Allele-specific Effects of Three-dimensional Genome Architecture in Hybrid Pigs

Yu Lin
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
https://orcid.org/0000-0003-4796-8409

Jing Li (lijing_2020@sicau.edu.cn)
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Yiren Gu
Animal Breeding and Genetics Key Laboratory of Sichuan Province, Sichuan Animal Science Academy, Chengdu, 610066, China

Long Jin
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Jingyi Bai
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Jiaman Zhang
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Yujie Wang
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Pengliang Liu
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Keren Long
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Mengnan He
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Diyan Li
School of Pharmacy, Chengdu University, Chengdu, 610106, China

Can Liu
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
Ziyin Han
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
Yu Zhang
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
Xiaokai Li
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
Bo Zeng
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
Lu Lu
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
Fanli Kong
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
Ying Sun
Institute of Geriatric health, Sichuan Provincial People's Hospital, University of Electronic Science and Technology of China, Chengdu, 610072, China
Yongliang Fan
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
Xun Wang
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
Tao Wang
School of Pharmacy, Chengdu University, Chengdu, 610106, China
An'an Jiang
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
Jideng Ma
Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China
linyuan Shen

Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Li Zhu

Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Yanzhi Jiang

Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Guoqing Tang

Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Qingyou Liu

Animal Molecular Design and Precise Breeding Key Laboratory of Guangdong Province, School of Life Science and Engineering, Foshan University, Foshan, 528225, China

Hua Li

Animal Molecular Design and Precise Breeding Key Laboratory of Guangdong Province, School of Life Science and Engineering, Foshan University, Foshan, 528225, China

Jinyong Wang

Pig Industry Sciences Key Laboratory of Ministry of Agriculture and Rural Affairs, Chongqing Academy of Animal Sciences, Chongqing, 402460, China

Liangpeng Ge

Pig Industry Sciences Key Laboratory of Ministry of Agriculture and Rural Affairs, Chongqing Academy of Animal Sciences, Chongqing, 402460, China

Xuewei Li

Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Qianzi Tang（✉ tangqianzi@sicau.edu.cn）

Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

Mingzhou Li（✉ mingzhou.li@sicau.edu.cn）

Livestock and Poultry Multi-omics Key Laboratory of Ministry of Agriculture and Rural Affairs, College of Animal Science and Technology, Sichuan Agricultural University, Chengdu, 611130, China

https://orcid.org/0000-0001-8681-7684

Research Article

Keywords: haplotype-specific chromatin conformation, homolog pairing, imprinting, sequence variants, gene regulation

Posted Date: December 21st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2392032/v2

Abstract

In diploid mammals, allele-specific three-dimensional (3D) genome architecture may lead to imbalanced gene expression. Through ultradeep *in situ* Hi-C sequencing of three representative somatic tissues (liver, skeletal muscle, and brain) from hybrid pigs generated by reciprocal crosses of phenotypically and physiologically divergent Berkshire and Tibetan pigs, we uncover extensive chromatin reorganization between homologous chromosomes across multiple scales. Haplotype-based interrogation of multi-omics data revealed the tissue-dependence of 3D chromatin conformation, suggesting that parent-of-origin-specific conformation may drive gene imprinting. We quantify the effects of genetic variations and histone modifications on allelic rewiring of long-range promoter-enhancer contacts, which likely contribute to the dramatic phenotypic differences between the parental pig breeds. This study also provides definitive evidence of structured homolog pairing in the pig genome which could facilitate regulatory interactions between homologous chromosomes. This work illustrates how allele-specific chromatin architecture facilitates concomitant shifts in allele-biased gene expression, and consequently phenotypic changes in mammals.

Introduction

It is well understood that the vast majority of somatic cells in mammals inherit two haploid genomes that unequally contribute to cellular function. Substantial differences have been identified between homologous chromosomes (homologs) in gene transcription[1], DNA methylation[2], and chromatin accessibility[3,4]. In addition, higher-order chromatin structure is now emerging as an important regulator of gene expression[5]. Advances in high-throughput chromatin conformation capture (Hi-C) and its derivatives have led to revelations in the full scope of three-dimensional (3D) chromatin dynamics during mammalian development and lineage specification[6]. However, conventional Hi-C data cannot generally distinguish between different allelic copies, and most previous studies infer an average chromatin architecture that incorporates the diploid genome without considering possible differences between homologs[7]. Variability in multi-level chromatin architecture between homologs remains largely unexplored; characterizing the 3D genome organization of a mammalian diploid genome remains challenging[8–11].

Sus scrofa (*i.e.*, pig or swine) is a primary source of calories for humans; the world's pig population reached almost 1 billion, with China alone accounting for 49 percent[12] Moreover, pig is rapidly emerging as a versatile and informative biomedical model for human developmental processes[13,14] and complex diseases[15,16], in addition to its utility in vaccine[17] and drug design[18] due to the relatively close anatomical[19,20], physiological[21,22], immunological[23] and genetic [24–26] similarities with humans. The ability to generate genome-editing mutations in pig combined with somatic nuclear cloning procedures has resulted in a number of new models for various human diseases[27] and suggests the potential use of pigs as a source for xenotransplantation[28]. Extensive genomic divergence is known to have occurred between European and Asian pigs, attributable to the relative isolation of European and Asian lineages (at

least one million years ago and their independent domestication in multiple locations across Eurasia in the past ~10,000 years[25]. Since animal breeding became more organized in the 18th century, and especially the strong genetic selection of inbred commercial lines within the last 70 years[25], many of these breeds exhibit remarkable phenotypic diversity and genetic adaptations that provide a sufficient number segregating sites to distinguish allele-specific information[29].

To reveal the role of allelic bias in chromatin architecture in transcriptional regulation, we introduce a pig model generated by reciprocal cross of highly selected European Berkshire and Asian aboriginal Tibetan pigs. Using these F1 hybrid progenies, we reconstructed the haplotype-resolved 3D structures of 14 diploid genomes using ultra-deep *in situ* Hi-C sequencing (Fig. 1a and **Supplementary Fig. 1a**). These Hi-C maps are then used to characterize the 3D nuclear organization of the diploid genomes and highlight features of the spatial organization of homologs. Further analysis of the haplotype-resolved Hi-C maps enabled systematic investigation of the tissue-, parent-of-origin-, and parental breed-specific differences in chromatin architecture at multiple scales. We also performed a genome-wide, allele-specific survey of the effects of genetic variations and histone modifications on the allelic reorganization of promoter-enhancer interactions (PEIs) linked to allelic imbalances in gene expression in these hybrid pigs. Ultimately, this study uncovers definitive evidence of the colocalization of homologs (known as 'homolog pairing') in the genome of somatic tissues, which is correlated with genomic compartmentalization and which represents a likely additional layer of transcriptional regulation between homologs.

Results

Construction of high-resolution diploid Hi-C contact maps

To generate chromosome-span haplotypes for use in assigning 'omics' data from diploid genomes to their parental origins, we carried out reciprocal crosses of two genetically distinct pig breeds (three families of Tibetan [ ] × Berkshire [ ], and three families of Berkshire [ ] × Tibetan [ ]) and sequenced the whole-genomes of six parent-child trios (~ 122.25 × coverage for each individual, $n = 18$) (Fig. 1a).

To understand the 3D structural basis of transcriptional control in a diploid genome, we performed *in situ* Hi-C for three tissues from newborn female filial 1 (F1), representing the three germ layers: the liver (endoderm, $n = 6$), the brain (ectoderm, $n = 4$), and skeletal muscle (mesoderm, $n = 4$). In total, ~ 48.95 billion valid Hi-C contacts (dataset 1: ~3.50 billion for each sample, $n = 14$) were generated, and the transcriptional profiles of corresponding samples were obtained using RNA-seq (Fig. 1a, **Supplementary Fig. 1a** and **Supplementary Fig. 2**). 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)[30].

Combining trio-based genomic and Hi-C data from F1 hybrids enabled construction of chromosome-span haplotypes that included 99.33% of all heterozygous single nucleotide variants (SNVs; ~4.54 per kb) for

each F1 hybrid (**Supplementary Fig. 3**; see Methods). Through SNV phasing combined with local imputation (**Supplementary Fig. 4**; see Methods), we built diploid Hi-C maps for 14 samples based on the unambiguous maternal and paternal contacts, each containing ~ 1.04 billion intra-chromosomal contacts (with maximum resolution of 2-kb), and ~ 100.13 million [M] inter-chromosomal contacts (~ 4.37% occurring between homologs) (Fig. 1b and **Supplementary Fig. 5**). The allelic expression of ~ 11,430 protein coding genes (59.14% of the autosomal genes) was estimated using Allelome.PRO[31] (**Supplementary Fig. 5;** see Methods).

We then reconstructed the 3D genome structures of the 14 diploid samples from their haplotype-resolved Hi-C maps at 20-kb resolution (Fig. 1c). Pig genome contained full sets of functional condensin II subunits (a determinant of architecture type)[32] (**Supplementary Fig. 1b**); its 3D structures had characteristically uneven mass distribution (*i.e.*, nucleotides were concentrated in the nucleus interior, except in the hollow nucleolus) (Fig. 1d) and displayed distinct chromosomal territories (Fig. 1e). Chromosomes preferred the centromere-facing-out (no Rabl-like) configuration (Fig. 1f) and sequence-dependent spatial organization of chromatin (**Supplementary Fig. 1c–h**). These findings agreed with human and rodent studies[8,9,11,33]. Additionally, homolog pairs were nearly equidistant from the center of the 3D nucleus (Fig. 1g and **Supplementary Fig. 1i**) and in closer spatial proximity (Fig. 1h–j), with highly similar intra- and inter-chromosomal contact patterns (indicated by multiple-chromosomal intermingling intensity; **Supplementary Fig. 1f**), compared to non-homologous chromosomes (heterologs).

Haplotype-resolved analyses of Hi-C maps revealed that different tissues displayed more distinct differences in 3D genome architecture than that between the same tissue in either parent-of-origin or between parental breeds (Table 1 and **Supplementary Fig. 1j–l**), which was recapitulated by allelic expression data (**Supplementary Fig. 1m,n**).

Table 1
Summary of allelic chromatin reorganization at multiple scales

Haplotype-resolved characteristics[a]	Measurements	Paired tissues	Parent-of-origin			Parental breed		
			Liver	Muscle	Brain	Liver	Muscle	Brain
Hi-C contact map	Similarity (QuASAR)[b]	~ 0.86	0.98	0.98	0.98	0.97	0.98	0.98
Compartment	A/B switched regions (Mb)[c]	~ 224.51	4.52	6.90	12.50	18.86	12.52	17.20
	A/B variable regions (Mb)[d]	/	4.52	4.88	4.22	61.42	29.30	14.80
TAD	No. of shifted boundaries	~ 184	1	2	2	0	0	0
PEI	No. of genes with differential RPS[e]	~ 4865	397	327	242	532	538	489

[a] See also **Supplementary Data 1–9**.

[b]. Correlations of haplotype-resolved Hi-C maps were determined using QuASAR scores. See **Supplementary Fig. 1j–v** for similarity measurements in GenomeDISCO and HiCRep.

[c]. Regions exhibiting distinct compartment status (*i.e.*, from A [positive value of A-B index] to B [negative value of A-B index], or from B to A) as A/B switched regions.

[d] Regions exhibiting the same compartment status but with the variable compartment scores (|ΔA-B index| > 0.5 and $P < 0.05$, paired Student's *t*-test) as A/B variable regions.

[e]. Significant RPS changes in genes between pairwise tissues (|ΔRPS| > 3) and between parent-of-origins (|ΔRPS| > 0.3), or between breeds (|ΔRPS| > 0.3) were determined by paired Student's *t*-test ($P < 0.05$) and |ΔRPS|.

Allelically compartmental rearrangements

In total, ~ 9.91% of the genome (~ 224.51 Mb regions) was found to have different compartmental status between tissues (*i.e.*, A/B switches; Fig. 2a, Table 1 and **Supplementary Data 1**). Genes located in tissue-restricted A compartment regions showed generally higher expression than the same genes in the B compartment in other tissues (**Supplementary Figs. 6 and 7a**). These tissue-dependent transcriptionally active genes were mainly involved in specialized functions of their respective tissue (**Supplementary Fig. 7b,c**).

Given the rarity of regions that show compartment switching between parent-of-origin (~ 0.35% of the genome; ~7.97 Mb) and between parental breeds (~ 0.71% of the genome; ~16.19 Mb), we compared scores (*i.e.*, A-B index values) between alleles with regions in same compartment and found that these regions had significantly different compartment scores (termed as A/B variables; $P < 0.05$, paired Student's *t*-test, and |ΔA-B index| > 0.5; Fig. 2a). This analysis showed that ~ 0.55% (~ 12.51 Mb) of the genome regions displayed allelic differences in compartmentalization (including A/B switches and A/B variables) between parents-of-origin (Table 1 and **Supplementary** Data **2**), which were enriched with 126 known, imprinted, mammalian genes ($P \leq 0.002$, Chi-square test; **Supplementary Fig. 8a,b** and **9**). However, genomic regions containing known, imprinted genes also had subtle but significantly higher variability in compartment scores between alleles (**Supplementary Fig. 8c**). These results indicated that parent-of-origin-specific compartmentalization could serve as an underlying mechanism for genome imprinting[34,35].

After comparing compartment switching in parents, we examined compartment switching between breeds and found that ~ 2.27% (~ 51.37 Mb) of the genome had allele-specific variable or switched compartments between the parental breeds (Table 1 and **Supplementary Data 3**). These regions had characteristically higher sequence divergence between alleles (*i.e.*, higher density of SNVs and short InDels, and thus lower pairwise similarity between haplotype identity score [IDS], see Methods; **Supplementary Fig. 8d**). Genes located in these regions showed agreement in their allelic bias between compartmentalization and expression (Fig. 2b). Although separating sequence variations from other confounding factors affecting compartmentalization remains challenging, genes located in breed-restricted A compartment regions potentially reflected some of the dramatic phenotypic differences between the parental breeds (**Supplementary Fig. 8e**). Notably, genes located in the Tibetan-restricted A compartment regions (more accessible and relatively active) were enriched for annotations associated with the non-shivering thermogenesis to resist cold stress ('brown fat cell differentiation', three genes), disease resistance ('adaptive immune response', 13 genes; 'immunoglobulin production', four genes), and resistance to intensive solar ultraviolet radiation-induced DNA damage ('regulation of DNA-dependent DNA replication', four genes)[36,37]. These results suggest that the compartmentation in hybrid pigs are separately inherited from their parents which exhibited the different adaption of parental breeds to their local environments.

