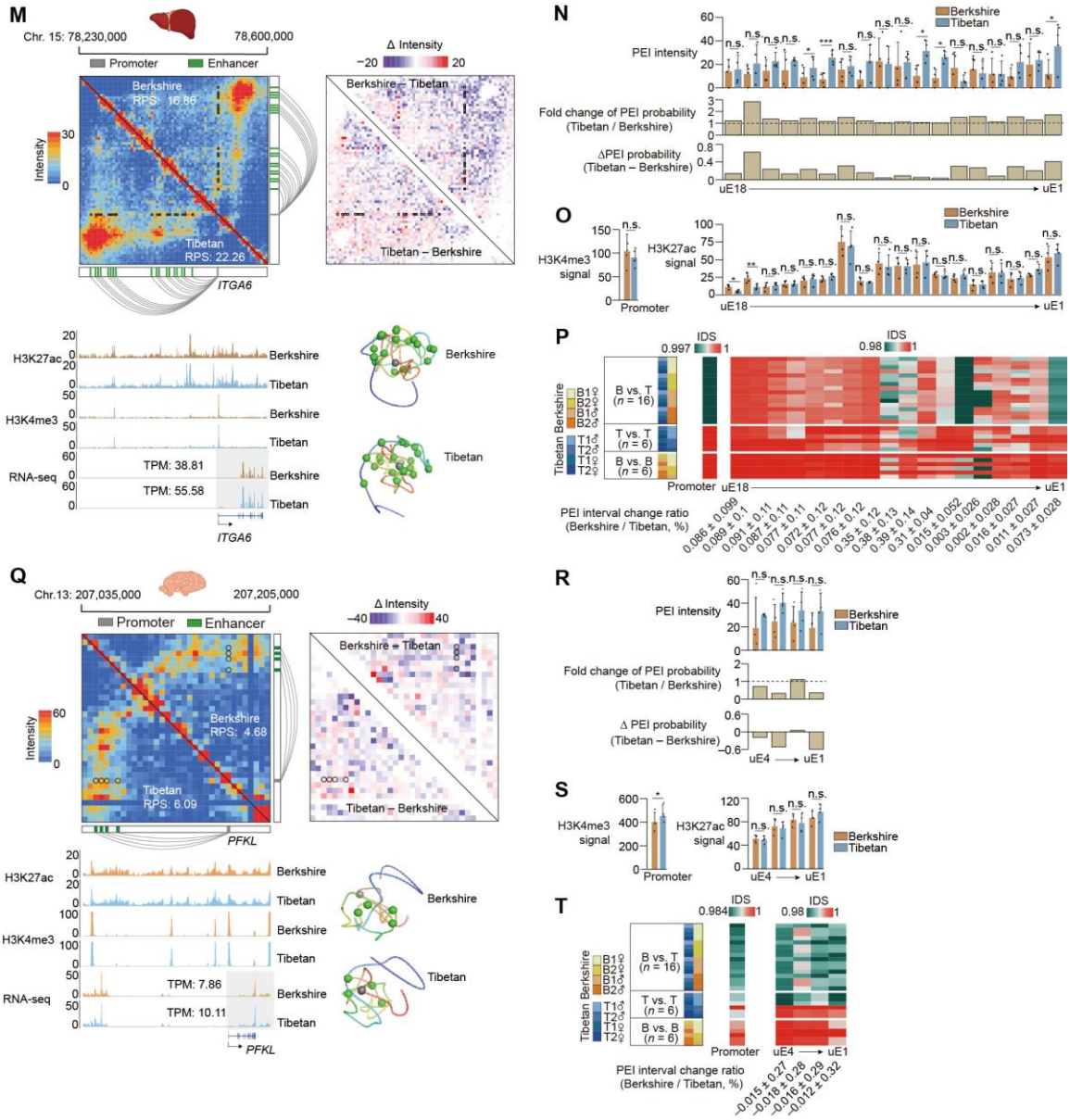

A
Hi-C contact matrices for 8 haplotypes in liver
Juicer HiCCUPS (v 1.8.9)
Loop identification in liver
Size: 40kb~1Mb Top 8000 loops remained
Collapsing loops with neighboring anchors
Convergent CTCF motifs at anchors
Loops in 8 haplotypes (n = 1804.63 ± 80.50)
CTCF ChIP-seq peaks at anchors in liver
Loops in 8 haplotypes (n = 1530.75 ± 92.19)
Loop identification
Loops detected in ≥ 1 haplotype
Non-redundant union loops in 8 haplotypes of liver (n = 5020)
Paired t-test between breeds based on distance normalized loop intensity
Loops with differential intensity between breeds (n = 329, P < 0.05)
Loops with consistent changes in intensity and CTCF motif occurrence
Breed-specific loops in liver (n = 2)
Breed-specific loop identification
B
Within TADs
Cross TADs
Percentage (%)
B1♀ B2♀ B1♂ B2♂ T1♂ T2♂ T1♀ T2♀
Haplotypes in liver
C
Phastcons (46 ways of mammals)
Gene boundaries
Non-gene boundaries
Non-loop boundaries
P-values
3.17 × 10⁻⁸
1.19 × 10⁻⁹
1.12 × 10⁻¹⁰
–30 bp
Motif start
Motif end
30 bp
D
A Tibetan-specific 1 bp deletion at the 10th postion of core CTCF motif
CTCF motif
Tibetan: CAGCCACAA GGGGGGCAG
Berkshire: CAGCCACAAGGGGGGCAG
Tibetan: CCACCAGCAGAGGGCGCCC
Berkshire: CCACCAGCAGAGGGCGCCC
Δ Intensity
–40 40
CTCF loop anchor
Chr. 2: 57,800,000
57,945,000
RPS
ZFP62: 10.91
MGAT1: 2.49
Tibetan
Berkshire
ZFP62: 12.05
MGAT1: 2.81
Tibetan – Berkshire
Berkshire – Tibetan
Intensity
MGAT1
ZFP62
Promoter
Enhancer
CTCF
H3K27ac
H3K4me3
RNA-seq
Berkshire
Tibetan
Hybrid F1
TPM: 11.47
TPM: 110.8
TPM: 11.37
TPM: 100.5
ZFP62
MGAT1
E
Consensus CTCF binding sites
Bits
No deletion
Heterozygous deletion
Homozygous deletion
Berkshire pigs (n = 21)
Tibetan pigs (n = 61)
F
Chr. 2: 57,670,000
57,970,000
4C-seq
Berkshire
Tibetan
G
Chr. 2: 57,800,000
57,945,000
Δ Intensity
–40 40
Tibetan
Berkshire
Tibetan – Berkshire
Berkshire – Tibetan
Intensity
H
Chr. 2: 57,800,000
57,945,000
Δ Intensity
–30 30
Tibetan
Berkshire
Tibetan – Berkshire
Berkshire – Tibetan
Intensity



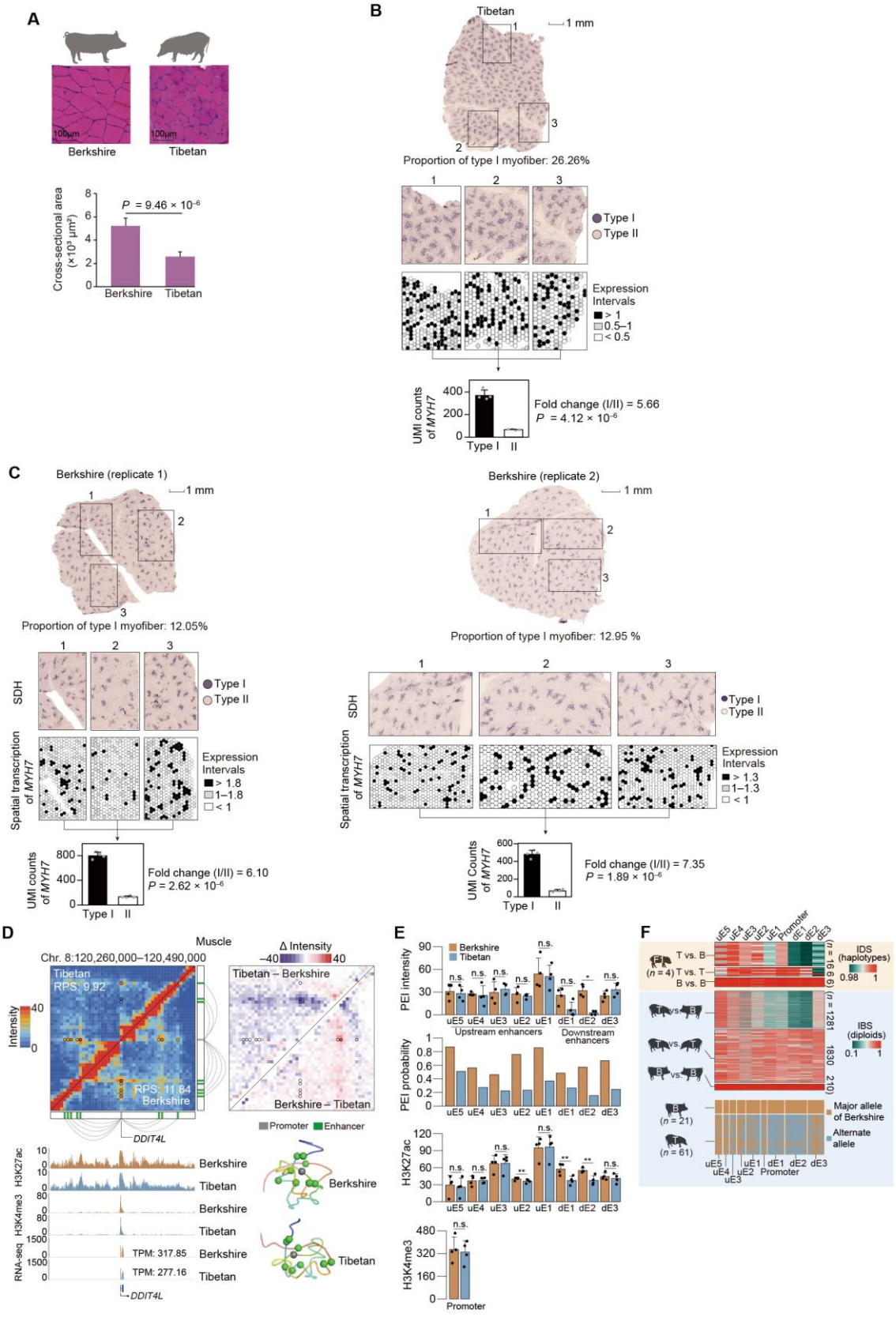
Supplemental Fig. S21. Breed-specific mutations induced allele-specific abolition of two CTCF-mediated loops. (**A**) Identification of breed-specific CTCF loops in the liver. Briefly, we applied Juicer HiCCUPS to identify loops in 8 haplotypes using their Hi-C contact matrices. Only loops with anchor distances between 40-kb and 1-Mb were kept and then ranked by detection significance, with the top 8000 loops retained. Loops with neighboring anchors were collapsed into a non-redundant set, and those without convergent CTCF motifs or CTCF ChIP-seq peaks at both anchors were removed. Loops detected in at least one of the 8 haplotypes were combined into a single non-redundant union set. For each loop set, the distance-normalized interaction intensity was retrieved from the Hi-C contact matrices for the 8 haplotypes. We identified loops with differential intensity between Berkshire and Tibetan haplotypes ($P < 0.05$, paired Student's *t*-test), and those with consistent intensity change and CTCF motif occurrence were considered as breed-specific CTCF loops. (**B**) Proportion of CTCF loops within and across TADs. (**C**) Evolutionary conservation (inferred by Phastcons) of consensus CTCF motifs at CTCF loop boundaries with or without genes, and non-boundary regions. The motif region (19 bp in length) is indicated by the dashed box. Data are presented as mean ± SD. The shaded area indicates the standard deviation. *P*-values were calculated using a paired Student's *t*-test. (**D**–**H**) A 1-bp deletion ('G', Chromosome 2: 57822013; the 10th position of the core CTCF binding motif at the 5' loop anchor) in Tibetan alleles most likely resulted in allele-specific abolition of the CTCF loop. (**D**) Hi-C maps and their allelic differences. Linear positions of the CTCF loop anchors (magenta squares) and sequences of the core CTCF binding site of Tibetan

and Berkshire alleles are shown (top). ChIP-seq signals of CTCF, H3K27ac and H3K4me3, and expression of two genes located within loop are also exhibited (bottom). (**E**) Sequence logos of the consensus CTCF binding motif (downloaded from https://jaspar.genereg.net/) indicate that the 10th position is highly conserved in vertebrates (top). The frequency of a 1-bp deletion ('G', Chromosome 2: 57822013) in populations of Berkshire (*n* = 21) and Tibetan pigs (*n* = 61) (bottom). Of the 82 pigs, six Berkshire and six Tibetan pigs were from six parent-child trios used in our study. The publicly available SNVs of other 60 pigs (including 15 Berkshire and 55 Tibetan pigs) were retrieved from the ISwine database (Fu et al., 2020). (**F**) 4C-seq signals confirmed the CTCF loop existing in Berkshire pigs were not present in Tibetan pigs. Histogram showing the genomic interactions across the 150 kb up- and down-stream the CTCF loop anchor. Red vertical arrows indicate the viewpoint. The grey shadow shows the differential interaction intensity between Tibetan and Berkshire pigs at the genomic loci where Berkshire had enriched chromatin interactions due to a breed-specific CTCF loop. (**G** and **H**) Hi-C maps around the Tibetan-specific 1-bp deletion ('G', Chromosome 2: 57822013) and their allelic differences in the muscle (**G**) and brain (**H**). Hi-C maps showing substantial differences in loop intensity between the two parental alleles (similar as shown in **E**). (**I**–**L**) A SNV ('G' > 'T', Chromosome 8: 53792054; the 13th position of the core CTCF binding motif of the 3' loop anchor) in Tibetan alleles mostly like resulted in allele-specific abolition of the CTCF loop. The schematics of I–L are similar to those shown in panels **D**–**G**.


A
Chr. 5: 94,430,000
94,505,000
Promoter
Enhancer
Δ Intensity
−20
20
Tibetan
RPS: 1.89
Berkshire
RPS: 3.84
Tibetan – Berkshire
Berkshire – Tibetan
Intensity
C12ORF29
H3K27ac
H3K4me3
RNA-seq
Berkshire
Tibetan
TPM: 25.93
TPM: 22.38
B
PEI intensity
Fold change of PEI probability (Berkshire / Tibetan)
Δ PEI probability (Berkshire – Tibetan)
n.s.
uE3 uE2 uE1
C
H3K4me3 signal
H3K27ac signal
Promoter
D
IDS
0.985
0.97
B vs. T (n = 16)
T vs. T (n = 6)
B vs. B (n = 6)
PEI interval change ratio (Berkshire / Tibetan, %)
0.38 ± 0.47
0.21 ± 0.058
0.22 ± 0.067
E
Chr.18: 29,635,000
29,935,000
CAV2
F
dE1
dE7
G
H
0.994
0.99
1.19 ± 0.05
1.01 ± 0.48
0.96 ± 0.46
0.83 ± 0.39
1.37 ± 1.01
1.33 ± 0.98
1.29 ± 0.95
I
Chr. 12: 25,295,000
25,445,000
Tibentan
RPS: 12.28
Berkshire
RPS: 7.32
ABI3
TPM: 49.73
TPM: 21.97
TPM: 2.19
TPM: 5.19
J
uE7
uE1
K
L
0.99
0.98
3.41 ± 1.61
3.58 ± 1.68
3.76 ± 1.78
3.87 ± 1.87
4.45 ± 2.09
4.73 ± 2.23
5.05 ± 2.39

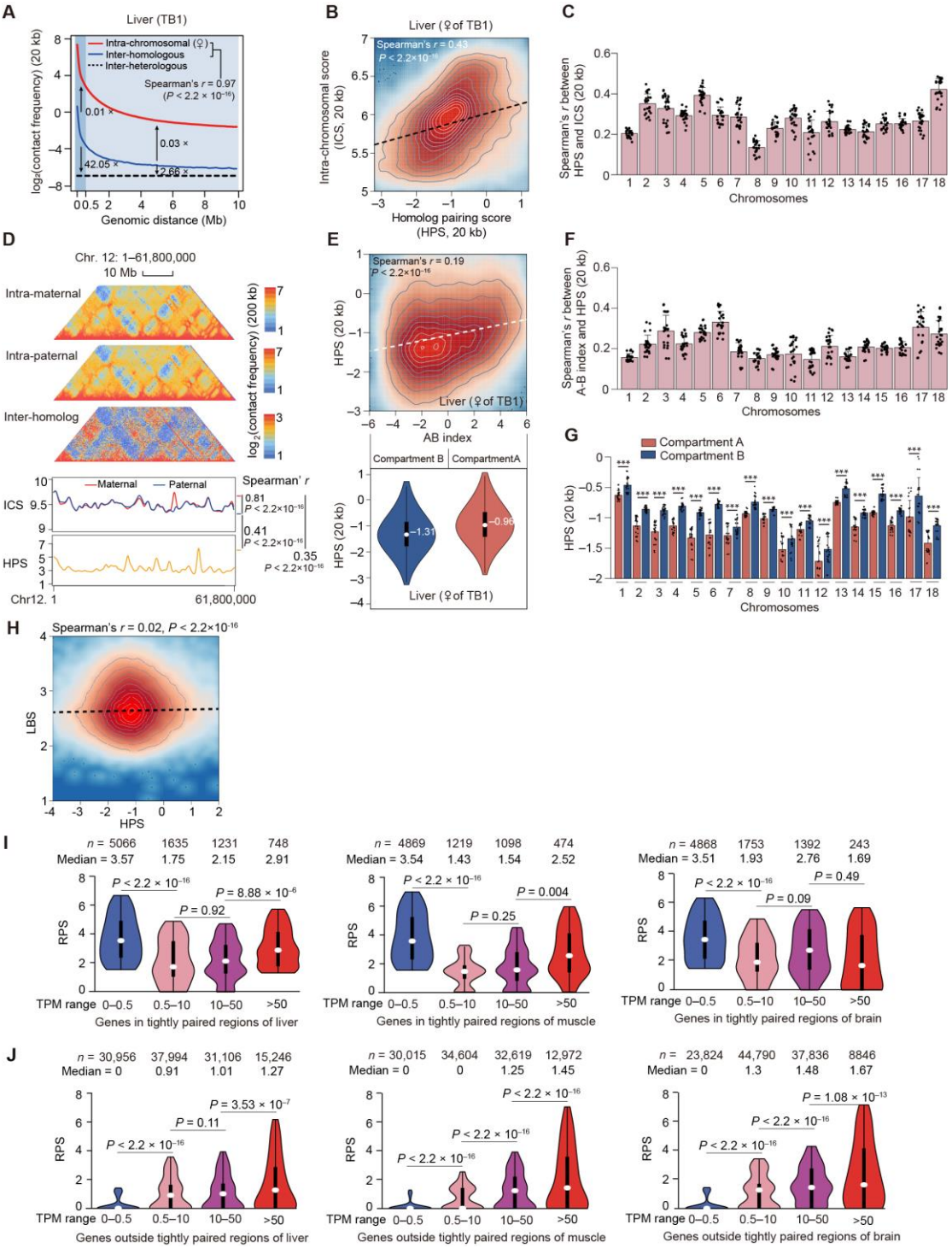


Supplemental Fig. S22. Representative examples of genes showing differential RPS between Berkshire and Tibetan alleles across the different tissues. (**A**–**D**) Allelically differential RPS of *C12ORF29* in the liver. (**A**) Hi-C maps (upper left) and their allelic differences (upper right). ChIP-seq signals of H3K27ac and H3K4me3, gene expression (lower left), and 3D structural models (lower right). Promoter (grey square), enhancers (green squares), and PEIs (connecting lines) are displayed. (**B**) Allelic differences of PEI intensity (mean ±SD, *n* = 4, top) and sequence-based PEI probability (middle and bottom). n.s., $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$. (**C**) Allelic differences of histone modification (H3K27ac in enhancers and H3K4me3 in promoter). Mean ±SD (*n* = 4). n.s., $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$. (**D**) Degree of haplotype similarities (measured by IDS) of promoter and enhancer in pairwise comparisons among 8 haplotypes. The allelic changes in the PEI interval length are also provided. Allelically differential RPS of *CAV2* (**E**–**H**), *ABI3* (**I**–**L**), and *ITGA6* (**M**–**P**) in the liver, and *PFKL* (**Q**–**T**) in the brain are similar to those shown in (**A**–**D**).



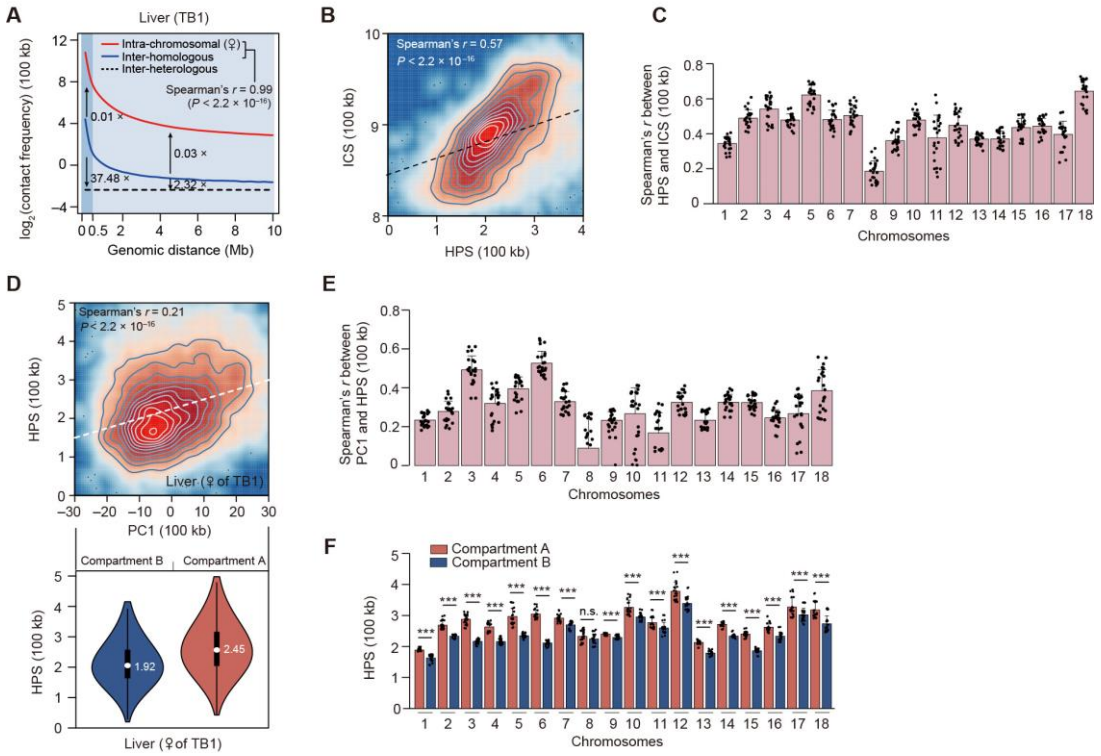
Supplemental Fig. S23. Differential RPS between Berkshire and Tibetan alleles in the skeletal muscle. (**A**) Comparison of myofiber cross-sectional area in representative longissimus dorsi muscle (hematoxylin-and-eosin staining) between adult purebred

Berkshire and Tibetan pigs (see Supplemental Methods). *P*-value is from paired Student's t-test. (**B** and **C**) Spatial transcriptomic profiles of type I myofiber marker *MYH7* expression in succinate dehydrogenase (SDH)-stained histological sections in Tibetan (**B**) and Berkshire pigs (**C**). The proportions of type I myofibers (dark color) were estimated using SDH-stained sections (top). Insets show three typical regions/images magnified for comparison (middle). Differential *MYH7* expression between type I and type II myofibers (masked from SDH-stained sections) was quantified using a 'pseudo-bulk' approach (see Supplemental Methods; bottom). (**D**–**F**) Allelic differences in PEIs and sequence divergence in promoter and enhancer regions of *DDIT4L* in muscle. (**D**) Hi-C maps (upper left) and a heatmap of corresponding allelic differences in PEI intensity (upper right). ChIP-seq signals of H3K27ac, H3K4me3, and gene expression levels (lower left), and 3D structural models of PEIs (lower right). Promoter (grey square), enhancers (green squares), and PEIs (connecting lines) are displayed beside the Hi-C maps. (**E**) Statistical analysis of allelic differences in PEI intensity (mean ± SD, n = 4, top) and sequence-based PEI probability (middle), and H3K27ac and H3K4me3 signals in enhancers and promoter (mean ± SD, n = 4; bottom). Paired Student's t-test, n.s., $P > 0.05$; *$P < 0.05$; **$P < 0.01$. (**F**) Degree of sequence divergence in promoter and enhancer regions among eight haplotypes (measured by IDS; top) and among 82 diploids (measured by IBS), as well as the frequency distribution of SNVs in 82 diploids (middle).