Allelically variable TAD boundaries

To identify allelic differences in 3D genome architecture that led to imbalanced expression at a finer scale than compartments, we partitioned the haploid genomes into ~ 3956 topologically associating domains (TADs; **Supplementary Fig. 6**). Pairwise comparison between tissues identified ~ 3.93% (~ 184) that had shifted boundaries between tissues (indicated by changes in boundary position with significantly different local boundary scores [LBS[10], see Methods; Fig. 2c, Table 1 and **Supplementary Data 4**). In addition, tissue-restricted boundaries showed substantial differences in compartmentalization, whereas boundaries that were the same between tissues did not (**Supplementary Fig. 10**). Genes near these tissue-restricted boundaries also showed greater differences in expression than those near boundaries that were

the same between tissue, and included critical marker genes associated with core functions of that tissue (**Supplementary Fig. 10**). These results highlight the function of TADs in facilitating local interactions and establishing regulatory context for their respective genes[38].

In total, three TAD boundaries were found to be altered between parent-of-origin across tissues that contained or were near to known imprinted genes (Table 1 and **Supplementary Data 5**). Two of these were paternal-specific boundaries, including one in brain and muscle that affected three imprinted genes (*NDN*, *MAGEL2* and *MKRN3*; Fig. 2d and **Supplementary Fig. 8f**), and the other in brain and liver affecting two imprinted genes (*CASD1* and *SGCE*; **Supplementary Fig. 8g,h**). One maternal-specific boundary was identified in muscle that influenced seven imprinted genes (*ASCL2*, *CD81, H19*, *IGF2*, *INS*, *TH* and *TSSC4*; **Supplementary Fig. 8i**). These imprinted genes around allele-specific boundaries generally showed increased transcriptional activity, supporting the possibility that parent-of-origin-specific TAD organization provides a structural context conducive to activation of imprinted genes[39].

Although no altered boundary positions were detected between haplotypes of the two parental breeds, ~ 190.97 Mb (~ 8.43%) of genomic regions had allelic differences in boundary strength (LBS; $P < 0.05$, paired Student's *t*-test, see Methods; Table 1 and **Supplementary Data 6**) and exhibited significantly higher sequence divergence among alleles than other regions ($P \leq 3.2\times10^{-12}$, Wilcoxon rank-sum test; Fig. 2e).

Haplotype-resolved interrogation of tissue- and imprinting-specific PEI organizations in gene expression control

The information necessary for regulating transcription is conveyed through long-range physical interactions between promoters and enhancers[40,41]. To compile an extensive but reliable genome-wide catalog of PEIs in the diploid pig genome, we conducted additional *in situ* Hi-C assays (dataset 2) for 12 of the 14 aforementioned samples (three tissues from four F1 hybrids, two F1 from the initial and two F1 from reciprocal crosses; Fig. 1a,b and **Supplementary Fig. 1a**). By combining both Hi-C datasets, we built diploid Hi-C maps for each of the 12 samples, including ~ 2.28 billion intra-chromosomal contacts (a maximum resolution of 1-kb), and ~ 239.91 M inter-chromosomal contacts (4.11% between homologs) per map (Fig. 1b and **Supplementary Figs. 2 and 5**). We identified 29,352 enhancers assigned to 7399 promoters (38.28% of the autosomal genes) across 24 haploid genomes at 5-kb resolution (**Supplementary Fig. 11a–f**).

To elucidate how this observed extensive rewiring of PEIs may contribute to the transcriptomic divergence across haploid genomes, we calculated the regulatory potential score (RPS, a spatial proximity-based index of the combined regulatory effects of multiple enhancers for a given gene) for each promoter using a unified set of PEIs from all haplotypes of a given tissue[27,42]; see Methods). As expected, genes with larger RPS had higher expression within haplotypes (**Supplementary Fig. 11g**). Beyond characterizing the spatial proximity of enhancers and promoters, we also measured allelic imbalances in the activity of

enhancers and promoters by generating H3K27ac and H3K4me3 ChIP-seq signal enrichment profiles[43,44] (**Supplementary Fig. 1a, 2g and 11h–o**).

Examination of the haplotype-resolved Hi-C maps revealed extensive rewiring of PEIs between tissues (**Supplementary Fig. 11p–r**). A total of 4865 genes exhibited significantly different RPS between pairs of tissues (Table 1, **Supplementary Data 7** and **Supplementary Fig. 12a**); the majority of which were also differentially transcribed between tissues (**Supplementary Fig. 12b**). These sets of tissue-specific active genes were enriched in annotations related to their respective core functions (**Supplementary Fig. 12c**). Representative examples are shown in **Supplementary Fig. 13**. Notably, examination of individual tissue-specific markers indicated that high RPS was a prevalent feature of germ layer markers and their expression was elevated in tissues originating from their respective germ layers (**Supplementary Fig. 12d–f**), which supported the hypothesis that transcriptional programming (as well as spatial proximity of genes and their enhancers) retains lineage- or differentiation-specific memory[11,45,46]. Representative examples are shown in **Supplementary Fig. 14**.

Genome imprinting represents a phenotypically consequential outcome of allele-specific expression[47]. By comparing data from F1 hybrids produced by reciprocal crosses, we were able to survey genome-wide allelic differences in chromatin conformation for 126 known imprinted genes (downloaded from https://geneimprint.com/site/home; **Supplementary Data 8**), which nearly half of which are located adjacently to each other as previously reported[48]. This analysis identified a set of 101 testable imprinted genes with informative SNVs and enhancer interactions in at least one tissue (Fig. 3a and **Supplementary Data 8**), among which 12 had strongly significant parent-of-origin-specific PEI organization (highly consistent between replicates) and generally corresponding parental expression bias in two or three tissues (Fig. 3b and **Supplementary Fig. 15**). Visual examination of a representative case showed the lower deviations of parent-of-origin-specific 3D structure for the region around the promoter (paternally-expressed *MAGEL2*) across replicates and different tissues (Fig. 3c). These results highlighted the mechanistic contribution of chromatin architecture in imprinting phenomenon[9,39].

Effects of genetic variations and histone modifications on allelic PEI rewiring

Since both genetic and epigenetic regulation shape gene expression patterns, we next explored whether and how inherited sequence variation and histone modifications (the environmentally-induced transgenerational epigenetic inheritance)[49] impacted allelic PEI rewiring in F1 hybrids. Considering the functional impacts of sequence variations on the genome, we comprehensively surveyed the SNVs, short and large InDels (*i.e.*, structural variants, SVs) to determine their effects on allelic PEI formation. For this analysis, we generated long-read sequence data for the four F1 hybrids (~ 125.22 × coverage per individual) using the Oxford Nanopore Technology (ONT) platform and identified their sequence-resolved large InDel alleles (**Supplementary Fig. 16**; see Methods). As expected, higher divergence between the alleles of promoters or enhancers was linked to more dramatic differences in PEI intensity between parental alleles (Fig. 4a,b). We then estimated the probability of long-range interactions for each PEI in

light of their respective, specific variants using the PEP algorithm[50]. We found that PEIs with higher intensity had a higher probability of interaction within alleles (**Supplementary Fig. 17a**) and between alleles (**Supplementary Fig. 17b**), supporting that genetic variations can affect dynamic PEI formation[51,52]. A representative example of the potential for PEI disruption by multiple SNVs and InDels are shown for *SMAD4* in Fig. 4c–e.

Additionally, comparison of allelic differences in the bridging distance between promoter and enhancer supported our general hypothesis that closer linear chromosomal proximity between a promoter and enhancer is associated with closer proximity in 3D nuclear space, and thus with higher PEI intensity, and *vice versa* (Fig. 4f). Simulation of PEI intensities based on the allelic variations in bridging distances using Huynh's algorithm[53] further supported this likelihood (**Supplementary Fig. 17c,d**). Moreover, the significantly greater activity of H3K4me3-enriched promoters or H3K27ac-enriched enhancers generally resulted in allele-specific increases in PEI intensity (**Supplementary Fig. 17e–h**).

Remarkably, we found that allele-specific PEIs was occurred more frequently in genes that contacted multiple enhancers (Fig. 4g), providing validation for the idea that enhancer function in a redundant manner; *i.e.*, the ubiquitous presence of multiple enhancers could serve as a mechanism to reinforce local regulatory circuitry necessary to buffer environmental stresses and genetic disruption, ultimately resulting in phenotypical robustness[54,55]. Importantly, these allele-specific PEIs exert relatively mild effects on RPS between parental alleles (Fig. 4h). In addition, we identified the allele-specific loss of two CCCTC-binding factor [CTCF]-mediated loops in a Tibetan allele in the diploid Hi-C maps of liver. In these cases, the CTCF binding motifs of the loop anchor were effectively disrupted by a 1-bp deletion ('G', chromosome 2: 57822013) and a SNV ('G' > 'T', chromosome 8: 53792054), respectively, but which only resulted in subtle impacts on the expression of genes within these loops (**Supplementary Fig. 18**). These results are analogous with findings in human cancers that 3D genome organization is strikingly disrupted by somatic genomic rearrangements, typically resulting in the fusion of discrete TADs, the gain or loss of CTCF-mediated loops, and enhancer 'hijacking' [53,56], whereas known SV polymorphisms in healthy human populations have significantly less impact on 3D organization[57].

Allelic differences in PEIs associated with phenotypic differences between parental breeds

We next asked whether the allelic rewiring of PEIs in F1 hybrids could contribute to phenotypic divergence between the parental breeds[36,58,59]. In total, ~ 520 genes were identified that showed differential RPS between the Berkshire and Tibetan alleles across tissues (Fig. 5a, Table 1 and **Supplementary Data 9**), which potentially could be an effect of breed-specific genetic variations. Population-level analysis of the 82 purebred diploid pig genomes (Berkshire [n = 21] and Tibetan [n = 61]) further suggested that variants in promoter and enhancers were almost fixed in the more genetically homogeneous Berkshire pigs (Fig. 5a), which could reflect an effect of rigorous selection during breeding[58,59], although functional analyses are required to validate the non-neutrality of these genes. Below, we examine some genes with

highly significant differential RPS between parental breeds that are likely candidates responsible for functional divergence in different tissues between parental breeds.

Liver

Since the liver is the metabolic hub for nutrient homeostasis[60], its function may contribute to major phenotypic differences between Tibetan (long-standing survived in highland at a free-range farming state)[36] and Berkshire pigs (enhanced, high efficiency energy storage under high calorie diets in captivity)[58,59]. As expected, a large suite of genes (254) with higher RPS in the Berkshire allele showed enrichment for 'maintenance of location' (six genes: *CAV1*, *INSIG2*, *LTBP1*, *PFN4, PLIN3* and *SPOUT1*) (Fig. 5a and **Supplementary Fig. 19**), which is necessary for proper subcellular localization of protein complexes or organelles and is therefore essential for lipid transport[61]. Among these genes, *CAV1* is known to function as a protective factor against non-alcoholic fatty liver disease by modulating lipid metabolism under high-fat diet conditions[62]; knockout of this gene in mice is associated with severe hepatic steatosis[63]. Notably, we found that the Berkshire allele had co-enhanced RPS along with the expression of *C12ORF29* (an anti-obesity factor that inhibits fatty acid uptake into hepatocytes)[64] and *CAV2* (a functionally similar *CAV1* paralog)[65] compared with that in Tibetan liver tissue (**Supplementary Fig. 20**). These results suggested a possible protective mechanism for Berkshire pigs against the deleterious effects of a 'diabetogenic' environment, *i.e.*, one that favors inactivity and energy abundance[66].

Conversely, Tibetan pigs are known to exhibit relatively high sensitivity to induction of obesity-related comorbidities (such as development of hepatic metabolic stress) when given excessive dietary energy in pig farms[36,67]. Our analysis identified 278 Tibetan alleles with higher RPS than the Berkshire alleles that could contribute to inflammatory processes, including 'chemokine signaling pathway' (seven genes: *CCL24*, *GNGT2*, *GRK3*, *GSK3B*, *SOS1*, *SOS2* and *STAT5B*) and 'positive regulation of cell-substrate adhesion' (six genes: *COL8A1*, *GSK3B*, *MAP4K4*, *MYADM*, *VEGFA* and *WNT4*; Fig. 5a and **Supplementary Fig. 19**). Of these, *VEGFA* reportedly plays essential roles in promoting angiogenesis and hypoxia-response under inflammation[68], while *MAP4K4* is a mediator of oxidative stress-induced cell death[69]. In addition, three typical inflammatory markers also displayed greater RPS and expression levels in the Tibetan allele compared to that of the Berkshire allele, as well as greater sequence divergence between alleles, including *ABI3* and *ITGA6* (two hypoxia-induced regulators)[70,71], and *CEBPD* (a hepatic lipogenic transcription factor that potentiates macrophage inflammatory response)[72] (**Supplementary Fig. 20**).

Skeletal muscle

Skeletal muscles represent the largest tissue (by weight) in the pig body (~ 63.50% carcass weight for Berkshire and ~ 41.68% for Tibetan)[73] and include the most economically valuable products. In muscle tissue, our analysis uncovered 265 Berkshire alleles with larger RPS than their Tibetan counterpart alleles that are involved in energy metabolism and muscle growth (*e.g.*, six genes related to 'insulin secretion':

ADCY6, *ATF2*, *CREB3*, *GCG*, *KCNU1* and *PLCB2*; Fig. 5a and **Supplementary Fig. 19**). Nonetheless, 273 genes were found with greater RPS in the Tibetan allele that are involved in 'mitochondrial fission' (five genes: *CARMIL1*, *DCN*, *MIEF1*, *PINK1* and *SPIRE1*). Among these genes, *MIEF1* and *CARMIL1* are well-characterized positive regulators of mitochondrial fission, defects in which may cause muscle wasting in adult mammals[74]. These findings reflected a consequence of 200 years of artificial selection in Berkshire pigs for faster growth (daily weight gain during 2–6 months of age, ~ 594.67 g·day^{-1}) and greater meat yield (myofiber cross-sectional area, ~ 5231.84 μm^2); in contrast to Tibetan pigs (~ 269.35 g·day^{-1} and ~ 2574.05 μm^2)[73] (Fig. 5b), which only recently attracted focused breeding attention for pork production at high altitudes.