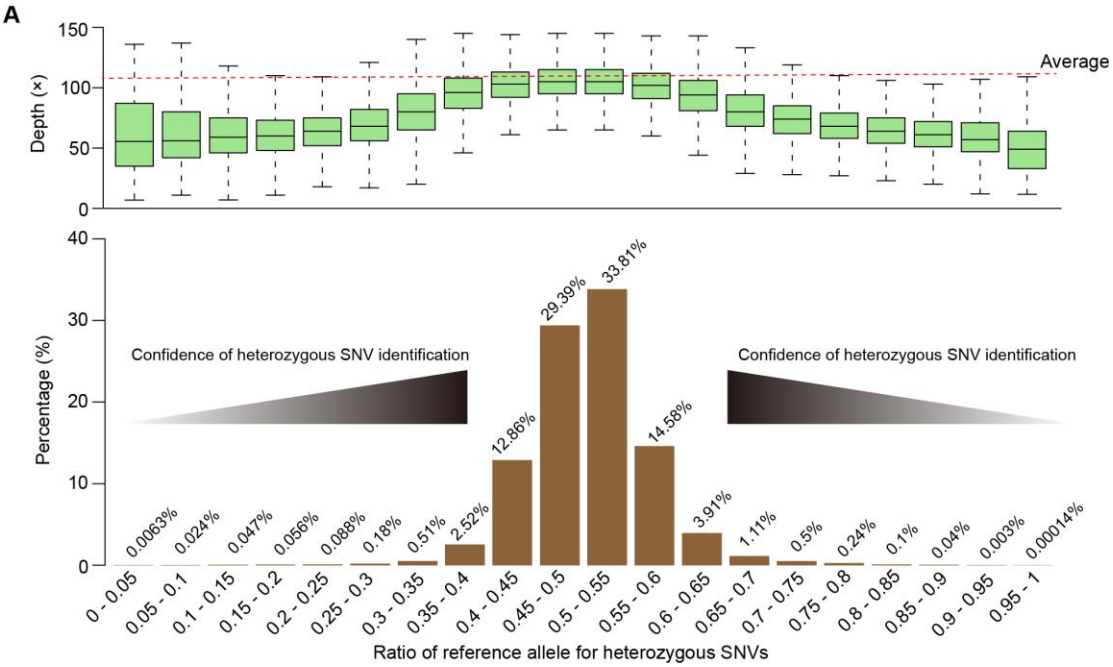


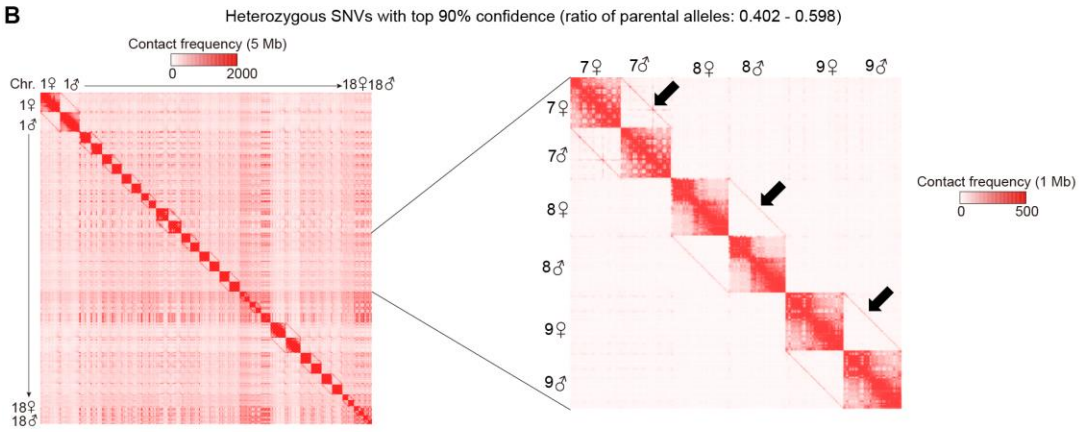
Supplemental Fig. S24. Characteristics of homolog pairing in the pig genome at 20-kb resolution. (**A**) Similar distance-dependent decay of intra-chromosomal (red) and inter-homologous (blue) contacts (Spearman's r = 0.97, P < 2.2×10^{-16}). The interactions between homologs are statistically weaker than within chromosomes (inter-homolog vs. intra-chromosome: average 0.01-fold between 20-kb to 500-kb, P < 0.01; and 0.036-fold between 500-kb to 10-Mb, P < 2.2×10^{-16}), but statistically stronger than between heterologs (dashed line; inter-homolog vs. inter-heterolog: average 35.31-fold between 20-

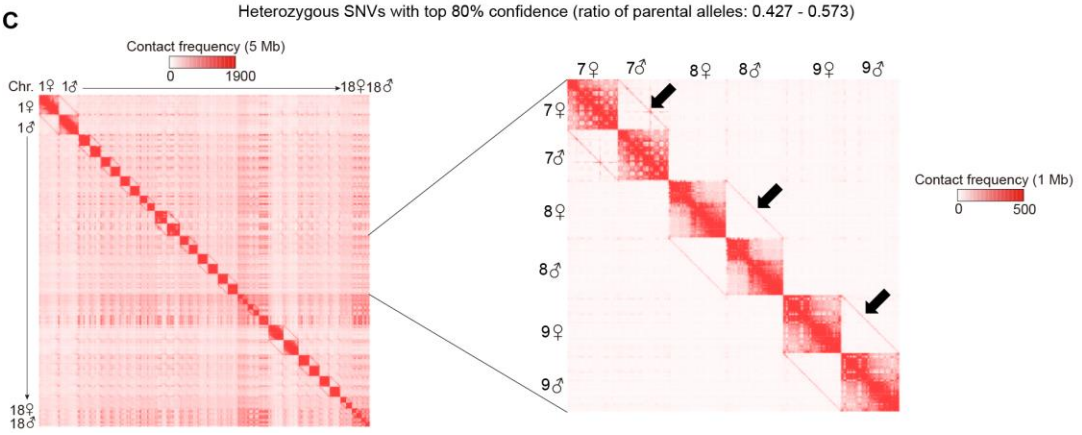
kb to 500-kb, $P < 9.73\times 10^{-16}$; and 2.41-fold between 500-kb to 10-Mb, $P < 2.2\times10^{-16}$) in 12 samples. The *P*-values were obtained using a paired Student's t-test. Data of a representative liver sample is shown. (**B**) Homolog pairing score (HPS) is positively correlated with the intrachromosomal score (ICS; Spearman's $r = 0.43$, $P < 2.2\times10^{-16}$). Data of a representative liver sample is shown. (**C**) Correlation between HPS and ICS for each of the 18 autosomes. (**D**) Homology pairing is correlated with compartmentalization. Top: HPS values are positively correlated with compartment scores (A-B index; Spearman's r = 0.19, $P < 2.2\times10^{-16}$). A representative liver sample is shown. Bottom: in comparison to compartment B regions, compartment A regions exhibited enhanced allelic pairing (reflected by statistically increased HPS values), which supports findings uncovered in flies (AlHaj Abed et al. 2019). (**E**) Correlation between HPS and compartment scores (A-B index) for each of the 18 autosomes. (**F**) Comparison of HPS between compartment A and B regions for each of the 18 autosomes. (**G**) Hi-C maps of Chromosome 12 in the pig genome (61.80 Mb in length) (top). The intra-chromosomal contacts (measured by ICS) of two parental alleles, and inter-homolog contacts (measured by HPS) are also shown (bottom). (**H**) Correlation between HPS and LBS. (**I** and **J**) Compared to other genes (**J**), those located in tightly paired regions (**I**) exhibited decreased covariations between allelic expression and RPS. Genes present in eight haplotypes within each tissue were classified into four classes based on their respective expression levels. This suggests a complicated transcriptional regulation program besides the intra-chromosomal PEI exists. In **C**, **E**, and **F**, the data are presented as mean ± SD ($n = 12$), and the dots represent the Spearman's correlation coefficients for each sample. The *P* values were calculated using a Wilcoxon rank-sum test. *$P < 0.05$; **$P < 0.01$; *** $P < 0.001$.

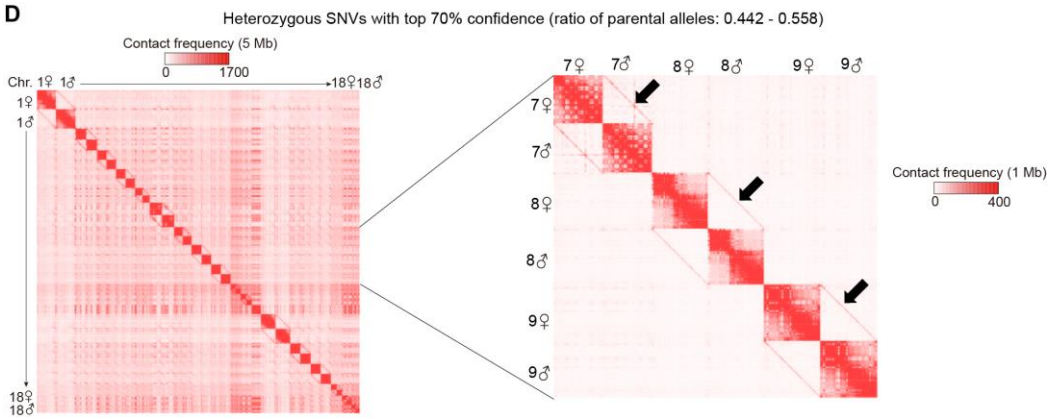


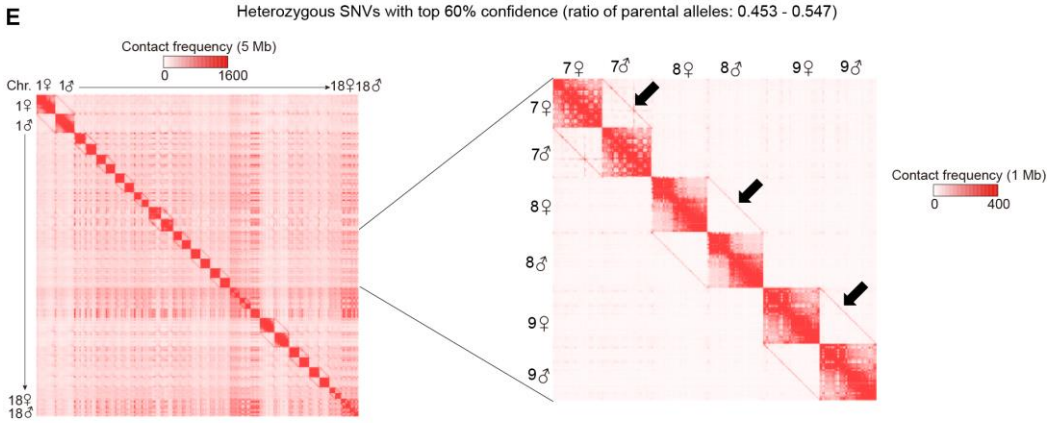
Supplemental Fig. S25. Characteristics of homolog pairing in the pig genome at 100-kb resolution, which recapitulates the findings obtained at 20-kb resolution. (**A**) Similar distance-dependent decay of intra-chromosomal (red) and inter-homologous (blue) contacts (mean Spearman's r = 0.99, $P < 2.2\times10^{-16}$). Although the interactions between homologs are statistically weaker than within chromosomes (inter-homolog versus intrachromosome: average 0.01-fold between 20-kb to 500-kb, $P < 0.12$; and 0.028-fold between 500-kb to 10-Mb, $P < 1.06\times10^{-8}$), they are still statistically stronger than between heterologs (dashed line; inter-homolog versus inter-heterolog: average 37.48-fold between 20-kb to 500-kb, $P < 0.007$; and 2.32-fold between 500-kb to 10-Mb, $P < 2.2\times10^{-16}$). P values were calculated using a paired Student's t-test. Data of a representative liver sample is shown. (**B**) Homolog pairing score (HPS) is positively correlated with the intra-chromosomal score (ICS) (Spearman's r = 0.57, $P < 2.2\times10^{-16}$). A representative liver sample is shown. (**C**) Correlations between HPS and ICS for each of the 18 autosomes. (**D**) Homology pairing is correlated with compartmentalization. Top: HPS values are positively correlated with A-B index (Spearman's r = 0.28, $P < 2.2\times10^{-16}$). A representative liver sample is shown. Bottom: In comparison to compartment B regions, compartment A regions exhibited enhanced allelic pairing (reflected by statistically increased HPS values). (**E**) Correlation between HPS and compartment scores (A-B index) for each of the 18 autosomes. (**F**) Comparison of HPS between compartments A and B regions for each of the 18 autosomes. In **C**, **E**, and **F**, the data are presented as mean ± SD (n = 12), and the dots represent the Spearman's correlation coefficients for every sample. P values were calculated using the Wilcoxon rank-sum test. *$P < 0.05$; **$P < 0.01$; *** $P < 0.001$.

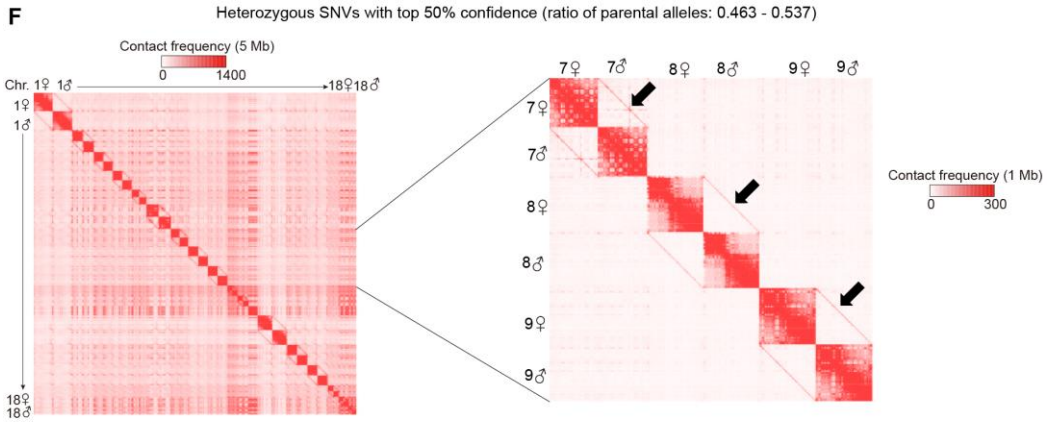

A
Depth (×)
Average
Percentage (%)
Confidence of heterozygous SNV identification
Confidence of heterozygous SNV identification
0.0063%
0.024%
0.047%
0.056%
0.088%
0.18%
0.51%
2.52%
12.86%
29.39%
33.81%
14.58%
3.91%
1.11%
0.5%
0.24%
0.1%
0.04%
0.003%
0.00014%
0 - 0.05
0.05 - 0.1
0.1 - 0.15
0.15 - 0.2
0.2 - 0.25
0.25 - 0.3
0.3 - 0.35
0.35 - 0.4
0.4 - 0.45
0.45 - 0.5
0.5 - 0.55
0.55 - 0.6
0.6 - 0.65
0.65 - 0.7
0.7 - 0.75
0.75 - 0.8
0.8 - 0.85
0.85 - 0.9
0.9 - 0.95
0.95 - 1
Ratio of reference allele for heterozygous SNVs

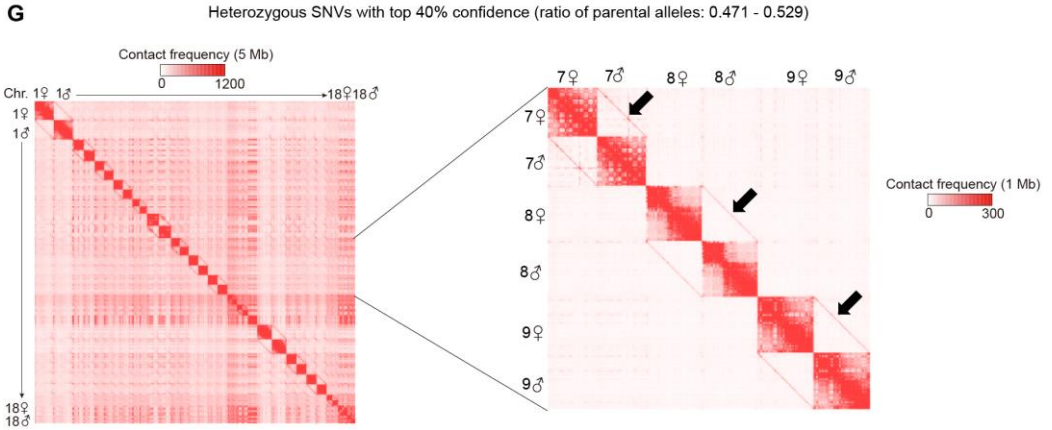

B
Heterozygous SNVs with top 90% confidence (ratio of parental alleles: 0.402 - 0.598)
Contact frequency (5 Mb)
Contact frequency (1 Mb)

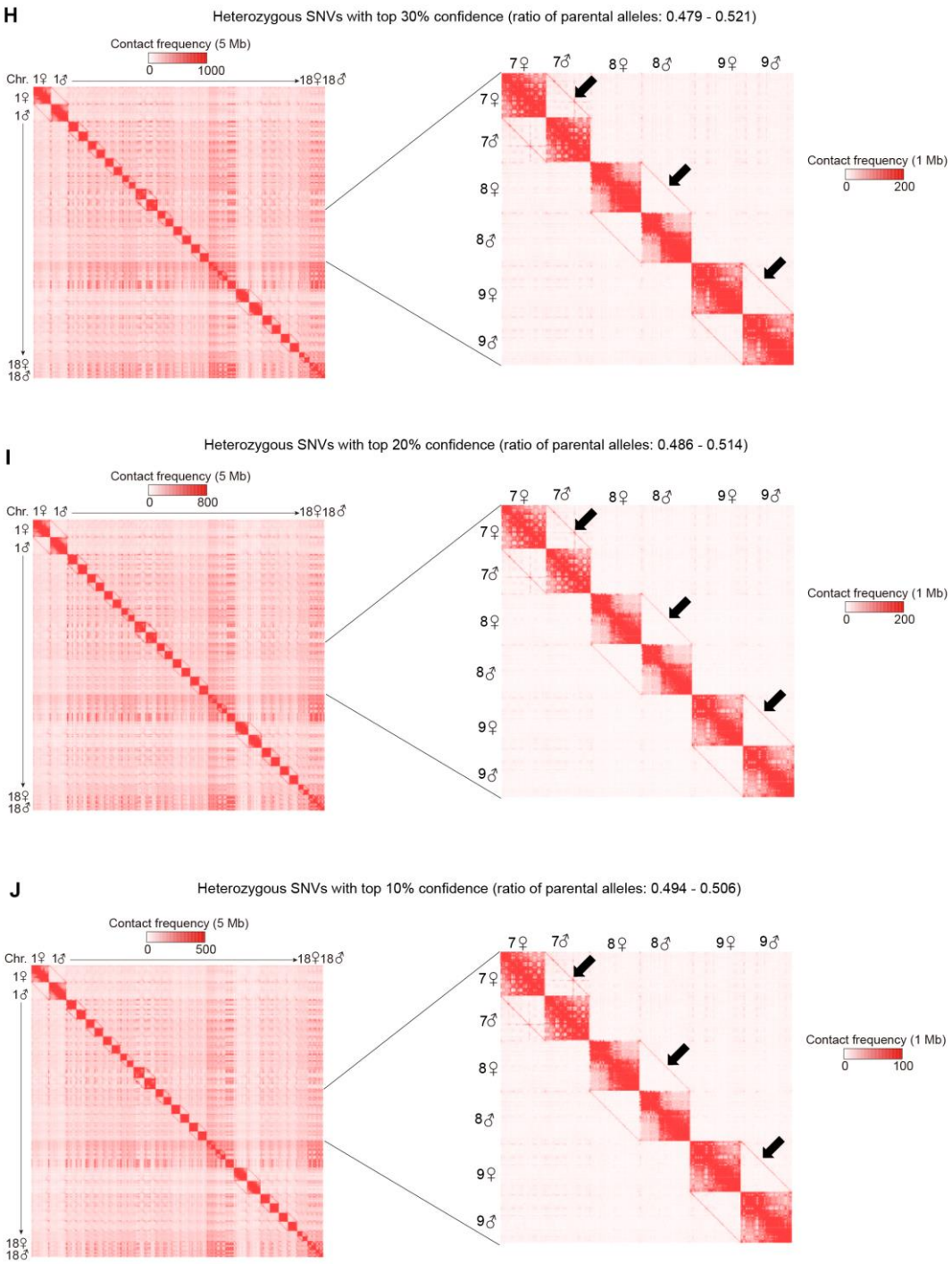

C
Heterozygous SNVs with top 80% confidence (ratio of parental alleles: 0.427 - 0.573)
Contact frequency (5 Mb)
Contact frequency (1 Mb)


D
Heterozygous SNVs with top 70% confidence (ratio of parental alleles: 0.442 - 0.558)
Contact frequency (5 Mb)
0
1700
Chr. 1♀ 1♂
18♀18♂
Contact frequency (1 Mb)
0
400

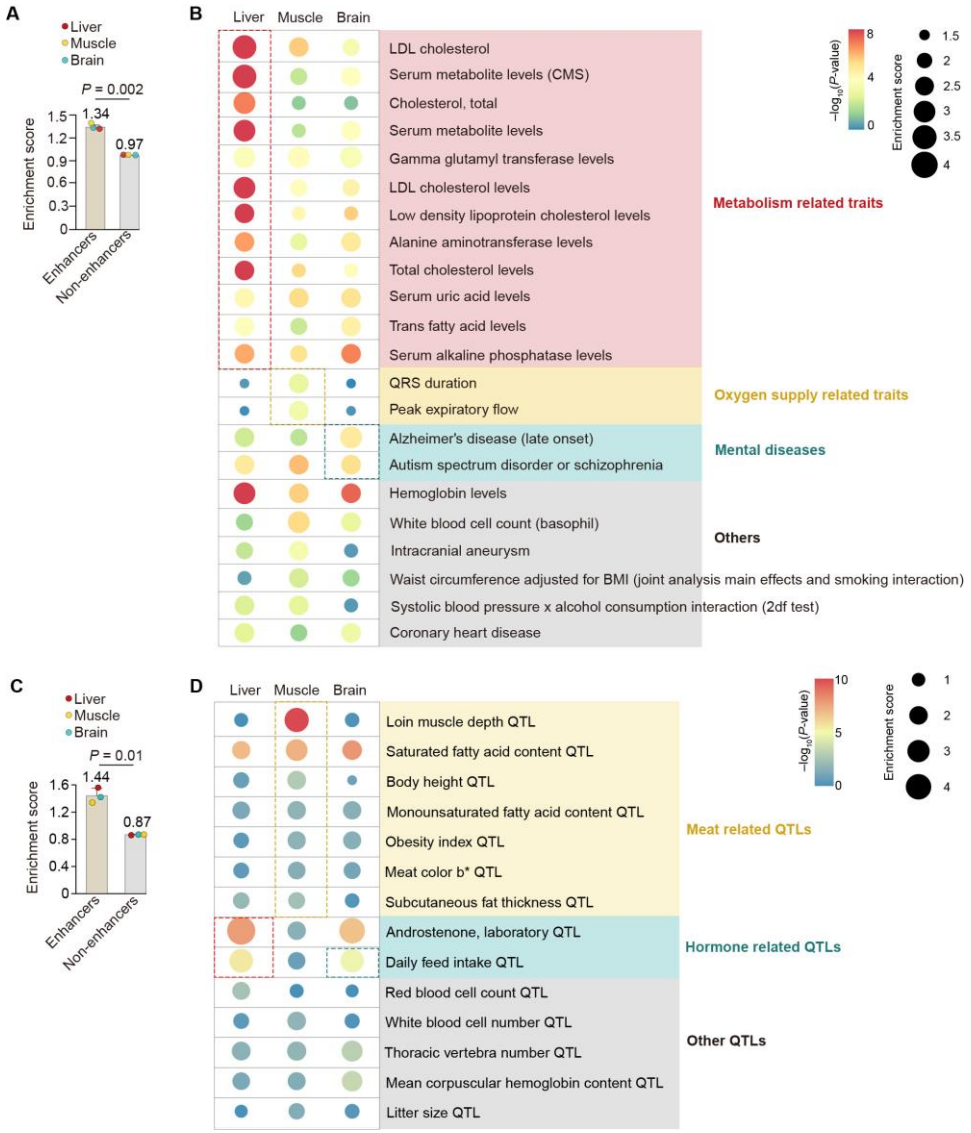

E
Heterozygous SNVs with top 60% confidence (ratio of parental alleles: 0.453 - 0.547)
Contact frequency (5 Mb)
0
1600
Chr. 1♀ 1♂
18♀18♂
Contact frequency (1 Mb)
0
400


F
Heterozygous SNVs with top 50% confidence (ratio of parental alleles: 0.463 - 0.537)
Contact frequency (5 Mb)
0
1400
Chr. 1♀ 1♂
18♀18♂
Contact frequency (1 Mb)
0
300


G
Heterozygous SNVs with top 40% confidence (ratio of parental alleles: 0.471 - 0.529)
Contact frequency (5 Mb)
0
1200
Chr. 1♀ 1♂
18♀18♂
Contact frequency (1 Mb)
0
300



Supplemental Fig. S26. Detection of homolog pairing with heterozygous SNVs at different confidence levels to assign parental origins of chromatin interactions. (**A**) The ratio of parental alleles for each heterozygous SNV. All heterozygous SNVs were divided into 20 intervals. Top: distribution of sequencing depths for SNVs within each interval. Bottom: proportion of SNVs within each interval. (**B**–**J**) Hi-C maps showing signals of genome-wide homolog pairing (arrows) in a representative somatic tissue (liver) of a hybrid pig sample based on heterozygous SNVs at different confidence levels.