The plasticity of skeletal muscle in mammals is largely attributable to myofiber heterogeneity (including two main types) which vary in metabolic capacity and contractive functionality[20]. In comparison to glycolytic type II, oxidative type I myofibers are more resistant to fatigue and rich in triglyceride and fatty droplets[75]. In muscle of our F1 hybrids, we observed that genes involved in lipid metabolism, such as 'regulation of fat cell differentiation' (six genes: *AAMDC*, *C1QL4*, *FERMT2*, *GPER1*, *SIRT2* and *ZBTB16*) were more structurally active (*i.e.*, greater PEI intensity) in the Tibetan allele compared to that in the Berkshire allele (Fig. 5a and **Supplementary Fig. 19**). This finding was consistent with a marked increase in the proportion of lipid-rich type I myofibers in muscle of Tibetan pigs (~ 26.26%) compared to that in Berkshire pigs (~ 12.05%; Fig. 5c and **Supplementary Fig. 21**). Consistent, with these data, three specific regulators of type II myofiber, *i.e.*, *ATL2*[76], *DDIT4L* and *KCNC4*[77] had lower RPS and expression in the Tibetan allele (Fig. 5a, d–f). In particular, *DDIT4L* (also known as *REDD2* or *SMHS1*) is essential for highly active glycolytic energy metabolism in specialized type II myofibers[78]. The triglycerides accumulating between or within myofibers represent a substantial energy source (contributing up to 20% of total energy turnover during physical exercise in humans)[79]. The higher proportion of type I myofibers in Tibetan pigs compared to that in Berkshire pigs may be both an effect of selection and husbandry practices as well as a contributing factor in the enhanced athletic performance of grazing Tibetan pigs compared to that of Berkshire pigs raised in limited space for activity in pig farms.

Brain

Given the essential contribution 3D genome organization regulating brain development[80], we also identified genes with differential RPS in brain that are involved in functional and behavioral differences between Berkshire and Tibetan pigs. In total, we identified 253 genes with differentially higher RPS in the Berkshire allele which were notably enriched in 'cellular response to nutrient' (three genes: *FES*, *LPL* and *SFRP1*; Fig. 5a and **Supplementary Fig. 19**), indicating a functional role in the control of appetite and food intake. Further, in the Berkshire allele, *CCK*, which is known to play essential roles in promoting satiety in the brain[81], had particularly high RPS and expression levels. In the Tibetan allele, 236 genes were found with significantly higher RPS that were enriched in 'response to carbohydrate' (seven genes: *ADIPOQ*, *MAP2K3*, *PFKL*, *PNPLA3*, *PTPRN2*, *SMAD4* and *TXN2*; Fig. 5a and **Supplementary Fig. 19**). Among these, *PFKL* reportedly evokes higher rates of glucose utilization in mice brain[82] (**Supplementary Fig. 20**).

These findings together show the distinct effects on central nervous system control of metabolism accomplished through breeding selection (*i.e.*, in the commercial Berkshire pig) versus the effects of natural selection in Tibetan pig (which is more similar to wild boar phenotypically and physiologically) in the absence of artificial selection[36]. To protect Berkshire against the deleterious effects of an extremely high-calorie concentrated diet, the central nervous system triggers a reduction in food intake upon sensing dietary excess[83]. By contrast, Tibetan pigs require high efficiency utilization of low-calorie carbohydrates to adapt to the limited nutritional resources typical of highland grazing[36,67]. This adaptive response potentially facilitates maintenance of brain homeostasis through a 'competent brain-pull' mechanism[84], which is indispensable to maintain high energy levels in the brain despite food deprivation.

Intriguingly, we also found four genes involved in 'peptide transport' (*ABCC4, DISP1, PTPRN2* and *VIP*) that were more active in 3D structure and transcription in the brain of the Tibetan allele compared to that in the Berkshire allele. These genes are responsible for synaptic transmission and plasticity, and learning and memory processes[85]. In particular, *VIP* is a regulatory peptide, the loss of which impairs recall ability and responses to fearful or dangerous conditions in mice[86]. The high RPS and expression of this gene is consistent with empirical observations that wild boars and Tibetan pigs exhibit a pugnacious and wild temperament, suggesting greater mental vigilance and responsiveness conducive to survival in the wild, compared to the relatively sluggish and docile Berkshire pigs, which are thus well-suited for growth in confinement typical of industrial pig production[36]. In addition, *FXN* and *SLC4A7* also had higher RPS and expression in the Tibetan allele (**Supplementary Fig. 20**), both of which have well-established roles in locomotor activity. Mutations in *FXN* have been causatively linked to Friedreich's ataxia in humans, characterized by progressive loss of coordination and balance, as well as ataxic gait in mice[87,88]. *SLC4A7* participates in visual and auditory sensory transmission, and its knockout negatively affects emotional response and exploratory behaviors[89].

These collective observations of differential RPS between Tibetan and Berkshire pigs in genes that are well-established to participate in metabolism, muscle development, and behavior provide strong evidence linking phenotype in pigs with regulatory control of development mediated by 3D chromatin architecture.

Structured homolog pairing in pig genome

Experimental evidence from flies and mammals suggests communication between homologs, which can contribute to transcriptional regulation[90]. Using ultra-deep Hi-C data obtained from 12 diploid samples, each containing ~ 9.86 M inter-homolog contacts (Fig. 1b), we assessed the extent of homolog proximity in pig genomes. The Hi-C maps revealed apparent inter-homolog interactions across the entire length of each autosome at 1-Mb resolution (Fig. 6a), and observed structurally coordinated inter-homolog and intra-chromosome interactions (**Supplementary Fig. 22a–d**). These results were consistent with findings in flies[91,92], which suggested similarity or conservation in the mechanisms of contact formation. Further, enhanced allelic pairing was evident in active A compartment regions, implying a likely functional role for homolog pairing in compartmentalization (**Supplementary Fig. 22e–g**). These results were recapitulated by analyses at 100-kb resolution that showed higher density of contacts (a 5.92-fold increase in inter-

homolog contacts in genomic bins from 20-kb to 100-kb), which suggested greater reliability (**Supplementary Fig. 23**).

We next investigated the infrequent homologous interactions in which an enhancer may act on the homolog of its corresponding promoter in an inter-chromosomal, *trans*-manner[93]. Examination of PEIs between enhancers and promoters from homologs showed that covariation between allelic expression and RPS decreased for genes located in tightly-paired loci with their enhancers (~ 74.35 Mb, or ~ 3.28% of the genome, see Methods; Fig. 6b and **Supplementary Fig. 22h–k**), indicating that RPS only reflects intra-chromosomal regulatory effects of PEI. This result suggests that although rare, tightly organized homolog pairing may result in an additional layer transcriptional regulation (Fig. 6c).

In one representative case, we found unbalanced inter-homolog interactions for the extremely tightly paired *PHOSPHO1* gene (top 0.53–0.59%) across the three tissues examined here (Fig. 6b). *PHOSPHO1* catalyzes hydrolysis of the lipid precursor phosphocholine into choline, which is highly abundant in brown adipose tissue and induced by cold exposure in mice[94,95]. Although the RPS of *PHOSPHO1* in the Tibetan allele was higher than that in the Berkshire allele in all three tissues, *PHOSPHO1* expression levels were lower than that predicted by RPS in the Tibetan allele and were actually down-regulated compared to its expression in the Berkshire allele (0.44-fold in brain, 0.43-fold in muscle, and TPM < 0.5 in liver) (Fig. 6d–f).

Analysis of inter-homolog interactions for the two copies of *PHOSPHO1* in the hybrid progeny revealed that Tibetan allele enhancers were spatially closer to Berkshire allele promoters on the paired homolog than the Berkshire enhancers were to the Tibetan promoter (mean FC of PEI intensities = 4.07, $P < 0.26$, paired Student's *t*-test; Fig. 6d–f). These findings suggested that enhancers located on one homolog could regulate the transcription of a target gene on the other homolog[93,96], which may explain the reduced correlation between intra-chromosomal contact-based RPS and expression of genes located in tightly paired regions. It is worth noting that although the activity of the *PHOSPHO1* promoter and its enhancers are allelically comparable (Fig. 6d–f), the divergence in their sequences and their bridging intervals was relatively larger than that of other genes in all three tissues (Fig. 6g). Similar scenarios were also identified in *B4GALNT2* and *GNGT2* (**Supplementary Figs. 24 and 25**), which can be pursued in future hypothesis-driven investigations. Taken together, these observations clearly demonstrate that non-mitotic homolog pairing occurs in mammalian somatic tissues and tight pairing may participate in modulating gene transcription via inter-homologous PEIs (known as 'transvection' or intragenic complementation[93]).

Discussion

Many long-standing, fundamental biological questions are linked to diploid 3D genome architecture in mammals[97,98]. Here, we generated high-resolution Hi-C diploid maps (at 1-kb maximum resolution) for three porcine tissues that represent distinct embryonic germ layers. These data were then used for in

depth, genome-wide, functional examination of haplotypic changes in 3D chromatin organizations at multiple scales between tissues, parents-of-origin, and parental breeds.

We found that 3D organization was different between tissues, and was highly correlated with the expression of genes involved in the specialized functions of each tissue and in the development of their respective germ layers. This study provides an allele-specific survey of 126 known imprinted genes, and confirms a parent-of-origin-specific 3D conformation for at least 12 of them. The results of this work show that homologs are differentially organized in 3D structures and that this organization is strongly correlated with widespread allelic imbalance in gene expression. This differential 3D organization thus likely contributes to both genetic and epigenetic regulation of phenotypic divergence among pig breeds. These observations expand our current understanding of the critical role of diploid genome architecture in allele-specific gene expression[7]. Homolog pairing was first described in *Drosophila* nearly a century ago[90]. Nonetheless, the detailed structures, the extent of pairing, and pairing-dependent gene regulation in mammals remains largely uncharacterized[91,92]. Using diploid Hi-C maps with unprecedented high resolution, homolog pairing could be obviously detected in porcine somatic tissues and pairing-dependent gene regulation could be observed at individual loci.

Considering that our bulk Hi-C data were obtained from cell population-scale assays, we cannot rule out that cellular heterogeneity may, to some extent, contribute to the observed allelic differences in chromatin features. Thus, single-cell analyses may be used to validate these scenarios in future studies[9,11,99]. In addition, although the allelic choice of X chromosome inactivation takes place individually in cells early in female development, the random choice between alleles is followed by a stable mitotic transmission[1,100]. Further analyses of clonal cells are required to decipher the dynamics of 3D organization between X chromosomes in females, which presents the most striking example of architectural differences between two homologs[101,102].

In recent decades, pigs, and miniature pig breeds in particular, have recently emerged as an attractive biomedical model for humans[19]. Over 730 genetically distinct pig breeds have been recognized worldwide (two thirds of which reside in Europe and China), and thus understanding the control of phenotype between genetically similar breeds is of enormous significance to animal husbandry[29,103]. Indeed, the F1 progeny of these breeding lines are known to exhibit remarkable phenotypic changes[25,103]. These progenies also contain a large number of segregating sites that can be used to distinguish parental haploids; the heterozygous SNP density of F1 hybrids in our pig model is ~ 4.54 per kb, but ~ 1 per kb in humans, ~ 5 per kb in fruit flies, and ~ 7.69 per kb in mice[91,104].

Our haplotype-based interrogation of PEI organization greatly expands the annotation of regulatory DNA elements (enhancers) currently available in the pig reference genome. Using the NHGRI-EBI GWAS catalog and the LiftOver tool (https://genome.ucsc.edu/cgi-bin/hgLiftOver), we found that DNA sequence variations (human noncoding SNPs) associated with specific traits or diseases were enriched in enhancers in the pig genome (mean SNP enrichment scores in enhancers vs. non-enhancer regions: 1.34

vs. 0.97, $P = 0.002$, paired Student's *t*-test, **Supplementary Fig. 26a**), confirming the functional importance of enhancers[105,106]. Supporting our findings of allele-specific PEIs between tissues associated with tissue-specific gene regulation related to core functions, we observed that SNPs associated with cholesterol level are enriched in metabolically active liver, SNPS linked to oxygen supply are enriched in high oxygen-consuming muscle, and Alzheimer's disease associated SNPs are enriched in neuron-rich brain tissue (**Supplementary Fig. 26b**). These results, combined with the preponderance of physiological similarities between pigs and humans, strongly support the use of pigs as a model for human disease[20,26,107].

Recent studies in humans have revealed that most disease- and phenotype-associated variants do not affect protein coding sequences but instead lie in non-coding regulatory regions (~ 98% of the genome)[108,109]. Nonetheless, establishing causal links between phenotypes in humans and the rapidly expanding list of non-coding variants poses an increasingly large challenge[51,52,110]. Our allele-specific survey of the impacts of (epi-)genetic variations on PEI rewiring between parental breeds in the F1 hybrid progeny greatly expands our knowledge of how 3D chromatin organization affects phenotype through transcriptional programming and assigns previously unrecognized function to large parts of the non-coding genome.

Methods

Pig breed crosses

We performed reciprocal crosses (parent-child trios, $n = 6$) of two pig breeds and generated trios of Tibetan ( ) × Berkshire ( ) ($n = 3$) and Berkshire ( ) × Tibetan ( ) ($n = 3$). Ear tissues from the two parental pigs of the six trios were collected ($n = 12$). A total of 14 samples comprising liver ($n = 6$), skeletal muscle ($n = 4$) and brain ($n = 4$) of new-born F1 females were collected and snap-frozen in liquid nitrogen for subsequent assays (see Fig. 1a and **Extended Data** Fig. 1a for details). All research involving living pigs were conducted according to the Regulations for the Administration of Affairs Concerning Experimental Animals (Ministry of Science and Technology, China, revised in March 2017), and approved by the Animal Ethical and Welfare Committee of Sichuan Agricultural University under permit No. 20220194. The animals were humanely euthanized as necessary to prevent suffering.

Whole-genome sequencing

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.

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[111]. 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.

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.

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[112]. 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. 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. We also carried out ChIP-seq using antibodies against CTCF (the most characterized insulator-binding protein) for liver samples from the four F1 hybrids.

Long-read DNA sequencing

We performed long-read sequencing (Oxford Nanopore Technologies) on four F1 hybrids. The genomic DNA from the liver was isolated using the phenol-chloroform extraction method. The DNA library was prepared by DNA repair and end-prep, followed by sequencing adapter ligation and clean-up. The PromethION platform was used for sequencing according to the effective concentration of the DNA library and data output requirements.

4C-seq library preparation and sequencing

Circular chromosome conformation capture sequencing (4C-seq) was performed for livers from the purebred Berkshire and Tibetan pigs (newborn females) as described previously[113,114] with minor

modifications. Briefly, frozen livers were crushed into powder in liquid nitrogen using pre-chilled mortar and pestle. The tissue powder was suspended in PBS and then filtered through 40 μm strainers to obtain single-cell suspensions. The cells were fixed with fresh 2% formaldehyde for 10 min at room temperature. The fixation was quenched with cold glycine at a final concentration of 125 mM, and the cells lysed and incubated in 1 mL cold lysis buffer (50 mM Tris, 150 mM NaCl, 5 mM EDTA, 0.5% NP-40, 1% Triton X-100, 1× protease inhibitors) for 10 min on ice. The nuclei were pelleted by centrifugation and washed twice with PBS. Primary digestion of Nuclear DNA was performed overnight at 37˚C with DpnII or NlaIII (New England Biolabs). DNA fragments were ligated with T4 DNA ligase overnight at 16˚C. Reverse crosslinking was carried out at 65˚C for 12 h with proteinase K, followed by RNaseA digestion and phenol/chloroform extraction precipitation. DNA was further digested using Csp6I (New England Biolabs) overnight, followed by proximity ligation and purification to prepare for the 4C library generation. The 4C libraries were obtained by performing a two-step PCR strategy. For each viewpoint, PCR was performed using Phusion DNA polymerase (Thermo Scientific) with 3.2 μg of the 4C template, and 16 individual PCR reactions performed on 200 ng of 4C template each. PCR products were pooled and purified using the QIAquick PCR Purification Kit (Qiagen), and run on a 4–20% gradient TBE Gel (Thermo Fisher Scientific). DNA smears from 200-bp to 800-bp were extracted and subjected to VAHTS DNA Clean Beads (Vazyme, China) using a bead-to-DNA ratio of 0.8 before high-throughput sequencing on the Illumina NovaSeq 6000 system was performed.

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.).

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[20]. 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.

Whole-genome sequencing data processing

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) [115]. 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)[116] 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)[117], and genetic structure inference using Structure (v 2.3.4)[118].

Hi-C data processing

The heterozygous SNVs of each F1 hybrid were masked by converting them to 'N' bases to reduce reference biases and mapping artifacts, in order to accurately distinguish allele-specific contacts[8]. High-quality Hi-C reads were aligned to the variant masked pig reference genome (Sscrofa 11.1), and sequences representing experimental Hi-C artefacts and other uninformative di-tags were removed, using the standard pipeline provided by HiC-Pro (v 2.9.0)[119]. Consequently, we obtained ~ 7.72 billion contacts for each of the 14 samples (dataset 1), and additional ~ 9.27 billion contacts for each of 12 samples (dataset 2; **Extended Data** Fig. 1a).

Construction Of Chromosomal-span Haplotypes

We generated two parental haplotypes for each autosome by combining the information of heterozygous SNVs obtained from whole-genome sequencing, and the Hi-C contacts. For each trio, we first phased the heterozygous SNVs of F1 hybrids when at least one parental genotype was homozygous (**Supplementary Fig. 2e**). We also phased heterozygous variants of unknown parental origin based on Hi-C contacts using HapCUT2 (v 1.3.1)[120] (**Supplementary Fig. 2f**). We constructed chromosomal-span haplotypes for each F1 hybrid by combining the results of the two strategies, incorporating the vast majority of all heterozygous SNVs. This generated two haplotypes for each chromosome, one for the maternal and another for the paternal alleles.

Reconstruction Of Haplotype-resolved Hi-c Maps

We reconstructed haplotype-resolved Hi-C maps for 14 diploid F1 pig samples using SNPsplit (v 0.3.4)[121], HaploHiC (v 0.32)[4] and the Juicer Tools[122] (**Supplementary Fig. 3**). The reconstruction analysis was separately applied to dataset 1 and the aggregation of datasets 1 and 2 (**Extended Data** Fig. 1a). For a diploid sample, we employed the SNPsplit (v 0.3.4) 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%; **Supplementary Fig. 3a**).

According to their parental origins, the informative contacts can be classified into six categories (**Supplementary Fig. 3b).** 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 97.18–98.35%. 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)[4] 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. Consequently, 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).

The Juicer Tools[121] software was used to generate Knight-Ruiz (KR) matrix balancing normalized[123] contact matrices for haplotype-resolved intra- (at 20-kb and 100-kb resolution) and inter-chromosomal (at 1-Mb resolution) Hi-C maps. Notably, normalized intra- and inter-chromosomal contact matrices (36 × 36) were generated for 18 pairs of homologous autosomes for each sample. After this, the intra- and inter-chromosomal matrices were further normalized across haplotypes using the quantile method[124] across samples using the counts per million (CPM, normalized to the average abundance across all samples) strategy.

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[31]. The Kallisto (v 0.44.0) software[125] 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)[126] 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[31] 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.

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[127].

Analysis Of Haplotype-resolved Hi-c Maps

Correlations between intra-chromosomal matrices

We used HiCRep[128], GenomeDISCO[129], and QuASR-Rep[130] with default parameters to assess the reproducibility of normalized intra-chromosomal contact matrices (at 100-kb resolution) across all haplotypes (n = 28).

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[131] 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 (**Extended Data** Fig. 1a).

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[132]. 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[132] 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 (**Extended Data** Fig. 1a).

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. 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. 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 (**Extended Data** Fig. 1a).

Topologically associated domains (TAD) calling

Haplotype-resolved TADs were identified at 20-kb resolution using the Directionality Index (DI)[132,133] and the Insulation Index (IS)[101] 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[133]. 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[101]. 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 (**Extended Data** Fig. 1a).

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)[134] and VI (measurement of the similarity of all subsets of the two TAD structures using a dynamic programming algorithm to compute the VI metrics)[135] were calculated to access the reproducibility of TADs between haplotypes.

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[10]; 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[136] 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 (**Extended Data** Fig. 1a).

Identification of promoter-enhancer interactions

To reliably identify haplotype-resolved promoter-enhancer interactions (PEIs), we performed additional *in situ* Hi-C assays (dataset 2) for 12 of the 14 aforementioned samples (three tissues from four F1 pigs, Fig. 1b and **Extended Data** Fig. 1a). By combining both Hi-C datasets 1 and 2, we obtained very high-resolution diploid Hi-C maps (a maximum resolution of 1-kb; **Supplementary Fig. 4h,i**). 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[137] with default parameters. 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.

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[138]. 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 (**Extended Data** Fig. 1a).

Quantification of promoter and enhancer activity

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[139] (*Q*-value < 0.05). Highly and moderately active enhancers were identified using the standard ROSE algorithm[140,141]. 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; **Extended Data** Fig. 3h); 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)[142] was used to visualize the ChIP-seq signals, gene expression data and gene locations.

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[31] 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[31] 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.

Effects of genetic variants on PEIs

Large-scale InDel calling using long-read DNA sequencing

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[143] 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)[144] and then aligned to the pig reference genome using minimap2 (v 2.17)[145]. 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[144]. The assembled contigs were then mapped to the reference genome using NGMLR (v 0.2.7)[146] and the high-confidence large-scale InDels were identified using Sniffles (v 1.0.11)[146] with at least two supporting contigs.

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

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[147]. 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)[148].

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[50]. 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[149]. 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 'pseudogenome'). We then created a map from each 5-kb enhancer or 2.7-kb promoter region of the reference genome to the haplotype-specific pseudogenome. The haplotype-resolved interacting and non-interacting promoter-enhancer pairs with both ends aligned from the reference genome to the respective haplotype-specific pseudogenome were retained for further analyses. We applied FIMO[150] for scanning motifs along the pseudogenome sequences of promoters and enhancers with the PWMs from HOCOMOCO Human v10 database[151]. 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[152].

We used XGBoost[153], a scalable and highly accurate implementation of gradient boosting[154], 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.

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[155]. 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

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

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:

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

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.

Calculation of inter-homolog pairing scores

To measure the strength of interactions between homologs, we generated inter-chromosomal normalized Hi-C maps at 20-kb resolution and calculated inter-homolog pairing scores (HPS) for each 20-kb bin. The PS of a 20-kb bin is the $\log_2$ transformed average contact frequency between homologs within a window of *W* bins up- and downstream of the specific 20-kb bin[91]. PS is defined as:

$$PS^{w}(i) = log_2\left(ContactFrequency_{m,n}\right)$$

where *m* and *n* correspond to the *i* − *W*-th and *i* + *W*-th bins. The intra-chromosomal score (ICS) was calculated using the intra-chromosomal Hi-C maps. This analysis was applied to aggregation of the Hi-C datasets 1 and 2 (**Extended Data** Fig. 1a).

ST data processing and analysis

The spatial transcriptomics sequencing data of skeletal muscle were analyzed as previously described[20]. 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)[156] package for removing low-quality spots or non-expressed genes, normalization (regularized negative binomial regression, 'SCTransform' function) of gene expression data[157], 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[158] using scater (v 1.24.0)[159] and limma (v 3.52.0)[160].

Functional enrichment analysis

Functional enrichment analysis was performed using the Metascape tool[161] 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.

Phylogenetic relationship of mammals estimated using condensin II subunits

The topological relationships of each of the five condensin II subunits (SMC2, SMC4, CAP-H2, CAP-G2, and CAP-D3)[162] across several mammals were constructed based on protein sequences using the multiple sequence alignment program Clustal Omega tool (v.1.2.1). This software uses seeded guide trees and HMM profile-profile techniques to generate the alignments[163]. The protein sequences used to generate the trees were downloaded from Uniprot (https://www.uniprot.org/)[164].

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)[165]. 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[105]. The significance of the enrichment score was calculated using a χ^2 test for each trait or disease that contained more than 50 SNPs.

Information related to statistical analysis

Statistical details were described in Fig. legends and the Results section. No methods were used to determine whether the data met the specific assumptions of each statistical approach.

Declarations

Code availability

All the code used for data analysis is available at https://github.com/QianZiTang/linyu.

Competing interests

The authors declare no competing interests.

Acknowledgments

We thank Dr. Isaac V. Greenhut for valuable discussion and feedback on the manuscript. We thank the High-Performance Computing Platform of Sichuan Agricultural University and the Ya'an Big Data Industrial Park for providing computing resources and support that have contributed to this research.

This work was supported by the National Key R & D Program of China (2020YFA0509500 to M.L., 2021YFA0805903 to L.J. and 2021YFD1300800 to L.L.), the Sichuan Science and Technology Program (2021ZDZX0008 and 2022JDJQ0054 to L.J. and Q.T., 2021YFYZ0009 to M.L., 2021YFYZ0030 to Y.G., 2021YFH0033 to J.M. and 2021YFS0008 to L.L.), the National Natural Science Foundation of China (U19A2036 to M.L., 32102507 to J.L., 32102512 to K.L. and 31872335 to X.W.L.), the China Agriculture Research System (CARS-35-01A to X.W.L.), the Ya'an Science and Technology Program (21SXHZ0022 to L.J.), the Post-doctoral Program of China (2020M683648XB to J.L. and 2021M692329 to L.L.), and the Foundation of Key Laboratory of Pig Industry Sciences (22519C to F.K.).

Author contributions

Conceptualization, M.L. and Y.G.; methodology, Y.L. and Q.T.; validation, Y.L., J.L., L.J., Y.Z., X.K.L.; formal analysis, Y.L., J.L., L.J., J.B., and J.Z.; investigation, Y.W., P.L., K.L., M.H., D.L., C.L., Z.H., B.Z., L.L., and F.K.; resources, Y.G., Y.S., Y.F., X.W., T.W., A.J., J.M., J.W., and L.G.; data curation, L.S., L.Z., Y.J., and G.T.; writing, M.L., J.L, L.J., and Q.T.; visualization, Q.L., H.L., and X.W.L.; supervision, M.L., Q.T., and J.L.; project administration, M.L.; funding acquisition, M.L.

References

1. Reinius, B. & Sandberg, R. Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. Nat. Rev. Genet. **16**, 653–664 (2015).

2. Xie, W. et al. Base-resolution analyses of sequence and parent-of-origin dependent DNA methylation in the mouse genome. Cell **148**, 816–831 (2012).
3. Rivera-Mulia, J. C. et al. Allele-specific control of replication timing and genome organization during development. Genome Res. **28**, 800–811 (2018).
4. Lindsly, S. et al. Functional organization of the maternal and paternal human 4D Nucleome. iScience **24**, 103452 (2021).
5. Oudelaar, A. M. & Higgs, D. R. The relationship between genome structure and function. Nat. Rev. Genet. **22**, 154–168 (2021).
6. Akgol Oksuz, B. et al. Systematic evaluation of chromosome conformation capture assays. Nat. Methods **18**, 1046–1055 (2021).
7. Li, J., Lin, Y., Tang, Q. & Li, M. Understanding three-dimensional chromatin organization in diploid genomes. Comput. Struct. Biotechnol. J. **19**, 3589–3598 (2021).
8. Du, Z. et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. Nature **547**, 232–235 (2017).
9. Tan, L. et al. Changes in genome architecture and transcriptional dynamics progress independently of sensory experience during post-natal brain development. Cell **184**, 741–758.e717 (2021).
10. Han, Z. et al. Diploid genome architecture revealed by multi-omic data of hybrid mice. Genome Res. **30**, 1097–1106 (2020).
11. Tan, L., Xing, D., Chang, C. H., Li, H. & Xie, X. S. Three-dimensional genome structures of single diploid human cells. Science **361**, 924–928 (2018).
12. FAO. *The Second Report on the State of the World's Animal Genetic Resources for Food and Agriculture*. (FAO Commission on Genetic Resources for Food and Agriculture Assessments, Rome; 2015).
13. Zhu, Q. et al. Specification and epigenomic resetting of the pig germline exhibit conservation with the human lineage. Cell Rep. **34**, 108735 (2021).
14. Liu, T. et al. Cross-species single-cell transcriptomic analysis reveals pre-gastrulation developmental differences among pigs, monkeys, and humans. Cell Discov. **7**, 1–17 (2021).
15. Yan, S. et al. A Huntingtin Knockin Pig Model Recapitulates Features of Selective Neurodegeneration in Huntington's Disease. Cell (2018).
16. Moretti, A. et al. Somatic gene editing ameliorates skeletal and cardiac muscle failure in pig and human models of Duchenne muscular dystrophy. Nat. Med. **26**, 207–214 (2020).
17. Decaro, N. & Lorusso, A. Novel human coronavirus (SARS-CoV-2): A lesson from animal coronaviruses. Vet. Microbiol. **244**, 108693 (2020).
18. Schelstraete, W., Devreese, M. & Croubels, S. Comparative toxicokinetics of Fusarium mycotoxins in pigs and humans. Food Chem. Toxicol. **137**, 111140 (2020).
19. Lunney, J. K. et al. Importance of the pig as a human biomedical model. Sci. Transl. Med. **13**, eabd5758 (2021).