Supplemental Fig. S27. Human GWAS and pig QTLs associated with specific traits or diseases were enriched in enhancers. (**A**) Comparison of enrichment for human trait-associated, noncoding SNPs (nucleotides that can be liftover from the human to the pig genome) in enhancers and other regions of the pig genome. (**B**) Bubble plot showing the enrichment of noncoding SNPs in enhancers that were separately identified in the liver, muscle, and brain. The traits or diseases shown were merged from the ten terms with the highest enrichment scores in each tissue ($P < 0.05$, χ^2 test), and can be empirically classified into four categories, including metabolism, oxygen supply, mental diseases, and others. (**C**) Comparison of enrichment for pig QTLs (~5 bp in length; Release 50; Apr 25, 2023; https://www.animalgenome.org/cgi-bin/QTLdb/SS/index) in enhancers and other regions of the pig genome. (**D**) Bubble plot showing the enrichment of pig QTLs in enhancers that were separately identified in the liver, muscle, and brain. The traits shown were merged from the terms with the highest enrichment scores in each tissue ($P < 0.05$, χ^2 test), and can be empirically classified into three categories, including meat, hormone and others.

Supplemental Methods

1. Generation of trio-based genomic and Hi-C data from F1 hybrids

1.1 Whole-genome sequencing and data processing

Genomic DNA was exacted from ear (F0, *n* = 12) and liver tissues (F1, *n* = 6) using the TIANamp Genomic DNA Kit (TIANGEN, DP304). Sequencing libraries were generated and sequenced on the Illumina HiSeq X ten platform with 150-bp paired-end reads, or on the BGISEQ-500 platform with 100-bp paired-end reads. High-quality whole-genome sequencing data were aligned to the pig reference genome (*Sscrofa* 11.1) using the Burrows-Wheeler Aligner (BWA, v 0.7.8; Li and Durbin 2009). Optical and PCR duplicates were removed using Picard MarkDuplicates (v 2.0.1, http://broadinstitute.github.io/picard). SNVs and indels were called using the Genome Analysis Toolkit (GATK, v 3.8; McKenna et al. 2010) HaplotypeCaller and stored in genomic variant call format (gVCF). For each trio, variant data from single sample gVCF files were aggregated into a multi-sample VCF file using GVCFGenotyper. Low-quality variants or genotypes were excluded using GATK, with the arguments 'QD < 10.0 || FS > 60.0 || MQ < 40.0 || MQRankSum < –12.5 || ReadPosRankSum < –8.0 || GQ < 30'. Unplaced scaffolds, sex chromosomes and the mitochondrial genome were removed from further analyses. For each trio, we discarded variants with the lowest (bottom ~1%) and highest (top ~1%) coverage depth (depth-based Z-score < –2.58 and > 2.58, respectively). We merged the SNV data for all individuals, and performed PCA using the GCTA software (v 1.93.2; Yang et al. 2011), and genetic structure inference using Structure (v 2.3.4; Pritchard et al. 2000).

1.2 *In situ* Hi-C library preparation and sequencing

We constructed 8–15 *in situ* Hi-C libraries (technical replicates) for each of the 14 samples, with minor modifications from previously described method (Rao et al. 2014). Compromising sample size and sequencing cost, we obtained a higher volume of Hi-C data (~7.29 billion contacts per sample, dataset 1 and 2) for each of the four skeletal muscle and four brain tissue samples compared to the six liver tissue samples (~3.15 billion contacts per sample, dataset 1) (**Fig. 1B**; **Supplemental Fig. S2B**). Briefly, the tissues were homogenized and fixed with a 4% formaldehyde solution at room temperature for 30 min. The chromatin was digested with 200 U of DpnII enzyme (R0543S, NEB, USA) at 37°C for 90 min, 65°C for 20 min and 25°C for 5 min. Nucleotide fill-in was conducted with 0.4 mM Biotin-14-dATP (19524-016, Invitrogen), 10 mM dCTP, 10 mM dGTP, 10 mM dTTP and 5 $U \cdot \mu L^{-1}$ Klenow Fragment (M0210L, NEB) at 37°C for 45 min. Ligation was performed by a T4 DNA ligase (L6030-HC-L, Enzymatics, USA) at 20°C for 30 min. DNA was sheared to the length of 300 to 500 bp and washed using M280 beads at 20°C for 20 min. The Hi-C libraries were amplified with 10 PCR cycles and sequenced with 150-bp paired-end

reads on the Illumina HiSeq X ten platform or 100-bp paired-end reads on the BGISEQ-500 platform. Compromising sample size and sequencing cost, we obtained a higher volume of Hi-C data (~7.29 billion contacts per sample, dataset 1 and 2) for each of the four skeletal muscle and four brain tissue samples compared to the six liver tissue samples (~3.15 billion contacts per sample, dataset 1) (**Fig. 1B**; **Supplemental Fig. S2B**)

2. Reconstruction and analysis of haplotype-resolved Hi-C maps

2.1 Reconstruction of haplotype-resolved Hi-C maps

We reconstructed haploid maternal and paternal Hi-C maps at a maximum resolution of 2-kb resolution for a total of 14 F1 samples (Hi-C dataset 1) and 1-kb resolution for 12 F1 samples (Hi-C datasets 1 and 2). Hi-C matrices are two-dimensional and thus the number of lattices increase exponentially with increasing resolution. For instance, the number of lattices increases 6.25-fold from 5-kb ($n = 2.12 \times 10^{11}$) to 2-kb ($n = 5.29 \times 10^{12}$), although the resolution only increases by 2.5-fold (see **Supplemental Fig. S5D**). Therefore, as resolution increases in Hi-C analysis, the resulting matrix will be sparser due to a higher percentage of bin pairs with low counts (see **Supplemental Fig. S5D**), potentially amplifying background noise in the data and reducing statistical power. Hence, for subsequent Hi-C analyses, we used a lower, more appropriate resolution (20-kb for A/B compartments and TADs, and 5-kb for PEIs) to ensure both sufficient statistical power and precision. Data from X Chromosomes were excluded from further analyses to avoid confounding factors related to the 'mosaic' 3D structural features of active and inactive X Chromosomes in tissues with highly heterogeneous cell populations (*i.e.*, different cell populations have different, random, X Chromosome inactivation; Deng et al. 2014).

For a diploid sample, we employed the SNPsplit (v 0.3.4; Krueger and Andrews 2016) to classify Hi-C contacts into three categories, which were termed informative (both reads containing haplotype-resolved SNVs), partial-informative (either read containing haplotype-resolved SNVs) and non-informative (neither read containing haplotype-resolved SNVs). The informative contacts provide the most accurate information to reconstruct a haplotype-resolved Hi-C map, but their application is limited by their relatively low proportion (8.19–12.03%; **Supplemental Fig. S4A**). According to their parental origins, the informative contacts can be classified into six categories (**Supplemental Fig. S4B**). Notably, we observed that only a small fraction (1.65–2.82%) of unphased intra-chromosomal contacts (including actual intra- and inter-homologous contacts) is inter-homologous. This indicates that for any given intra- or inter-homologous contact, the probability of it originating from intra-maternal or intra-paternal category ranged from 97.18% to 98.35%. This allowed us to directly classify the partial-informative intra- or inter-homologous contacts to either the intra-maternal or intra-paternal category with acceptable misclassification rates (1.65–2.82%). Nonetheless, these haplotype-resolved intra-

chromosomal contacts were highly correlated with SNV density (Spearman's r = 0.93, $P < 2.2\times10^{-16}$) due to non-uniform SNV distribution. To address this, we employed HaploHiC (v 0.32; Lindsly et al. 2021) to phase the parental origin of non-informative intra- or inter-homologous contacts with local imputation based on the distribution of informative and partial-informative intra-chromosomal contacts. This allowed us to generate 28 nearly complete haplotype-resolved intra-chromosomal Hi-C maps (comprising ~99.37% intra- or inter-homologous contacts for each sample). In contrast, inter-heterologs contacts occupy ~50.20% of nonhomologous contacts. Accordingly, the partial-informative and non-informative nonhomologous contacts cannot be assigned due to a high misclassification rate, with only ~9.63% of nonhomologous contacts were successfully assigned to their parental origins.

To make the Hi-C data sets comparable, we normalized the Hi-C matrixes at 5-kb, 20-kb and 100-kb resolutions using a BNBC (band-wise normalization and batch correction) quantile normalization strategy (Fletez-Brant et al. 2021). In addition, we validated the normalization efficiency using the multiHiCcompare (Stansfield et al. 2019) for Hi-C matrixes of Chromosome 18.

2.2 Correlations between intra-chromosomal matrices

We used HiCRep (Yang et al. 2017a), GenomeDISCO (Ursu et al. 2018), and QuASR-Rep (Yardımcı et al. 2019) with default parameters to assess the reproducibility of normalized intra-chromosomal contact matrices (at 20-kb resolution) across all haplotypes (n = 28).

2.3 3D modelling of diploid pig genomes

We reconstructed 3D genome organization for each sample based on the normalized intra- (at 20-kb resolution) and inter-chromosomal (at 1-Mb resolution) contact matrices for 18 pairs of homologous autosomes using an approximation of multidimensional scaling (MDS) method implemented in the miniMDS (Rieber and Mahony 2017) program. The software PYMOL (The PyMOL Molecular Graphics System, v 2.5.2 Schrödinger, LLC.) was used for visualization. This analysis was applied to the Hi-C dataset 1 (see **Supplemental Fig. S1**).

2.4 Allelically compartmental rearrangements

2.4.1 Identification of A/B compartments

Haplotype-resolved A/B compartments at 20-kb resolution were identified using both principal component analysis (PCA) and A-B index, as previously described (Rowley et al. 2017). Briefly, PCA was performed to generate PC1 vectors for each autosome per sample at 100-kb resolution. Spearman's r between PC1 and genomic characteristics including gene density and GC content were then calculated. Bins with positive Spearman's r were

defined as compartments A, and the remainder as compartments B. The A-B index was then calculated as previously described (Rowley et al. 2017) at 20-kb resolution, which represents the likelihood of a genomic segment interacting with the A or B compartments defined at 100-kb resolution, as described above. Bins of 20-kb length with positive or negative A-B index were considered as either A or B compartments, respectively. The reproducibility of A/B compartments between the two alleles was assessed using Pearson' *r* correlation based on the A-B index values. This analysis was applied to the Hi-C dataset 1 (see **Supplemental Fig. S1**).