20. Jin, L. et al. A pig BodyMap transcriptome reveals diverse tissue physiologies and evolutionary dynamics of transcription. Nat. Commun. **12**, 3715 (2021).
21. Karlsson, M. et al. Genome-wide annotation of protein-coding genes in pig. BMC Biol. **20**, 25 (2022).
22. Sjöstedt, E. et al. An atlas of the protein-coding genes in the human, pig, and mouse brain. Science **367**, eaay5947 (2020).
23. Dawson, H. D. et al. Structural and functional annotation of the porcine immunome. BMC Genom. **14**, 332 (2013).
24. Warr, A. et al. An improved pig reference genome sequence to enable pig genetics and genomics research. Gigascience **9**, giaa051 (2020).
25. Groenen, M.A. et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature **491**, 393–398 (2012).
26. Zhao, Y. et al. A compendium and comparative epigenomics analysis of *cis*-regulatory elements in the pig genome. Nat. Commun. **12**, 1–17 (2021).
27. Zhi, M. et al. Generation and characterization of stable pig pregastrulation epiblast stem cell lines. Cell Res, **32**, 383–400 (2021).
28. Yue, Y. et al. Extensive germline genome engineering in pigs. Nat. Biomed. Eng. **5**, 134–143 (2021).
29. Li, M. et al. Comprehensive variation discovery and recovery of missing sequence in the pig genome using multiple *de novo* assemblies. Genome Res. **27**, 865–874 (2017).
30. Deng, Q., Ramsköld, D., Reinius, B. & Sandberg, R. Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. Science **343**, 193–196 (2014).
31. Andergassen, D. et al. Allelome.PRO, a pipeline to define allele-specific genomic features from high-throughput sequencing data. Nucleic Acids Res. **43**, e146 (2015).
32. Hoencamp, C. et al. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. Science **372**, 984–989 (2021).
33. Stevens, T. J. et al. 3D structures of individual mammalian genomes studied by single-cell Hi-C. Nature **544**, 59–64 (2017).
34. Dixon, J. R. et al. Chromatin architecture reorganization during stem cell differentiation. Nature **518**, 331–336 (2015).
35. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell **159**, 1665–1680 (2014).
36. Li, M. et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nat. Genet. **45**, 1431–1438 (2013).
37. Witt, K. E. & Huerta-Sánchez, E. Convergent evolution in human and domesticate adaptation to high-altitude environments. Philos. Trans. Roy. Soc. B **374**, 20180235 (2019).
38. Beagan, J. A. & Phillips-Cremins, J. E. On the existence and functionality of topologically associating domains. Nat. Genet. **52**, 8–16 (2020).

39. Monk, D., Mackay, D. J., Eggermann, T., Maher, E. R. & Riccio, A. Genomic imprinting disorders: lessons on how genome, epigenome and environment interact. Nat. Rev. Genet. **20**, 235–248 (2019).
40. Song, W., Sharan, R. & Ovcharenko, I. The first enhancer in an enhancer chain safeguards subsequent enhancer-promoter contacts from a distance. Genome Biol. **20**, 197 (2019).
41. Oh, S. et al. Enhancer release and retargeting activates disease-susceptibility genes. Nature **595**, 735–740 (2021).
42. Li, D. et al. Dynamic transcriptome and chromatin architecture in granulosa cells during chicken folliculogenesis. Nat. Commun. **13**, 1–14 (2022).
43. Villar, D. et al. Enhancer evolution across 20 mammalian species. Cell **160**, 554–566 (2015).
44. Pérez-Rico, Y. A. et al. Comparative analyses of super-enhancers reveal conserved elements in vertebrate genomes. Genome Res. **27**, 259–268 (2017).
45. Hon, G. C. et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat. Genet. **45**, 1198–1206 (2013).
46. Javierre, B. M. et al. Lineage-Specific genome architecture links enhancers and non-coding disease variants to target gene promoters. *Cell* **167**, 1369–1384 e1319 (2016).
47. Tucci, V., Isles, A. R., Kelsey, G. & Ferguson-Smith, A. C. Genomic imprinting and physiological processes in mammals. Cell **176**, 952–965 (2019).
48. Santini, L. et al. Genomic imprinting in mouse blastocysts is predominantly associated with H3K27me3. Nat. Commun. **12**, 1–16 (2021).
49. Fitz-James, M. H. & Cavalli, G. Molecular mechanisms of transgenerational epigenetic inheritance. Nat. Rev. Genet. (2022).
50. Yang, Y., Zhang, R., Singh, S. & Ma, J. Exploiting sequence-based features for predicting enhancer-promoter interactions. Bioinformatics **33**, i252–i260 (2017).
51. Kvon, E. Z. et al. Comprehensive *in vivo* interrogation reveals phenotypic impact of human enhancer variants. Cell **180**, 1262–1271. e1215 (2020).
52. Spielmann, M., Lupiáñez, D. G. & Mundlos, S. Structural variation in the 3D genome. Nat. Rev. Genet. **19**, 453–467 (2018).
53. Huynh, L. & Hormozdiari, F. TAD fusion score: discovery and ranking the contribution of deletions to genome structure. Genome Biol. **20**, 1–13 (2019).
54. Kvon, E. Z., Waymack, R., Gad, M. & Wunderlich, Z. Enhancer redundancy in development and disease. Nat. Rev. Genet. **22**, 324–336 (2021).
55. Laverré, A., Tannier, E. & Necsulea, A. Long-range promoter-enhancer contacts are conserved during evolution and contribute to gene expression robustness. Genome Res. **32**, 280–296 (2022).
56. Weischenfeldt, J. et al. Pan-cancer analysis of somatic copy-number alterations implicates IRS4 and IGF2 in enhancer hijacking. Nat. Genet. **49**, 65–74 (2017).
57. Akdemir, K. C. et al. Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. Nat. Genet. **52**, 294–305 (2020).

58. Li, M. et al. Whole-genome sequencing of Berkshire (European native pig) provides insights into its origin and domestication. Sci. Rep. **4**, 4678 (2014).
59. Jeong, H. et al. Exploring evidence of positive selection reveals genetic basis of meat quality traits in Berkshire pigs through whole genome sequencing. BMC Genet. **16**, 104 (2015).
60. Ben-Moshe, S. & Itzkovitz, S. Spatial heterogeneity in the mammalian liver. Nat. Rev. Gastroenterol. Hepatol. **16**, 395–410 (2019).
61. Wong, L. H., Gatta, A. T. & Levine, T. P. Lipid transfer proteins: the lipid commute via shuttles, bridges and tubes. Nat. Rev. Mol. Cell Biol. **20**, 85–101 (2019).
62. Han, M. et al. Hepatocyte caveolin-1 modulates metabolic gene profiles and functions in non-alcoholic fatty liver disease. Cell Death Dis. **11**, 1–14 (2020).
63. Li, M. et al. Caveolin1 protects against diet induced hepatic lipid accumulation in mice. PLoS ONE **12**, e0178748 (2017).
64. Wang, L., Tran, A., Lee, J. & Belsham, D. D. Palmitate differentially regulates Spexin, and its receptors Galr2 and Galr3, in GnRH neurons through mechanisms involving PKC, MAPKs, and TLR4. Mol. Cell. Endocrinol. **518**, 110991 (2020).
65. Yang, X. et al. Twist1-induced miR-199a-3p promotes liver fibrosis by suppressing caveolin-2 and activating TGF-beta pathway. Signal Transduct. Target. Ther. **5**, 75 (2020).
66. Gerstein, H. C. & Waltman, L. Why don't pigs get diabetes? Explanations for variations in diabetes susceptibility in human populations living in a diabetogenic environment. Can. Med. Assoc. J. **174**, 25–26 (2006).
67. Zhu, L. et al. Distinct expression patterns of genes associated with muscle growth and adipose deposition in tibetan pigs: a possible adaptive mechanism for high altitude conditions. High Alt. Med. Biol. **10**, 45–55 (2009).
68. Apte, R. S., Chen, D. S. & Ferrara, N. VEGF in signaling and disease: beyond discovery and development. Cell **176**, 1248–1264 (2019).
69. Roth Flach, R.J. et al. Map4k4 impairs energy metabolism in endothelial cells and promotes insulin resistance in obesity. Am. J. Physiol. Endocrinol. Metab. **313**, E303–E313 (2017).
70. Richter, M. et al. Single-nucleus RNA-seq2 reveals functional crosstalk between liver zonation and ploidy. Nat. Commun. **12**, 1–16 (2021).
71. Verma, P. et al. Transcriptome analysis of circulating PBMCs to understand mechanism of high altitude adaptation in native cattle of Ladakh region. Sci. Rep. **8**, 1–15 (2018).
72. Spek, C. A., Aberson, H. L., Butler, J. M., de Vos, A. F. & Duitman, J. CEBPD potentiates the macrophage inflammatory response but CEBPD knock-out macrophages fail to identify cebpd-dependent pro-inflammatory transcriptional programs. Cells **10**, 2233 (2021).
73. Wang, L. Y. et al. Animal genetic resources in China: pigs (ed. China National Commission of Animal Genetic Resources). (China Agricultural Press, 2011).

74. Kraus, F., Roy, K., Pucadyil, T. J. & Ryan, M. T. Function and regulation of the divisome for mitochondrial fission. Nature **590**, 57–66 (2021).
75. Malenfant, P. et al. Fat content in individual muscle fibers of lean and obese subjects. Int. J. Obes. Relat. Metab. Disord. **25**, 1316–1321 (2001).
76. Murgia, M. et al. Protein profile of fiber types in human skeletal muscle: a single-fiber proteomics study. Skelet. Muscle **11**, 1–19 (2021).
77. Miranda, D.R. et al. Mechanisms of altered skeletal muscle action potentials in the R6/2 mouse model of Huntington's disease. Am. J. Physiol. Cell Physiol. **319**, C218–C232 (2020).
78. Pisani, D. F., Leclerc, L., Jarretou, G., Marini, J. F. & Dechesne, C. A. SMHS1 is involved in oxidative/glycolytic-energy metabolism balance of muscle fibers. Biochem. Biophys. Res. Commun. **326**, 788–793 (2005).
79. Roepstorff, C., Vistisen, B. & Kiens, B. Intramuscular triacylglycerol in energy metabolism during exercise in humans. Exerc. Sport Sci. Rev. **33**, 182–188 (2005).
80. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. Nature **538**, 523–527 (2016).
81. Cawthon, C. R. & Claire, B. The critical role of CCK in the regulation of food intake and diet-induced obesity. Peptides **138**, 170492 (2021).
82. Dienel, G.A. Brain glucose metabolism: integration of energetics with function. Physiol. Rev. **99**, 949–1045 (2019).
83. Mancini, G. & Horvath, T. L. Viral vectors for studying brain mechanisms that control energy homeostasis. Cell Metab. **27**, 1168–1175 (2018).
84. Herhaus, B., Ullmann, E., Chrousos, G. & Petrowski, K. High/low cortisol reactivity and food intake in people with obesity and healthy weight. Transl. Psychiatry **10**, 1–8 (2020).
85. Goff, K. M. & Goldberg, E. M. A role for vasoactive intestinal peptide interneurons in neurodevelopmental disorders. Dev. Neurosci. **43**, 168–180 (2021).
86. Stack, C. M. et al. Deficits in social behavior and reversal learning are more prevalent in male offspring of VIP deficient female mice. Exp. Neurol. **211**, 67–84 (2008).
87. Savelieff, M. G. & Feldman, E. L. Lessons for clinical trial design in Friedreich's ataxia. Lancet Neurol. **20**, 330–332 (2021).
88. Chandran, V. et al. Inducible and reversible phenotypes in a novel mouse model of Friedreich's ataxia. Elife **6**, e30054 (2017).
89. Choi, I. et al. Sodium bicarbonate cotransporter NBCn1/Slc4a7 affects locomotor activity and hearing in mice. Behav. Brain Res. **401**, 113065 (2021).
90. Child, M. B. 6th. et al. Live imaging and biophysical modeling support a button-based mechanism of somatic homolog pairing in *Drosophila*. Elife **10** (2021).
91. AlHaj Abed, J. et al. Highly structured homolog pairing reflects functional organization of the *Drosophila* genome. Nat. Commun. **10**, 4485 (2019).

92. Erceg, J. et al. The genome-wide multi-layered architecture of chromosome pairing in early *Drosophila* embryos. Nat. Commun. **10**, 4486 (2019).
93. Galouzis, C. C. & Prud'homme, B. Transvection regulates the sex-biased expression of a fly X-linked gene. Science **371**, 396–400 (2021).
94. Jiang, M., Chavarria, T. E., Yuan, B., Lodish, H. F. & Huang, N. J. Phosphocholine accumulation and PHOSPHO1 depletion promote adipose tissue thermogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 15055–15065 (2020).
95. Gliniak, C. M. & Scherer, P. E. PHOSPHO1 puts the breaks on thermogenesis in brown adipocytes. *Proc. Natl. Acad. Sci.* **117**, 16726–16728 (2020).
96. Lim, B., Heist, T., Levine, M. & Fukaya, T. Visualization of transvection in living *Drosophila* embryos. *Mol. Cell* **70**, 287–296. e286 (2018).
97. Zhang, D., Lam, J. & Blobel, G.A. Engineering three-dimensional genome folding. Nat. Genet. **53**, 602–611 (2021).
98. Jerkovic, I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. Nat. Rev. Mol. Cell Biol. **22**, 511–528 (2021).
99. Zhou, T., Zhang, R. & Ma, J. The 3D genome structure of single cells. Annu. Rev. Biomed. Data Sci. **4**, 21–41 (2021).
100. Galupa, R. et al. A conserved noncoding locus regulates random monoallelic *Xist* expression across a topological boundary. Mol. Cell **77**, 352–367. e358 (2020).
101. Crane, E. et al. Condensin-driven remodelling of X chromosome topology during dosage compensation. Nature **523**, 240–244 (2015).
102. Bauer, M. et al. Chromosome compartments on the inactive X guide TAD formation independently of transcription during X-reactivation. Nat. Commun. **12**, 1–21 (2021).
103. Chen, K., Baxter, T., Muir, W. M., Groenen, M. A. & Schook, L. B. Genetic resources, genome mapping and evolutionary genomics of the pig (*Sus scrofa*). Int. J. Biol. Sci. **3**, 153–165 (2007).
104. de Wit, E. Capturing heterogeneity: single-cell structures of the 3D genome. Nat. Struct. Mol. Biol. **24**, 437–438 (2017).
105. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. Cell **155**, 934–947 (2013).
106. Maurano, M. T. et al. Systematic localization of common disease-associated variation in regulatory DNA. Science **337**, 1190–1195 (2012).
107. Pan, Z. et al. Pig genome functional annotation enhances the biological interpretation of complex traits and human disease. Nat. Commun. **12**, 1–15 (2021).
108. Nurk, S. et al. The complete sequence of a human genome. Science **376**, 44–53 (2022).
109. Aganezov, S. et al. A complete reference genome improves analysis of human genetic variation. Science **376**, eabl3533 (2022).
110. Nasser, J. et al. Genome-wide enhancer maps link risk variants to disease genes. Nature **593**, 238–243 (2021).

111. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell **159**, 1665–1680 (2014).
112. Han, K., Ren, R., Cao, J., Zhao, S. & Yu, M. Genome-wide identification of histone modifications involved in placental development in pigs. Front. Genet. **10**, 277 (2019).
113. Krijger, P. H. L., Geeven, G., Bianchi, V., Hilvering, C. R. E. & de Laat, W. 4C-seq from beginning to end: A detailed protocol for sample preparation and data analysis. Methods **170**, 17–32 (2020).
114. van de Werken, H. J. et al. 4C technology: protocols and data analysis. Meth. Enzymol. **513**, 89–112 (2012).
115. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25**, 1754–1760 (2009).
116. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. **20**, 1297–1303 (2010).
117. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. **88**, 76–82 (2011).
118. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. Genetics **155**, 945–959 (2000).
119. Servant, N. et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. **16**, 259 (2015).
120. Edge, P., Bafna, V. & Bansal, V. HapCUT2: robust and accurate haplotype assembly for diverse sequencing technologies. Genome Res. **27**, 801–812 (2017).
121. Krueger, F. & Andrews, S. R. SNPsplit: allele-specific splitting of alignments between genomes with known SNP genotypes. F1000Res. **5**, 1479 (2016).
122. Durand, N. C. et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. **3**, 95–98 (2016).
123. Knight, P. A. & Ruiz, D. A fast algorithm for matrix balancing. IMA J. Numer. Anal. **33**, 1029–1047 (2013).
124. Fletez-Brant, K., Qiu, Y., Gorkin, D. U., Hu, M. & Hansen, K. D. Removing unwanted variation between samples in Hi-C experiments. *BioRxiv*, 214361 (2021).
125. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. **34**, 525–527 (2016).
126. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics **29**, 15–21 (2013).
127. Liu, S. et al. Genome Architecture Mediates Transcriptional Control of Human Myogenic Reprogramming. iScience **6**, 232–246 (2018).
128. Yang, T. et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. Genome Res. **27**, 1939–1949 (2017).
129. Ursu, O. et al. GenomeDISCO: a concordance score for chromosome conformation capture experiments using random walks on contact map graphs. Bioinformatics **34**, 2701–2707 (2018).

130. Yardimci, G. G. et al. Measuring the reproducibility and quality of Hi-C data. Genome Biol. **20**, 57 (2019).
131. Rieber, L. & Mahony, S. miniMDS: 3D structural inference from high-resolution Hi-C data. Bioinformatics **33**, i261–i266 (2017).
132. Rowley, M.J. et al. Evolutionarily conserved principles predict 3D chromatin organization. Mol. Cell **67**, 837–852.e837 (2017).
133. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature **485**, 376–380 (2012).
134. Zufferey, M., Tavernari, D., Oricchio, E. & Ciriello, G. Comparison of computational methods for the identification of topologically associating domains. Genome Biol. **19**, 217 (2018).
135. Sauerwald, N. & Kingsford, C. Quantifying the similarity of topological domains across normal and cancer human cell types. Bioinformatics **34**, i475–i483 (2018).
136. Li, M. et al. An atlas of DNA methylomes in porcine adipose and muscle tissues. Nat. Commun. **3**, 850 (2012).
137. Ron, G., Globerson, Y., Moran, D. & Kaplan, T. Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. Nat. Commun. **8**, 2237 (2017).
138. Zhi, M. et al. Generation and characterization of stable pig pregastrulation epiblast stem cell lines. Cell Res. **32**, 383–400 (2022).
139. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP-seq enrichment using MACS. Nat. Protoc. **7**, 1728–1740 (2012).
140. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell **153**, 307–319 (2013).
141. Loven, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell **153**, 320–334 (2013).
142. Thorvaldsdóttir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. **14**, 178–192 (2013).
143. Koren, S. et al. De novo assembly of haplotype-resolved genomes with trio binning. Nat. Biotechnol. (2018).
144. Chen, Y. et al. Efficient assembly of nanopore reads via highly accurate and intact error correction. Nat. Commun. **12**, 60 (2021).
145. Li, H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**, 3094–3100 (2018).
146. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods **15**, 461–468 (2018).
147. Li, H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics **34**, 3094–3100 (2018).

148. Sedlazeck, F. J. et al. Accurate detection of complex structural variations using single-molecule sequencing. Nat. Methods **15**, 461–468 (2018).
149. Whalen, S., Truty, R. M. & Pollard, K. S. Enhancer-promoter interactions are encoded by complex genomic signatures on looping chromatin. Nat. Genet. **48**, 488–496 (2016).
150. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. Bioinformatics **27**, 1017–1018 (2011).
151. Kulakovskiy, I. V. et al. HOCOMOCO: expansion and enhancement of the collection of transcription factor binding sites models. Nucleic Acids Res. **44**, D116–125 (2016).
152. Yang, Y., Zhang, R., Singh, S. & Ma, J. Exploiting sequence-based features for predicting enhancer-promoter interactions. Bioinformatics **33**, i252–i260 (2017).
153. Chen, T., Guestrin, C. 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).
154. Friedman, J. H. Greedy function approximation: a gradient boosting machine. Annals of Statistics **29**, 1189–1232 (2001).
155. Huynh, L. & Hormozdiari, F. TAD fusion score: discovery and ranking the contribution of deletions to genome structure. Genome Biol. **20**, 60 (2019).
156. Stuart, T. et al. Comprehensive integration of single-cell data. Cell **177**, 1888–1902 e1821 (2019).
157. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. **20**, 296 (2019).
158. Maynard, K. R. et al. Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. Nat. Neurosci. **24**, 425–436 (2021).
159. McCarthy, D. J., Campbell, K. R., Lun, A. T. & Wills, Q. F. Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. Bioinformatics **33**, 1179–1186 (2017).
160. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. **43**, e47 (2015).
161. Zhou, Y. et al. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat. Commun. **10**, 1523 (2019).
162. Hoencamp, C. et al. 3D genomics across the tree of life reveals condensin II as a determinant of architecture type. Science **372**, 984–989 (2021).
163. Sievers, F. & Higgins, D. G. The clustal omega multiple alignment package. Methods Mol. Biol. **2231**, 3–16 (2021).
164. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res. **49**, D480–D489 (2021).
165. Buniello, A. et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. **47**, D1005–D1012 (2019).

Figures

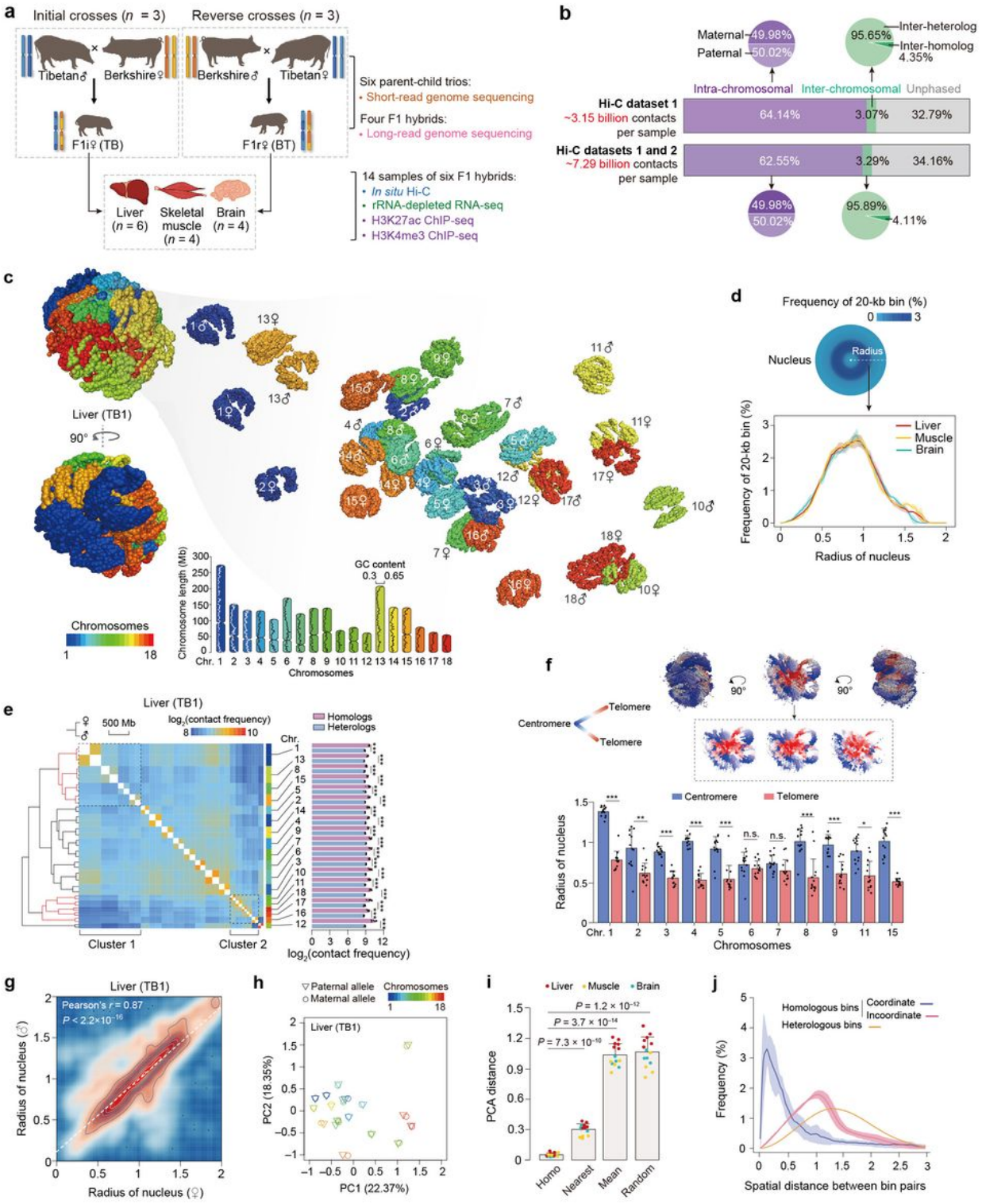

Figure 1

Three-dimensional (3D) structures of diploid genome in hybrid pigs.

a, Schematic representation of reciprocal crosses between genetically distinct Berkshire and Tibetan pig breeds. Liver, skeletal muscle, and brain of F1 hybrids were collected for multi-omics assays (see **Supplementary Fig. 1a** for details). F1i and F1r denote the F1 hybrids of initial and reverse crosses, respectively.

b, Allele assignment of *in situ* Hi-C contacts in F1 hybrids. A total of ~48.95 billion valid Hi-C contacts were generated, including ~3.15 billion autosomal contacts for each of the 14 samples (six liver, four muscle, and four brain samples; dataset 1). To reliably identify PEIs throughout the diploid genome, we carried out additional *in situ* Hi-C assays for 12 samples (dataset 2), and resulting in a final total of ~97.53 billion valid Hi-C contacts (~7.29 billion autosomal contacts for each of the 12 samples, including four liver, four muscle, and four brain samples) in the combined datasets 1 and 2 (see **Supplementary Fig. 1a** for details).

c, 3D genome structure of a representative liver sample (see Methods). Left: whole genome. Right*:* the 18 autosome pairs visualized separately.

d, 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.

e, 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., $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$).

f, 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 were compared. Data are presented as mean values ± SD ($n = 14$). Wilcoxon rank-sum test, n.s., $P \geq 0.05$; *$P < 0.05$; **$P < 0.01$; ***$P < 0.001$.

g, Homologs exhibited highly similar distances to the center of 3D nucleus (Pearson's $r = 0.87$, $P < 2.2 \times 10^{-16}$). The center of the 3D nucleus was calculated as the mean coordinate of all autosomes.

h, PCA analysis of autosomal interactions in a representative liver sample in which homologs that are closer in the plot have more similar interaction patterns.

i, 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 Student's *t*-test.

j, 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.