2.4.2 A/B compartment switches and variables

To identify the distinct compartment status (*i.e.*, A/B switches) between haplotypes in different categories (*i.e.*, tissues, parental-of-origins, and parental breeds), we defined a set of common A/B compartments (with more than 75% of haplotypes exhibiting the same chromatin status) for each category. Considering the number of biological replicates differed among categories (i.e., n = 12 for liver and n = 8 for skeletal muscle and brain in analyses identifying A/B switches between pairwise tissues; or n = 6 for liver and n = 4 for skeletal muscle and brain in analyses of A/B switches between parents-of-origin and parental breeds), we therefore used 75% as a strict threshold, rather than absolute number of replicates, to define the chromatin state for each biological category. For each tissue, the common A/B compartments were defined if the same chromatin status was presented in more than 9 haplotypes for liver, and more than 6 haplotypes for skeletal muscle and brain. For categories of parental-of-origin and paternal breeds, the common A/B compartments were defined if the same chromatin status was present in more than 5 haplotypes for liver, and at least 3 haplotypes for both skeletal muscle and brain tissues. Thus, A/B switches referred to the genomic regions with different common compartment status between haplotypes in different categories. We also identified A/B compartment switches using HOMER (http://homer.ucsd.edu/homer/; Lin et al. 2012) to validate the reliability of the switched compartments detected using the above-mentioned method.

In addition, we recognized regions with the same compartment status between haplotypes but have statistically significant differences in compartment scores (*i.e.*, the A-B index values) ($|\Delta$A-B index$| > 0.5$ and $P < 0.05$, paired Students' *t*-test) between parent-of-origins and paternal breeds (termed as A/B variables). Both analyses were applied to the Hi-C dataset 1 (**Supplemental Fig. S1**).

2.5 Allelically variable topologically associated domains (TADs)

2.5.1 TAD calling

Haplotype-resolved TADs were identified at 20-kb resolution using the Directionality Index (DI; Rowley et al. 2017) and the Insulation Index (IS; Crane et al. 2015) as previously

described. Briefly, the DI value of each 20-kb bin was calculated using the number of reads that map from a given bin to the upstream and downstream 2-Mb regions (Dixon et al. 2012). A hidden Markov model (HMM) was then applied to the DI values to infer domains and anchors. In addition, the IS value was calculated and normalized for each 20-kb bin (Crane et al. 2015). Bins with minimal IS along the normalized IS vector were interpreted as the TAD anchors. Finally, large TADs identified by DI were further split into small TADs based on IS, and then the two sets of TADs were merged for further analyses. This analysis was applied to the Hi-C dataset 1 (see **Supplemental Fig. S1**).

2.5.2 Measurement of TAD concordance

MoC (assessment of the overlap between each pair of TADs by measuring in number of base pairs and considering the overall size of both TADs; Zufferey et al. 2018) and VI (measurement of the similarity of all subsets of the two TAD structures using a dynamic programming algorithm to compute the VI metrics; Sauerwald and Kingsford 2018) were calculated to access the reproducibility of TADs between haplotypes.

2.5.3 Identification of TAD boundary shifts

TAD boundary was defined as the anchor bin along with its 60-kb flanking segments and two boundaries were merged as a larger boundary if they were overlapped. We identified the shifts of haplotype-resolved TAD boundaries between tissues, parent-of-origin and parental breeds as the bins containing changed boundary positions that exhibited significantly different local boundary scores (LBS; Han et al. 2020); quantitatively reflecting the strength of TAD boundary).

Specifically, 15 continuous 20-kb bins formed a locus (locus = 300 kb). For a given bin, the interactions within its up- and downstream loci (represented as 300 kb × 300 kb triangles in the contact matrix) and right loci were defined as its intra-loci interactions. The interactions between these two loci (300 kb × 300 kb diamonds in the contact matrix) are inter-loci iterations. The $\log_2$-transformed ratio of intra- to inter-loci interactions was calculated as LBS of the bin.

We further used a custom method modified from our previous work (Li et al. 2012) to identify regions with statistically differential LBS (D-LBS). Briefly, we scanned the genome from 5' to 3' and, if the difference between the two haplotype groups for a 20-kb bin was significant ($P < 0.05$, paired Students' *t*-test), then that bin was considered as the seed site of a candidate D-LBS region. After this, a 3' downstream adjacent 20-kb bin was concatenated to this seed site, and the average LBS of these two sites was subjected to another round of paired Students' *t*-test. The same process was repeated for the next 20-kb bin until a low-variance 20-kb bin was identified ($P > 0.05$). In order to eliminate 'trailing smear' (*i.e.*, a low variance bin being incorporated into a D-LBS region due to very high

variance of its preceding bins), we repeated the above test procedures from 3' to 5' across the genome. If a genomic region containing three or more bins have statistically significant ($P < 0.05$) different LBS across the haplotypes, then this region was considered as a D-LBS region. This analysis was applied to the Hi-C dataset 1 (see **Supplemental Fig. S1**).

Additionally, we used a dedicated tool for TAD boundary analysis, TADCompare (https://github.com/dozmorovlab/TADCompare), to calculate TAD boundary scores (in which high values indicate TAD boundaries) for the shifted boundaries identified between parents-of-origin using the above-mentioned method, to assess the reliability of our results.

2.6 Haplotype-resolved interrogation of promoter-enhancer interactions (PEIs)

2.6.1 Identification of PEIs

To reliably identify haplotype-resolved PEIs, we performed additional *in situ* Hi-C assays (dataset 2) for 12 of the 14 aforementioned samples (three tissues from four F1 pigs, see **Supplemental Fig. S1**). By combining both Hi-C datasets 1 and 2, we obtained very high-resolution diploid Hi-C maps (a maximum resolution of 1-kb). We generated KR-normalized intra-chromosomal contact maps at 5-kb resolution for each haplotype, and further quantile normalized them across 24 haplotypes. The promoter regions of genes were defined as the 5-kb bin covering the transcription start site (TSS). The normalized contact maps at 5-kb resolution were split into smaller matrices (20 Mb × 20 Mb) with a step size of 10 Mb to accelerate the identification of PEIs using the PSYCHIC algorithm (Ron et al. 2017) with default parameters. PEI intensity is calculated as follows:

PEI intensity = observed contact frequency – expected contact frequency

in which the observed contact frequency is the number of promoter-centered interactions within 10 Mb genomic distance upstream and downstream, and the expected contact frequency is defined as the normalized contact frequency according to the domain-specific background model calculated by PSYCHIC software.

We reserved high-confidence haplotype-resolved PEIs using the following parameters: (i) FDR ≤ 10^{-4}; (ii) interaction distances ≥ 25-kb; (iii) for each tissue, more than two occurrences identified in the same parental breed or parent-of-origin; (iv) for each tissue, more than three occurrences identified across all haplotypes. To determine the appropriate FDR threshold for PEI identification, we calculated the average PEI number per gene using eight different FDR thresholds (**Supplemental Fig. S13A**), and found that the threshold of 10^{-4} can ensure maximum PEI discovery with a low false positive rate.

2.6.2 Calculation of regulatory potential scores

To explore the regulatory effects of multiple enhancers on a gene, we calculated regulatory

potential score (RPS) for each gene as previously reported (Zhi et al. 2022). The RPS was calculated as,

$$\sum_{n=1}^{N} log_{10}\,(I_n + 1)$$

where I_n is the distance-normalized interaction intensity of the n^{th} enhancer (*i.e.*, the observed contacts minus the expected contacts). If a promoter does not interact with an enhancer, then the RPS is set to zero. This analysis was applied to aggregation of the Hi-C datasets 1 and 2 (**Supplemental Fig. S1**). As we observed that pairwise haplotypes within each tissue shared a higher correlation than pairwise tissues (**Supplemental Fig. S7N-P**), the PEIs identified in each haplotype were merged to calculate RPS, consequently allowing the identification of differential RPS genes between tissues, parents-of-origin, and breeds using a paired Student's *t*-test.

3. Allele-specific gene expression

3.1 rRNA-depleted RNA-seq library preparation and sequencing

Total RNA was extracted from 14 samples used for *in situ* Hi-C assay using the RNeasy Mini Kit (Qiagen). We used an rRNA depletion protocol (Ribo-Zero kit, Epicentre) coupled with the Illumina TruSeq RNA-seq library protocol to construct the strand-specific RNA-seq libraries. All libraries were quantified using the Qubit dsDNA High Sensitivity Assay Kit (Invitrogen) and sequenced with 150-bp paired-end reads on the Illumina HiSeq X Ten or 100-bp paired-end reads on the BGISEQ-500 platform.

3.2 Quantitation of allele-specific gene expression

We quantified the allelic expression of ~11,430 protein-coding genes that were covered by informative SNVs and had evidence of transcription (transcripts per million [TPM] ≥ 0.5 in at least one allele) using the Allelome.PRO (Andergassen et al. 2015). The Kallisto (v 0.44.0) software (Bray et al. 2016) was used to quantify the total gene-level expression for both alleles of 19,328 autosomal protein-coding genes as transcripts per million (TPM). Briefly, high-quality RNA-seq reads were aligned to the pig reference genome (*Sscrofa* 11.1) using STAR (v 2.6.0c; Dobin et al. 2013) with parameters '--outSAMattributes NH HI NM MD --alignEndsType EndToEnd'. We sought to quantify the allele-specific expression of 19,328 autosomal genes, ~88.35% of which contain haplotype-resolved exonic SNVs for each F1 hybrid. We employed the Allelome.PRO (Andergassen et al. 2015) to distinguish the parental origin of uniquely mapped reads that covered haplotype-resolved exonic SNVs. For each diploid sample, we measured the number of reads assigned to each haplotype of a gene, and calculated the allelic ratio for the maternal or paternal haplotype, when the total number of reads assigned to a gene was larger than 10. We

obtained allele-specific TPM by multiplying the maternal against paternal allelic ratios and the total TPM for both alleles. In addition, the TPM of non- and low-expressed (TPM < 0.5) genes was divided by 2 and assigned to each haplotype.

Although the Western and Chinese pigs showed large genetic divergence when the Duroc reference genome (Sscrofa11.1) was used as a framework for analyses, no reference bias in allele-specific expression was detected between Berkshire and Tibetan haplotypes, as estimated by aligning RNA-seq reads against the reference genome and against the N-masked genome (**Supplemental Fig. S5H**).

Similarities in gene expression were assessed by Spearman's correlation coefficients among haplotypes (n = 28). The combined similarities in gene expression ('function') and chromatin architecture ('form') among haplotypes (n = 28) were estimated using a chromosome phase portrait approach (Liu et al. 2018).

4. Histone modifications on allelic PEI rewiring

4.1 ChIP-seq library preparation and sequencing

We performed ChIP-seq using antibodies against H3K27ac (a canonical histone marker of active enhancers) and H3K4me3 (a histone marker of promoter activation) for 12 samples (liver, skeletal muscle, and brain from each of four F1 hybrids). The ChIP-seq experiments were performed as previously described (Han et al. 2019). Briefly, the chromatin was prepared from formaldehyde fixed tissues and fragmented with a sonicator to an average fragment size of 200–500 bp. Half of the soluble chromatin was stored at –20°C as input for DNA sequencing and the remaining was used for immunoprecipitation reaction with H3K27ac (ab4729, Abcam) and H3K4me3 (9751, CST) antibodies. We also carried out ChIP-seq using antibodies against CTCF (A13272, Abclonal) for liver samples from the four F1 hybrids. For both input DNA and immunoprecipitated DNA, each ChIP-seq library was sequenced on an Illumina HiSeq X Ten platform to generate 150-bp paired-end reads.

4.2 ChIP-seq data processing

High-quality ChIP-seq data were aligned to the pig reference genome (*Sscrofa* 11.1) using BWA (v 0.7.8) and the potential PCR duplicates removed using Picard's MarkDuplicates (v 2.0.1). The aligned single sample ChIP-seq data were aggregated into a multi-sample file for each tissue, and the H3K27ac, H3K4me3 and CTCF peaks identified using MACS2 (Q-value < 0.05; Feng et al. 2012). Highly and moderately active enhancers were identified using the standard ROSE algorithm (Lovén et al. 2013; Whyte et al. 2013). Briefly, neighboring H3K27ac peak regions within 12.5-kb of one another were stitched together, ranked by increasing H3K27ac signal and visualized as a curve (with x-axis representing the rank and y-axis representing the signal); We identified the x-axis point for which a line

with a slope of 1 was tangent to the curve, and this inflection point split all enhancers to highly active enhancers (above the point) and moderately active enhancers (below the point). IGV (v 2.3.91; Thorvaldsdóttir et al. 2013) was used to visualize the ChIP-seq signals, gene expression data and gene locations.

4.3 Allele-specific activity of promoters and enhancers

The allele-specific activity of promoters and enhancers for each sample were qualified by separately assigning the parental origin of H3K4me3 and H3K27ac reads. Specifically, we first identified ~12,772 active promoters (5-kb bins containing TSS) for each tissue. We then employed the Allelome.PRO algorithm to distinguish the parental origin of the uniquely mapped H3K4me3 reads that covered haplotype-resolved SNVs. We measured the number of reads assigned to each haplotype of a promoter, and calculated the allelic ratio for the maternal or paternal haplotype (estimated as maternal or paternal against the total read count), when the total number of reads assigned to a promoter was larger than 10.

All putative enhancers (5-kb in length) involved in PEIs were classified into three categories, including ~2380 highly active enhancers (covered by the H3K27ac peak), ~7476 moderate active enhancers (covered by the H3K27ac peak) and ~26,538 lower active enhancers (not covered by the H3K27ac peak) for each tissue. For each highly and moderately active enhancer, we employed Allelome.PRO to distinguish the parental origin of the uniquely mapped H3K27ac reads that covered haplotype-resolved SNVs. We measured the number of reads assigned to each haplotype of an enhancer, and calculated the allelic ratio for the maternal or paternal haplotype, when the total number of reads assigned to an enhancer was larger than 10.

5. Large-scale indel calling using long-read DNA sequencing

We performed long-read sequencing (Oxford Nanopore Technologies) on four F1 hybrids. The PromethION platform was used for sequencing according to the effective concentration of the DNA library and data output requirements. High-quality long-read DNA sequencing data was obtained by removal of low-quality or short reads (< 1-kb). For each F1 hybrid, the high-quality reads were assigned to each parental haplotype using the parental reads of whole-genome sequencing data, based on a trio binning algorithm (Koren et al. 2018) as implemented in the the Canu software (v 1.8). The haplotype-resolved long reads were first error-corrected using the NECAT software (v 0.0.1; Chen et al. 2021) and then aligned to the pig reference genome using minimap2 (v 2.17; Li 2018). We used a local assembly strategy to identify large-scale indels. Briefly, the reference genome regions were split into 60-kb sliding windows with 20-kb increments. For each 60-kb bin, the aligned reads were retrieved and assembled using the NECAT software. The assembled contigs were then mapped to the reference genome using NGMLR (v 0.2.7; Sedlazeck et al. 2018)

and the high-confidence large-scale indels were identified using Sniffles (v 1.0.11; Sedlazeck et al. 2018) with at least two supporting contigs.

6. Genetic variations on allelic PEI rewiring

6.1 Calculation of identify score (IDS) and identity-by-state (IBS)

To measure the degree of sequence divergence among haplotypes, we calculated the identify score (IDS) to assess the pairwise similarity between haplotypes. For a given genomic region (*e.g.*, 20-kb or 5-kb bin), we estimated

$$IDS = 1 - (Diff_{nuc}/Total_{nuc})$$

where the $Diff_{nuc}$ was defined as the number of different nucleotides (combining the number of SNVs and the length of indels) between the two haplotypes, and $Total_{nuc}$ was defined as the total nucleotide number for a given region. Notably, insertions occurring in either haplotype increased the value of $Total_{nuc}$, whereas deletions occurring in both haplotypes had an opposite effect.

Degree of sequence similarity for promoters and enhancers in pairwise comparisons among the Berkshire and Tibetan purebred pigs (n = 82) was measured by pairwise identity-by-state (IBS) distances. In this population-level analysis, 12 are from the six trios in this study (Berkshire [n = 6] and Tibetan [n = 6]) and 70 are publicly available (Berkshire [n = 15] and Tibetan [n = 55]). Publicly available SNVs were retrieved from the ISwine database (http://iswine.iomics.pro/pig-iqgs/iqgs/index).

6.2 Prediction of PEIs based on sequence features alone

To quantitatively measure the disruptive effects of variants embedded in enhancers and promoters on the formation of PEIs, we assessed the probability of allelic PEIs based solely on their respective sequence features using the PEP algorithm with minor modifications (Yang et al. 2017b).

Specifically, we first built a supervised classification model based on gradient tree boosting to predict promoter-enhancer interactions using only sequences for each haplotype-resolved PEI dataset. We included the 24 haplotype-resolved PEI dataset of eight haplotypes derived from four F1 hybrids for liver, skeletal muscle, and brain. For each positive dataset, we generated a set of non-interacting promoter-enhancer pairs by randomly pairing putative promoters and enhancers, and sampled a negative set with the same size as the positive samples from these simulated non-interacting pairs. We also ensured that the selected non-interacting pairs were not detected by Hi-C and followed the same distance constraints between the positive pairs (Whalen et al. 2016). Supplied with the reference genome assembly and the relevant genetic variants (SNVs, short- and large-

scale indels) for each haplotype, we first reconstructed a haplotype-specific genome (termed as 'pseudo-genome'). We then created a map from each 5-kb enhancer or 2.7-kb promoter region of the reference genome to the haplotype-specific pseudo-genome. The haplotype-resolved interacting and non-interacting promoter-enhancer pairs with both ends aligned from the reference genome to the respective haplotype-specific pseudo-genome were retained for further analyses.

We applied FIMO (Grant et al. 2011) for scanning motifs along the pseudo-genome sequences of promoters and enhancers with the PWMs from HOCOMOCO Human v10 database (Kulakovskiy et al. 2016). This allowed us to identify motif matches using a *P*-value threshold of 1×10^{-4}. For a certain motif, we calculated the normalized motif occurrences for each of the enhancer or promoter sequences, used these frequencies for all motifs as feature vectors, and concatenated the feature vectors of the enhancer region and the paired promoter region to form the feature representation of an enhancer-promoter pair (Yang et al. 2017b).

We used XGBoost (Chen et al, 2016), a scalable and highly accurate implementation of gradient boosting, as a predicting algorithm to predict PEIs based on feature representation generated by motif occurrences using the R package XGBoostExplainer (v 0.1), which makes XGBoost as transparent as a single decision tree. We obtained the predicted classification represented as a 0–1 probability of samples with known labels by performing one-fold cross validations. Samples with unknown labels were classified based on a complete model built from all the samples with known labels.

6.3 Simulation of PEI intensity based on interval lengths

To test the hypothesis promoters and enhancers in closer proximity on the linear genome tend to be spatially closer (thus with the elevated PEI intensity), and *vice versa*, we simulated the PEI intensities of 24 haplotypes based on the allelically variable bridging distance between promoter and enhancer (consequence of incorporating short- and large-scale indels) using the Huynh's algorithm (Huynh and Hormozdiari 2019). Specifically, we used the opposite haplotype within the same individual as reference (termed 'H matrix'), to predict a new Hi-C contact matrix for the haplotype of interest (termed 'G matrix'). To this end, we first modeled contact frequencies of the H matrix by introducing the parameters *β, α* and *r*, and the following equation

$$H_{i,j}^{'} = \frac{e^{(\alpha_i+\alpha_j)/2} L_{original}^{\beta}}{e^{r_{i+1}} e^{r_{i+2}} \ldots e^{r_j}}$$

where *α* captures the genomic properties (*e.g.*, GC content, mappability) of each bin; *β* models the power-law scaling of contact frequencies based on the genomic distance; *r* represents the putative existence of an insulator within the specific bin; $H_{i,j}^{'}$ indicates the

simulated contact frequency between bin_i and bin_j; and $L_{original}$ denotes the linear genomic distance between bin_i and bin_j in the pig reference genome. The unknown parameters were estimated through optimization. We assumed that deletions and insertions either decrease or increase the genomic distances, eliminate or add insulators, and do not alter the *α, β,* or *r* values. We thus estimated the contact frequency of the G matrix as follows:

$$G_{i,j} = \frac{e^{(\alpha_i+\alpha_j)/2} L_{new}^{\beta}}{e^{r_{i+1}} \ldots e^{r_{x_1-1}} e^{r_{y_1+1}} \ldots e^{r_{x_N-1}} e^{r_{y_N+1}} \ldots e^{r_j}}$$

where x_n , y_n (*n* = 1~*N*) denotes the start and end coordinates (in the unit of bin number) for *N* deletions between bin_i and bin_j in the haplotype of interest relative to the pig reference genome; L_{new} represents the linear genomic distance between bin_i and bin_j considering the effects of all deletions and insertions in the haplotype of interest relative to the reference haplotype. It should be noted that insertions in the reference haplotype relative to the reference genome were considered as deletions in the haplotype of interest relative to the reference haplotype, and *vice versa*. Since we were not able to estimate *r* values for inserted sequences with respect to the reference genome, we set *r* values to zero in these bins.

7. Morphological measurements and spatial transcriptomics of skeletal muscles

7.1 H & E and SDH staining of skeletal muscles

The *longissimus dorsi* muscle from purebred Berkshire and Tibetan pigs (6-month-old, *n* = 2 for each breed) were collected and fixed in a 10% neutral buffered formalin solution, embedded in paraffin using TP1020 semi-enclosed tissue processor (Leica), sliced at a thickness of 6 μm using RM2135 rotary microtome (Leica), and finally stained with hematoxylin and eosin (H & E). The myofiber cross-sectional area was measured as an average of 100 fibers in randomly selected fields.

The proportion of types I (dark staining) and II (light staining) myofibers was assessed using the succinate dehydrogenase (SDH) staining. Briefly, the isolated samples were embedded with the OCT solution (TissueTek, Japan), and then frozen in –80 °C. The 10 μm frozen sections were made with a Leica frozen microtome, and stained with a SDH staining kit (Beijing Solarbio Science & Technology Co., Ltd.).

7.2 Spatial transcriptomic assays

To verify the classification of types I and II myofibers in *longissimus dorsi* muscle using the histochemical SDH staining approach, we dissected the transcriptional differences of myofiber-specific markers (*MYH7* for type I) using spatial transcriptomics (ST) (10X Genomics) as previously described (Jin et al. 2021). Briefly, fresh muscle tissues were

sliced into roughly 6.5 mm × 6.5 mm × 1 cm pieces and snap-frozen in liquid nitrogen. Tissue samples were embedded using OCT media at –80°C. For the ST analysis, the samples were sectioned on a cryostat at a thickness of 15 μm, and each section was placed within a capture area on a Visium Spatial slide, which was permeabilized according to the protocol provided by 10X Genomics for 18 min. Fragmented and barcoded RNA was used as the carrier material. The spike-in constituted approximately 25% of the libraries. ST cDNA libraries were diluted to 4 nM and sequenced on the Illumina NovaSeq 6000 platform with paired-end sequencing reads.

7.3 ST data processing and analysis

The spatial transcriptomics sequencing data of skeletal muscle were analyzed as previously described (Jin et al. 2021). Briefly, we first manually aligned the image to the slide area and removed unreliable spots using the 10X Genomics Loupe Browser (v 4.1.0). Sample demultiplexing and expression quantification were carried out using Space Ranger (v 1.1) according to the 10X Genomics spatial gene expression analysis pipeline. We retrieved the reference genome sequence and gene annotation file (GTF) from Ensembl *Sscrofa* 11.1 (Release 102). Read alignment and gene expression quantification was performed using Space Ranger. We further applied the Seurat (v 3.2; Stuart et al. 2019) package for removing low-quality spots or non-expressed genes, normalization (regularized negative binomial regression, 'SCTransform' function) of gene expression data (Hafemeister and Satija 2019), and spatial visualization of features. To avoid sparsity and greatly increase unique molecular index (UMI) coverage of genes, we performed a 'pseudo-bulk' approach. We manually grouped spots into types I and II muscle fiber clusters according to the SDH staining results, summed the UMI counts for 100 randomly sampled spots from each myofiber cluster, and generated four replicates of 'pseudo-bulk' expression matrices for each cluster. The normalization and differential expression analysis of the pseudo-bulk data was performed as previously described (Maynard et al. 2021) using scater (v 1.24.0; McCarthy et al. 2017) and limma (v 3.52.0; Ritchie et al. 2015).

8. Collection of imprinted genes in the pig genome

We first retrieved a total of 36 imprinted genes that were previously identified in the pig genome from the genomic imprinting website (https://geneimprint.com/site/home). In addition, we also retrieved the previously reported imprinted genes from another 11 mammalian genomes, including cow, dog, human, laboratory opossum, mouse, opossum, rabbit, rat, rhesus macaque, sheep, and wallaby (see **Supplemental Fig. S11**). These imprinted genes were then assigned to a total of 90 ortholog genes in the pig genome using ortholog information downloaded from the Ensembl website, resulting in a total of 126

imprinted genes.