For Hi-C maps, **d**, **f**, **g**, and **j** show 20-kb resolution; C uses 100-kb resolution, 1-Mb resolution was used in **e**, **h**, and **i**. See also **Supplementary Fig. 1–5**.

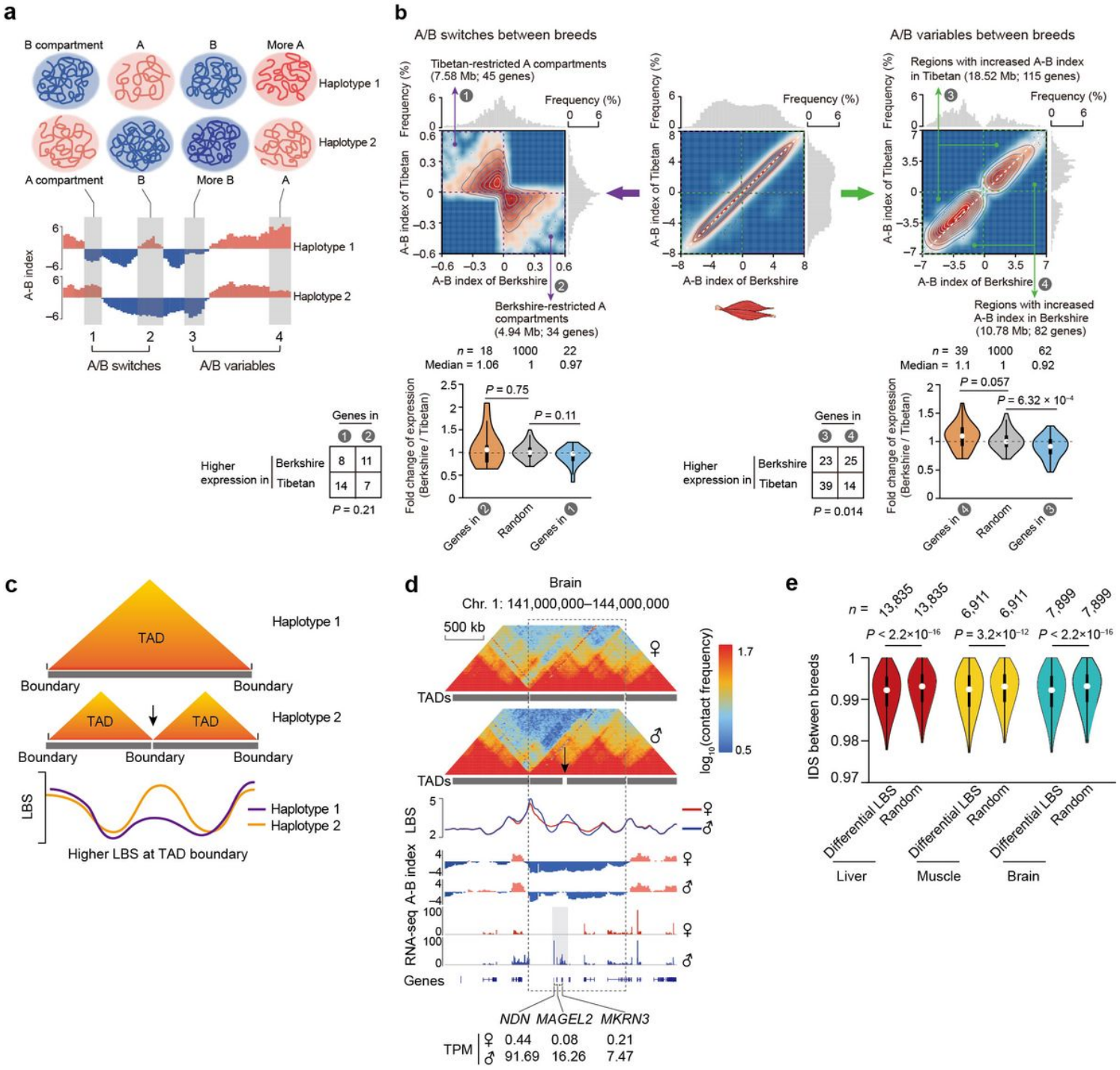

Figure 2

Compartmental rearrangements and variable TAD boundaries among haploid genomes.

a, Identification of compartmental rearrangements. Regions with distinct compartment status were defined (*i.e.*, from B [negative value of A-B index] to A [positive value of A-B index], or from A to B; scenarios 1 and 2, respectively) as A/B switched regions. Additionally, compartment scores (*i.e.*, the A-B index values) for regions with the same status were compared between haploid genomes (not between those with distinct switches in compartment status), and identified regions with significantly higher compartment scores as A/B variable regions ($P < 0.05$, paired Student's *t*-test, and $|\Delta$A-B index$| > 0.5$; scenarios 3 and 4).

b, Allelic compartmental rearrangements in muscle.We identified compartment A regions exclusively found in the Tibetan (7.58 Mb) or Berkshire (4.94 Mb) alleles (top left), and regions with significantly elevated compartment scores in of the Tibetan (18.52 Mb) and Berkshire (10.78 Mb) alleles (top right). Genes with testable differences in allele (*i.e.*, with informative SNVs and TPM ≥ 0.5 in at least one allele) tended to be located in the more accessible chromatin regions and generally showed increased expression; violin plots show changes in allelic expression (Wilcoxon rank-sum test) and counts (Fisher's exact test) of genes located in regions with allelic differences in compartmentalization (*i.e.*, A/B switches, left; and A/B variables, right).

c, Identification of shifted TAD boundaries (see **Methods**).

d, Paternal-specific TAD boundaries adjacent to three paternally-expressed imprinted genes (*NDN*, *MAGEL2* and *MKRN3*).

e, In comparison to genomic background, regions with allelic differences in boundary strength (LBS, $P < 0.05$, paired Student's *t*-test) exhibited greater allelic divergence in sequence divergence (reflected by the lower pairwise haplotype similarities between two parental breeds determined by identity score [IDS]; **see Methods**). *P* values are from Wilcoxon rank-sum test.

For Hi-C maps, 20-kb resolution were used. See also **Supplementary Fig. 6–10** and **Supplementary Data 1–6**.

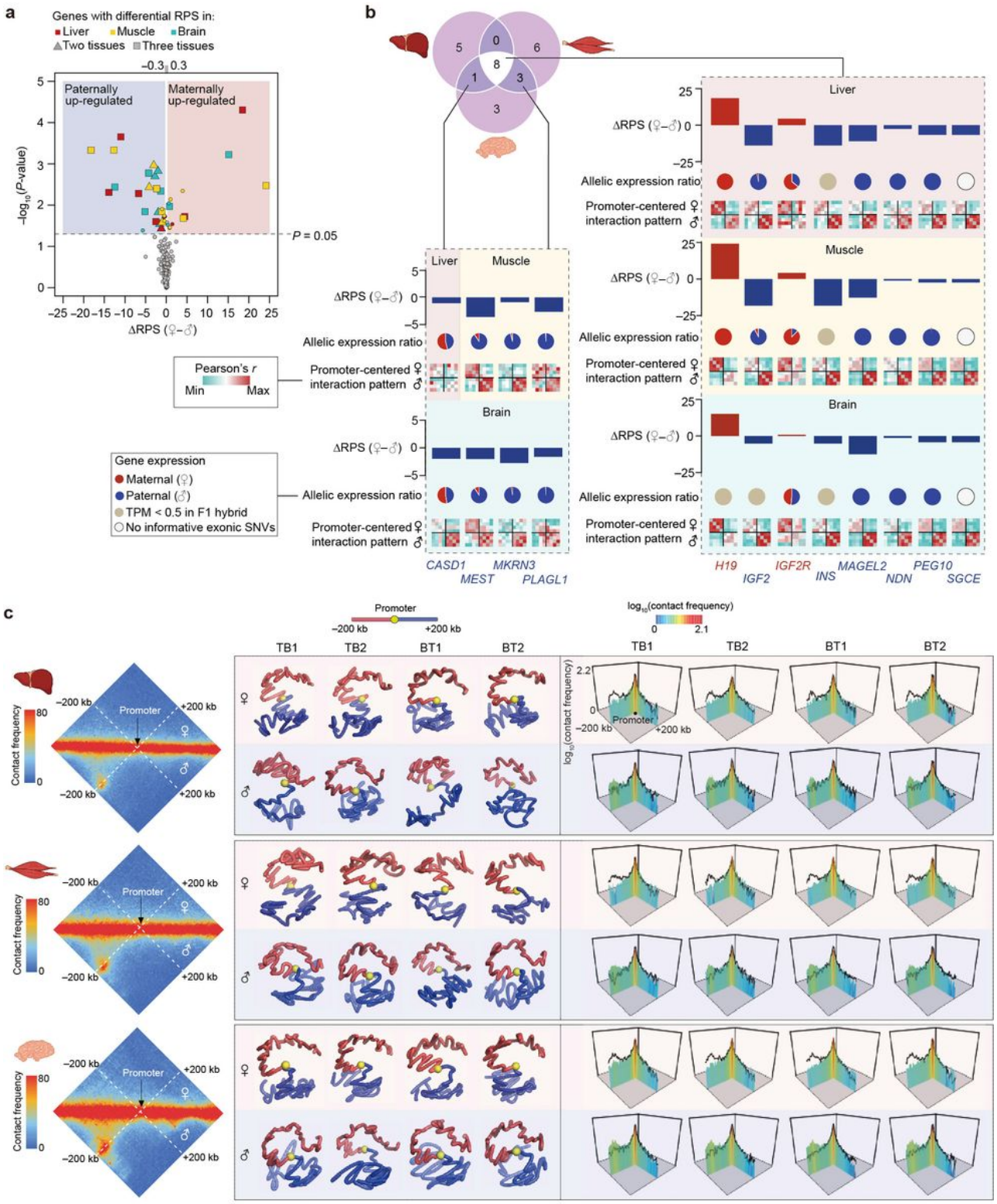

Figure 3

Parent-of-origin-specific PEI organization of imprinted genes.

a, Quantification of PEIs in 126 previously described imprinted genes identified 101 genes with informative SNVs which interacted with enhancers in at least one tissue. Plots of differential RPS

between parents-of-origin in at least two tissues are shown for 12 representative imprinted genes (|ΔRPS| > 0.3 and $P < 0.05$, paired Students *t*-test).

b, Allelic differences in RPS and expression of 12 imprinted genes with differential RPS listed in (A). The similarity (Pearson's *r*) of interactions in promoter-centered regions (200 kb up- and downstream from the promoter bin) among eight haplotypes are shown for each tissue.

c, Allelic promoter-centered interactions in the representative imprinted gene *MAGEL2*. Haplotype-resolved Hi-C maps (left*)*, 3D structural models (middle), and interaction metaplots (right) of promoter-centered regions across replicates and tissues. Enhanced paternal allele-specific interactions between the promoter and its 200 kb upstream regions. Black lines in metaplots show interactions of the other parental allele.

Hi-C maps in **a**–**c** are shown at 5-kb resolution. See also **Supplementary Fig. 11** and **15**, and **Supplementary Data 8**.

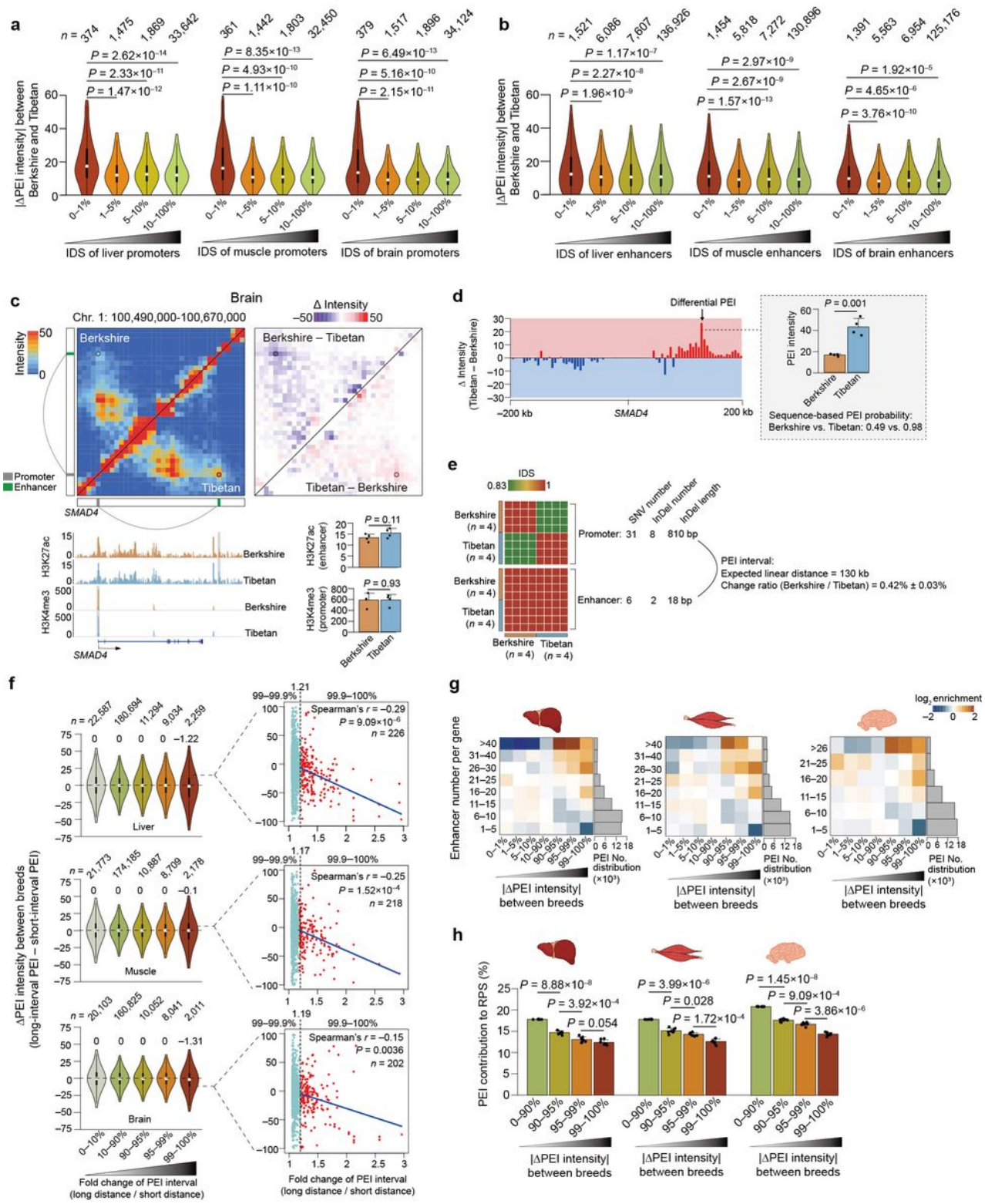

Figure 4

Influence of allelic sequence variations on PEIs.

a, **b,** Promoters (**a**) and enhancers (**b**), with higher allelic sequence divergence (especially those with the lowest 1% of IDS between parental breeds; see **Methods**) showed greater differences in PEI intensity. *P*-values are from 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 **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, Effects on PEI intensity caused by allelic variation-induced differences in PEI bridging distances. PEIs with the top 1% of increases in bridging distance showed significantly decreased PEI intensity (0–99% vs. 99–100%: $P = 5.03 \times 10^{-12}$ in liver, $P = 2.21 \times 10^{-6}$ in muscle, and $P = 3.26 \times 10^{-10}$ in brain; Wilcoxon rank-sum test; left). This trend is more prominent in PEIs with the greatest extension (allelic fold change in linear bridging distance of the top 0.1%: 1.17–1.21), which exhibited strongly negative correlations between bridging distance and PEI intensity of PEIs (Spearman's $r = -0.15$ to -0.29, $P \leq 0.0036$; right).

g, PEIs with enrichment for differential allelic intensity (x-axis) among genes that contact a variable number of enhancers (y-axis) in the three tissues.

h, PEIs showing greater differential allelic intensity tend to contribute less to RPS (including the combined regulatory effects of multiple enhancers for a given gene), and thus exert the least impacts on gene expression. *P*-values are from paired Student's *t*-test.