9. Functional enrichment analysis

Functional enrichment analysis was performed using the Metascape tool (Zhou et al. 2019) with default parameters. Genes in the pig genome were converted to human orthologs, which were used as inputs for the enrichment analysis. Human (*Homo sapiens*) was chosen as the target species, and enrichment analysis was performed against all genes in the genome as the background set, with Gene Ontology-biological processes (GO-BP) and KEGG as the test sets. The statistically significant terms were selected as outputs.

10. Trait-associated SNP enrichment analysis

We downloaded 146,690 unique human trait-associated SNPs from the NHGRI-EBI GWAS Catalog (https://www.ebi.ac.uk/gwas/, last access on June 1, 2021; Buniello et al. 2019). These SNPs were assigned to 77,917 loci in the pig reference genome (*Sscrofa* 11.1) using the UCSC liftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver). A total of 73,363 noncoding SNPs (or 94.16%) linked to 4,514 traits or diseases were used for subsequent analyses. We compared the enrichment scores (*i.e.*, relative density) of noncoding SNPs between enhancers and other regions. The significance of the enrichment score was calculated using a χ^2 test for each trait or disease that contained more than 50 SNPs.

Supplemental References

AlHaj Abed J, Erceg J, Goloborodko A, Nguyen SC, McCole RB, Saylor W, Fudenberg G, Lajoie BR, Dekker J, Mirny LA, et al. 2019. Highly structured homolog pairing reflects functional organization of the Drosophila genome. *Nat Commun* **10**: 4485. doi:10.1038/s41467-019-12208-3.

Andergassen D, Dotter CP, Kulinski TM, Guenzl PM, Bammer PC, Barlow DP, Pauler FM, Hudson QJ. 2015. Allelome.PRO, a pipeline to define allele-specific genomic features from high-throughput sequencing data. *Nucleic Acids Res* **43**: e146. doi: 10.1093/nar/gkv727.

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**: 525-527. doi:10.1038/nbt.3519.

Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, et al. 2019. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* **47**: D1005-d1012. doi:10.1093/nar/gky1120.

Chen, T., Guestrin, C. 2016. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. New York, NY, USA: ACM, 785–794 (2016).

Chen Y, Nie F, Xie SQ, Zheng YF, Dai Q, Bray T, Wang YX, Xing JF, Huang ZJ, Wang DP,

et al. 2021. Efficient assembly of nanopore reads via highly accurate and intact error correction. *Nat Commun* **12**: 60. doi:10.1038/s41467-020-20236-7.

Crane E, Bian Q, McCord RP, Lajoie BR, Wheeler BS, Ralston EJ, Uzawa S, Dekker J, Meyer BJ. 2015. Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**: 240-244. doi:10.1038/nature14450.

de Wit E. 2017. Capturing heterogeneity: single-cell structures of the 3D genome. *Nat Struct Mol Biol* **24**: 437-438. doi:10.1038/nsmb.3404.

Deng Q, Ramsköld D, Reinius B, Sandberg R. 2014. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**: 193-196. doi:10.1126/science.1245316.

Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. 2012. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**: 376-380. doi:10.1038/nature11082.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15-21. doi:10.1093/bioinformatics/bts635.

Edge P, Bafna V, Bansal V. 2017. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. *Genome Res* **27**: 801-812. doi:10.1101/gr.213462.116.

Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc* **7**: 1728-1740. doi:10.1038/nprot.2012.101.

Fletez-Brant K, Qiu Y, Gorkin DU, Hu M, Hansen KD. 2021. Removing unwanted variation between samples in Hi-C experiments. *bioRxiv* **214361**. doi:10.1101/214361

Fu Y, Xu J, Tang Z, Wang L, Yin D, Fan Y, Zhang D, Deng F, Zhang Y, Zhang H, et al. 2020. A gene prioritization method based on a swine multi-omics knowledgebase and a deep learning model. *Commun Biol* **3**: 502. doi:10.1038/s42003-020-01233-4.

Grant CE, Bailey TL, Noble WS. 2011. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**: 1017-1018. doi:10.1093/bioinformatics/btr064.

Hafemeister C, Satija R. 2019. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**: 296. doi:10.1186/s13059-019-1874-1.

Han K, Ren R, Cao J, Zhao S, Yu M. 2019. Genome-wide identification of histone modifications involved in placental development in pigs. *Front Genet* **10**: 277. doi:10.3389/fgene.2019.00277.

Han Z, Cui K, Placek K, Hong N, Lin C, Chen W, Zhao K, Jin W. 2020. Diploid genome architecture revealed by multi-omic data of hybrid mice. *Genome Res* **30**: 1097-1106. doi:10.1101/gr.257568.119.

Hoencamp C, Dudchenko O, Elbatsh AMO, Brahmachari S, Raaijmakers JA, van Schaik T, Cacciatore AS, Contessoto VG, van Heesbeen R, van den Broek B, et al. 2021. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. *Science* **372**: 984-989. doi:10.1126/science.abe2218.

Huynh L, Hormozdiari F. 2019. TAD fusion score: discovery and ranking the contribution of deletions to genome structure. *Genome Biol* **20**: 60. doi:10.1186/s13059-019-1666-7.

Jin L, Tang Q, Hu S, Chen Z, Zhou X, Zeng B, Wang Y, He M, Li Y, Gui L, et al. 2021. A pig BodyMap transcriptome reveals diverse tissue physiologies and evolutionary dynamics of transcription. *Nat Commun* **12**: 3715. doi:10.1038/s41467-021-23560-8.

Koren S, Rhie A, Walenz BP, Dilthey AT, Bickhart DM, Kingan SB, Hiendleder S, Williams JL, Smith TPL, Phillippy AM. 2018. *De novo* assembly of haplotype-resolved genomes with trio binning. *Nat Biotechnol* **36**: 1174-1182. doi:10.1038/nbt.4277.

Krueger F, Andrews SR. 2016. SNPsplit: Allele-specific splitting of alignments between genomes with known SNP genotypes. *F1000Res* **5**: 1479. doi:10.12688/f1000research.9037.2.

Kulakovskiy IV, Vorontsov IE, Yevshin IS, Soboleva AV, Kasianov AS, Ashoor H, Ba-Alawi W, Bajic VB, Medvedeva YA, Kolpakov FA, et al. 2016. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. *Nucleic Acids Res* **44**: D116-125. doi:10.1093/nar/gkv1249.

Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094-3100. doi:10.1093/bioinformatics/bty191.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**: 1754-1760. doi:10.1093/bioinformatics/btp324.

Li M, Wu H, Luo Z, Xia Y, Guan J, Wang T, Gu Y, Chen L, Zhang K, Ma J, et al. 2012. An atlas of DNA methylomes in porcine adipose and muscle tissues. *Nat Commun* **3**: 850. doi:10.1038/ncomms1854.

Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, Chandra V, Bossen C, Glass CK, Murre C. 2012. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat Immunol* **13**:1196-204.

Lindsly S, Jia W, Chen H, Liu S, Ronquist S, Chen C, Wen X, Stansbury C, Dotson GA, Ryan C, et al. 2021. Functional organization of the maternal and paternal human 4D Nucleome. *iScience* **24**: 103452. doi:10.1016/j.isci.2021.103452.

Liu S, Chen H, Ronquist S, Seaman L, Ceglia N, Meixner W, Chen PY, Higgins G, Baldi P, Smale S, et al. 2018. Genome architecture mediates transcriptional control of human myogenic reprogramming. *iScience* **6**: 232-246. doi:10.1016/j.isci.2018.08.002.

Lin YC, Benner C, Mansson R, Heinz S, Miyazaki K, Miyazaki M, Chandra V, Bossen C, Glass CK, Murre C. 2012. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nat Immunol*. **13**: 1196-1204. doi:10.1038/ni.2432.

Lovén J, Hoke HA, Lin CY, Lau A, Orlando DA, Vakoc CR, Bradner JE, Lee TI, Young RA. 2013. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**: 320-334. doi:10.1016/j.cell.2013.03.036.

Maynard KR, Collado-Torres L, Weber LM, Uytingco C, Barry BK, Williams SR, Catallini JL, 2nd, Tran MN, Besich Z, Tippani M, et al. 2021. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* **24**: 425-436. doi:10.1038/s41593-020-00787-0.

McCarthy DJ, Campbell KR, Lun AT, Wills QF. 2017. Scater: pre-processing, quality control,

normalization and visualization of single-cell RNA-seq data in R. *Bioinformatics* **33**: 1179-1186. doi:10.1093/bioinformatics/btw777.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297-1303. doi:10.1101/gr.107524.110.

Pritchard JK, Stephens M, Donnelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945-959. doi:10.1093/genetics/155.2.945.

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, et al. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**: 1665-1680. doi:10.1016/j.cell.2014.11.021.

Rieber L, Mahony S. 2017. miniMDS: 3D structural inference from high-resolution Hi-C data. *Bioinformatics* **33**: i261-i266. doi:10.1093/bioinformatics/btx271.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**: e47. doi:10.1093/nar/gkv007.

Ron G, Globerson Y, Moran D, Kaplan T. 2017. Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nat Commun* **8**: 2237. doi:10.1038/s41467-017-02386-3.

Rowley MJ, Nichols MH, Lyu X, Ando-Kuri M, Rivera ISM, Hermetz K, Wang P, Ruan Y, Corces VG. 2017. Evolutionarily Conserved principles predict 3D chromatin organization. *Mol Cell* **67**: 837-852.e837. doi:10.1016/j.molcel.2017.07.022.

Sauerwald N, Kingsford C. 2018. Quantifying the similarity of topological domains across normal and cancer human cell types. *Bioinformatics* **34**: i475-i483. doi:10.1093/bioinformatics/bty265.

Sedlazeck FJ, Rescheneder P, Smolka M, Fang H, Nattestad M, von Haeseler A, Schatz MC. 2018. Accurate detection of complex structural variations using single-molecule sequencing. *Nat Methods* **15**: 461-468. doi:10.1038/s41592-018-0001-7.

Stansfield JC, Cresswell KG, Dozmorov MG. 2019. multiHiCcompare: joint normalization and comparative analysis of complex Hi-C experiments. *Bioinformatics* **35**:2916-2923.

Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y, Stoeckius M, Smibert P, Satija R. 2019. Comprehensive integration of single-cell data. *Cell* **177**: 1888-1902.e1821. doi:10.1016/j.cell.2019.05.031.

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178-192. doi:10.1093/bib/bbs017.

Ursu O, Boley N, Taranova M, Wang YXR, Yardimci GG, Stafford Noble W, Kundaje A. 2018. GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs. *Bioinformatics* **34**: 2701-2707. doi:10.1093/bioinformatics/bty164.

Whalen S, Truty RM, Pollard KS. 2016. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. *Nat Genet* **48**: 488-496. doi:10.1038/ng.3539.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA. 2013. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**: 307-319. doi:10.1016/j.cell.2013.03.035.

Yang J, Lee SH, Goddard ME, Visscher PM. 2011. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**: 76-82. doi:10.1016/j.ajhg.2010.11.011.

Yang T, Zhang F, Yardımcı GG, Song F, Hardison RC, Noble WS, Yue F, Li Q. 2017a. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res* **27**: 1939-1949. doi:10.1101/gr.220640.117.

Yang Y, Zhang R, Singh S, Ma J. 2017b. Exploiting sequence-based features for predicting enhancer-promoter interactions. *Bioinformatics* **33**: i252-i260. doi:10.1093/bioinformatics/btx257.

Yardımcı GG, Ozadam H, Sauria MEG, Ursu O, Yan KK, Yang T, Chakraborty A, Kaul A, Lajoie BR, Song F, et al. 2019. Measuring the reproducibility and quality of Hi-C data. *Genome Biol* **20**: 57. doi:10.1186/s13059-019-1658-7.

Zhi ML, Zhang JY, Tang QZ, Yu DW, Gao S, Gao DF, Liu PL, Guo JX, Hai T, Gao J, et al. 2022. Generation and characterization of stable pig pregastrulation epiblast stem cell lines. *Cell Res* **32**: 383-400. doi:10.1038/s41422-021-00592-9.

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. 2019. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Commun* **10**: 1523. doi:10.1038/s41467-019-09234-6.

Zufferey M, Tavernari D, Oricchio E, Ciriello G. 2018. Comparison of computational methods for the identification of topologically associating domains. *Genome Biol* **19**: 217. doi:10.1186/s13059-018-1596-9.