Hi-C maps in **a**–**e** and **g**, **h** are shown at 5-kb resolution. See also **Supplementary Fig. 16** and **17**, and **Supplementary Data 9**.

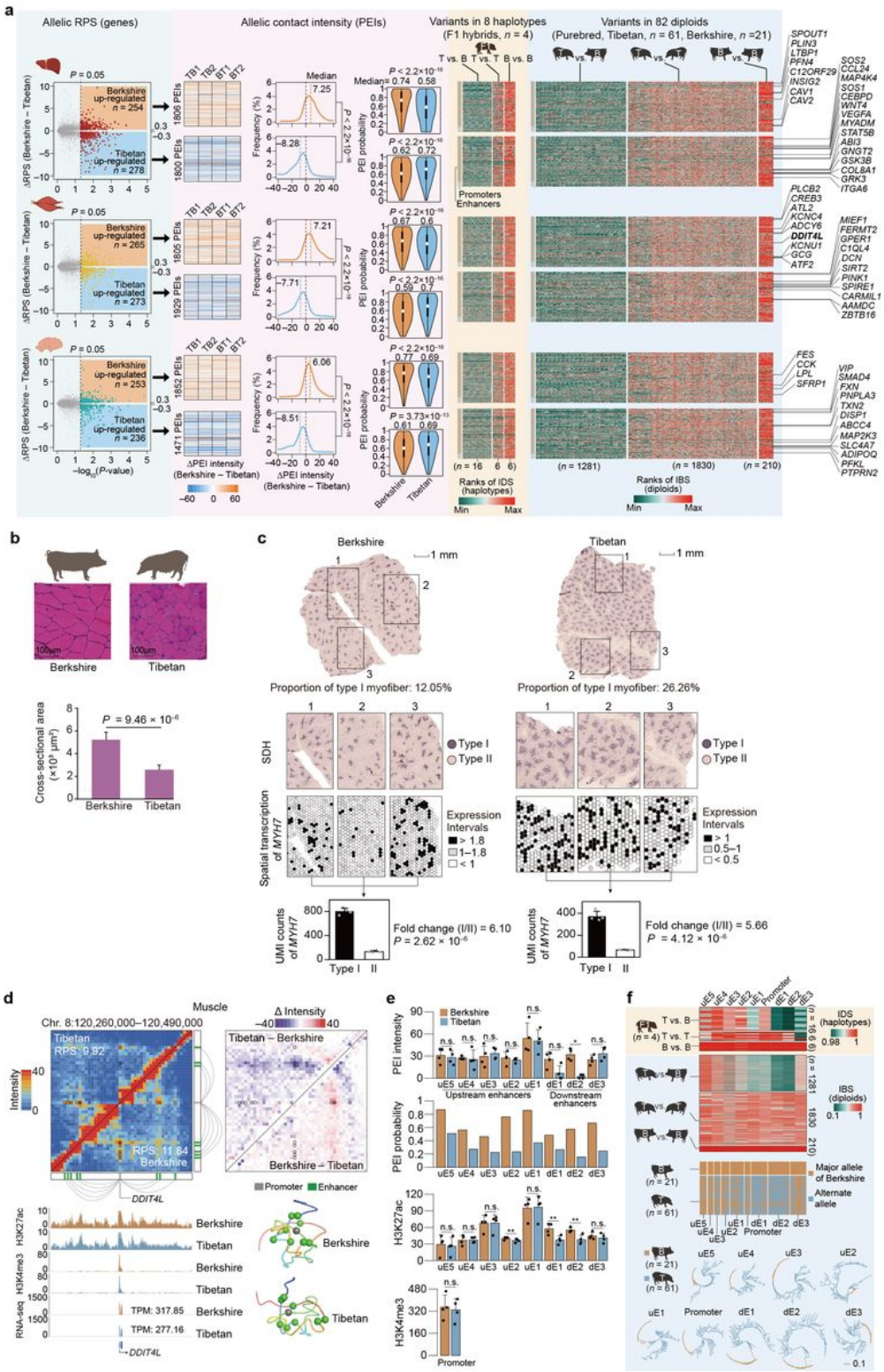

Figure 5

Allelic rewiring of PEIs in F1 hybrids.

a, Volcano plots of genes with differential RPS (|ΔRPS| > 0.3 and $P < 0.05$, paired Student's *t*-test) between Berkshire and Tibetan alleles in three tissues (left). Differential PEI intensities, distributions of differential PEI intensities, and estimated probabilities of PEIs based on their sequence variants (see

Methods) are shown (middle). Degree of sequence similarity for promoters and enhancers in pairwise comparisons among eight haplotypes (Tibetan [*n* = 4] and Berkshire [*n* = 4]; measured by IDS) and among 82 diploid genomes in population-level analysis of purebred pigs (measured by pairwise IBS distances; see Methods) (right). Among the 82 diploids, 12 arefrom 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]); SNVs were retrieved from the ISwine database [http://iswine.iomics.pro/pig-iqgs/iqgs/index]). Representative functional genes are labelled.

b, Comparison of myofiber cross-sectional area in representative *longissimus dorsi*muscle (hematoxylin-and-eosin staining) between adult purebred Berkshire and Tibetan pigs (see Methods). *P*-value is from paired Student's *t*-test.

c, Spatial transcriptomic profiles of type I myofiber marker *MYH7* expression in succinate dehydrogenase (SDH)-stained histological sections. 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 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). Neighbor-joining phylogenetic trees of a promoter and eight enhancers were separately constructed based on pairwise IBS distances (bottom). Note that the genetically distinct clusters between breeds, with high homozygosity in the Berkshire population. See also **Supplementary Fig. 19–21** and **Supplementary Data 9**.

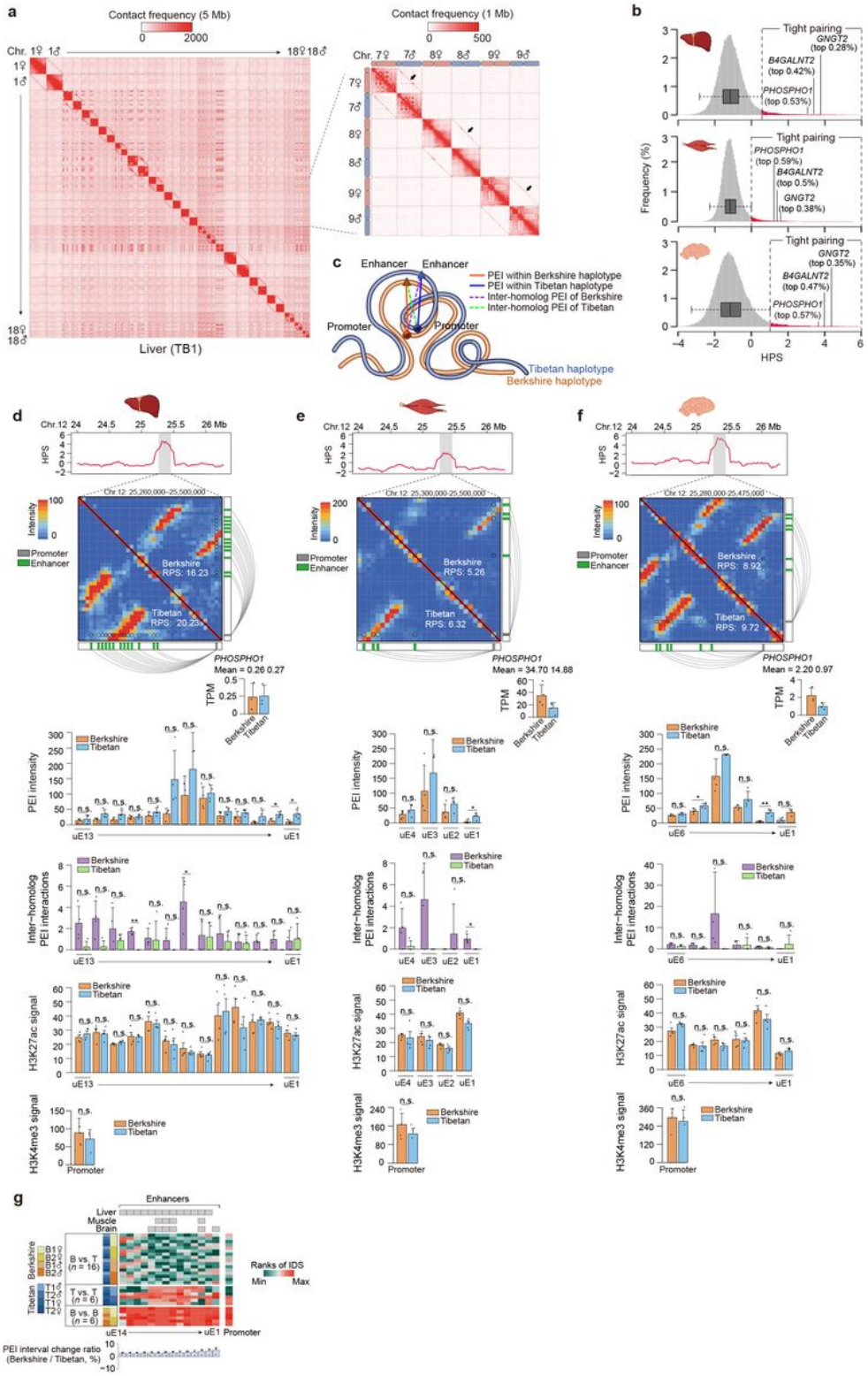

Figure 6

Highly structured homolog pairing facilitates regulatory interactions between homologs.

a, Hi-C maps showing signals of genome-wide homolog pairing (arrows) in representative somatic tissue (liver) of a hybrid pig. Magnified Hi-C maps of chromosomes 7, 8 and 9 are shown on the right.

b, Identification of tightly-paired loci in all three somatic tissues of hybrid pig based on distribution of homolog pairing score (HPS) values. In the boxplot, the internal line indicates the median, the box limits indicate the 25^{th} and 75^{th} quartiles and the whiskers extend to 1.5 × IQR from the quartiles. Rare tightly-paired loci (liver, 76.74 Mb; muscle, 72.12 Mb; brain, 74.18 Mb) are defined as those with HPS above Q3 + 1.5 × IQR. Percentiles of HPS for *PHOSPHO1* as well as *B4GALNT2* and *GNGT2* are shown.

c, Schematic representation of intra-chromosomal (solid lines) and inter-homolog (dashed lines) promoter-enhancer interactions in tightly-paired loci; the former are generally much stronger than the latter.

d–**f**, Evidence of unbalanced inter-homolog *PHOSPHO1* PEIs in the genomes of liver (**d**), muscle (**e**), and brain (**f**) in hybrid pigs. Top to bottom: HPS values and Hi-C maps of the corresponding genomic regions (with schematic representation of allelic PEI organization). Allelic differences in gene expression, intra-chromosomal and inter-homolog PEI intensity, and ChIP-seq signals of H3K27ac (enhancers) and H3K4me3 (promoter) are also shown (mean ± SD, $n = 4$). *P*-values are from paired Student's *t*-test. *PHOSPHO1* is in a tightly-paired locus in all three tissues of (**b**). The intra-chromosomal PEIs of *PHOSPHO1* were significantly greater in the Tibetan allele compared with in the Berkshire allele of the three tissues, 2 (of 13, liver), 1 (of 4, muscle), and 2 (of 6, brain), resulting in increased RPS for *PHOSPHO1* in the Tibetan allele, although its expression was not correspondingly increased in the allele. This effect may be attributable to the increased inter-homolog PEIs between enhancers in the Tibetan allele and the promoter in the Berkshire allele. Data in E–G show means ± SD ($n = 4$), and dots indicate values of each haplotype ($n = 4$). *P* values are from paired Student's *t*-test. n.s., $P > 0.05$; *$P < 0.05$; **$P < 0.01$.

g, Degree of haplotype similarities (measured by IDS) of promoter and enhancers in pairwise comparisons among eight haplotypes. The changes in PEI interval length (mean ± SD, $n = 4$) are also shown. Top: Schematic representation of enhancers across three tissues organized from upstream to downstream of the promoter.

Hi-C maps in **b**, and **c**, are shown at 20-kb resolution, 5-kb resolution in **d**–**f**, and 1-Mb resolution in **a**. See also **Supplementary Fig. 22**–**25**.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Supplementaryinformation1123OK.docx