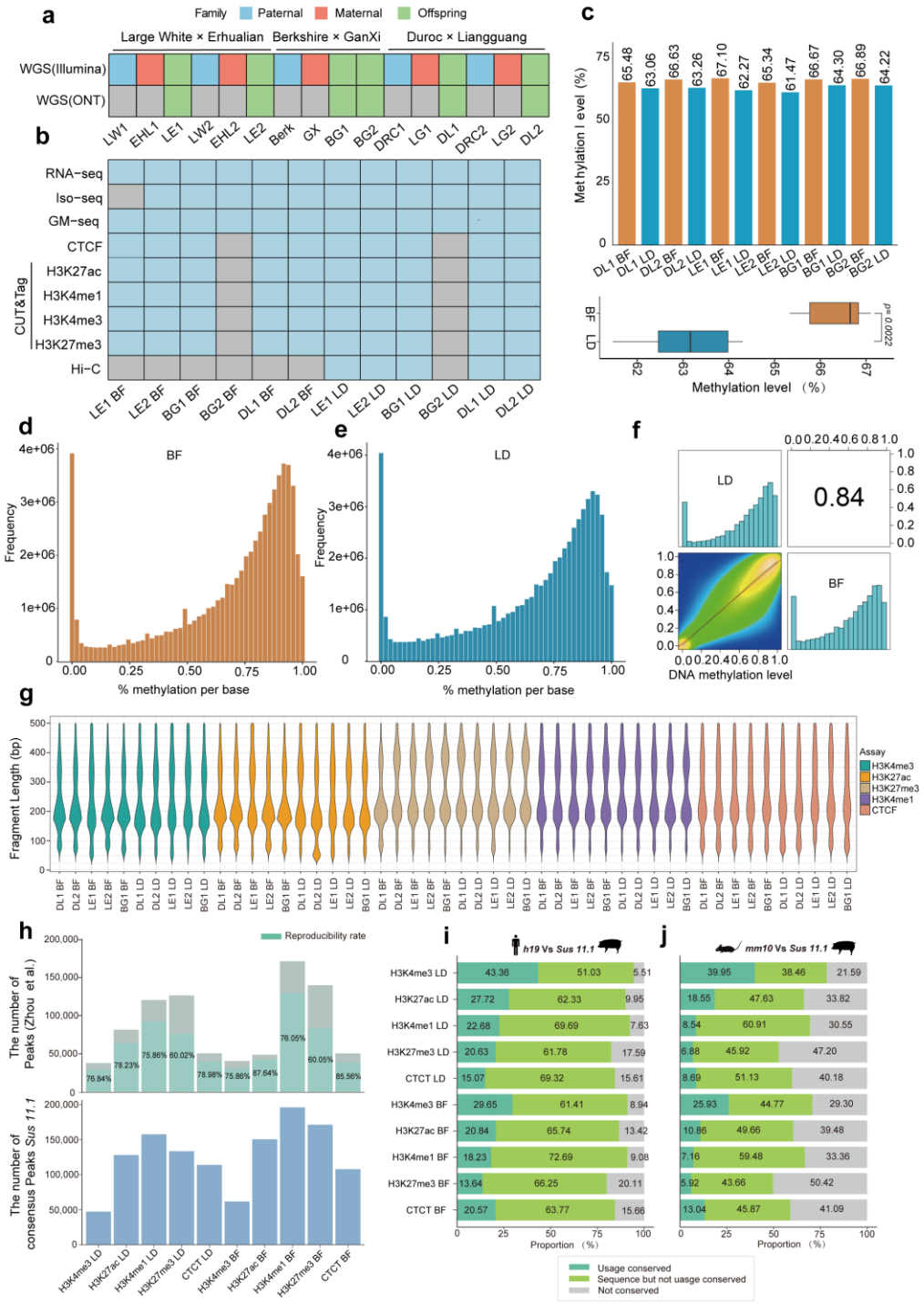
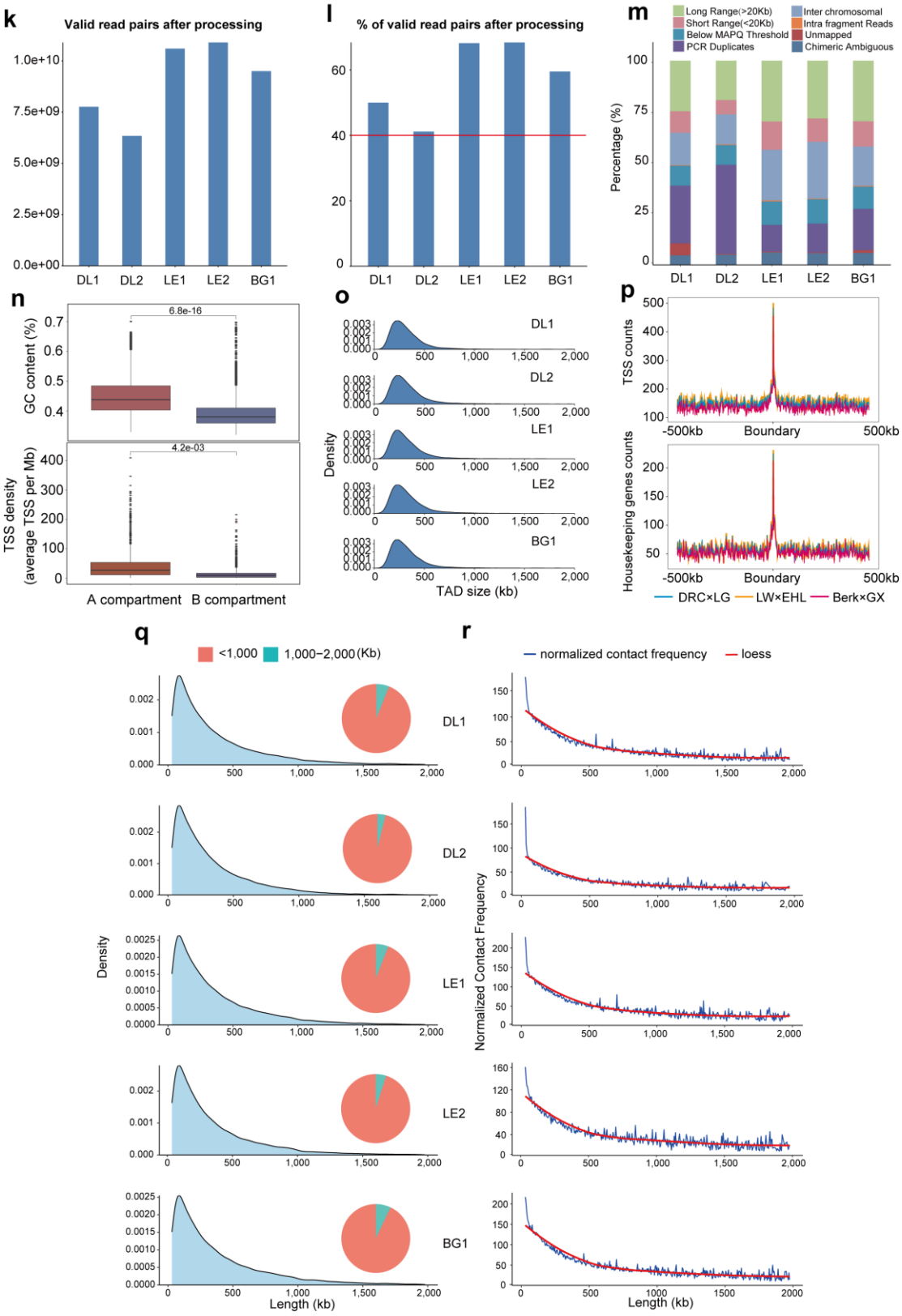
Mechanism of Parent-of-origin Effects revealed by Multi-omic data in Euro-Chinese Hybrid Pigs

The supplementary information contains:

Supplementary Figures

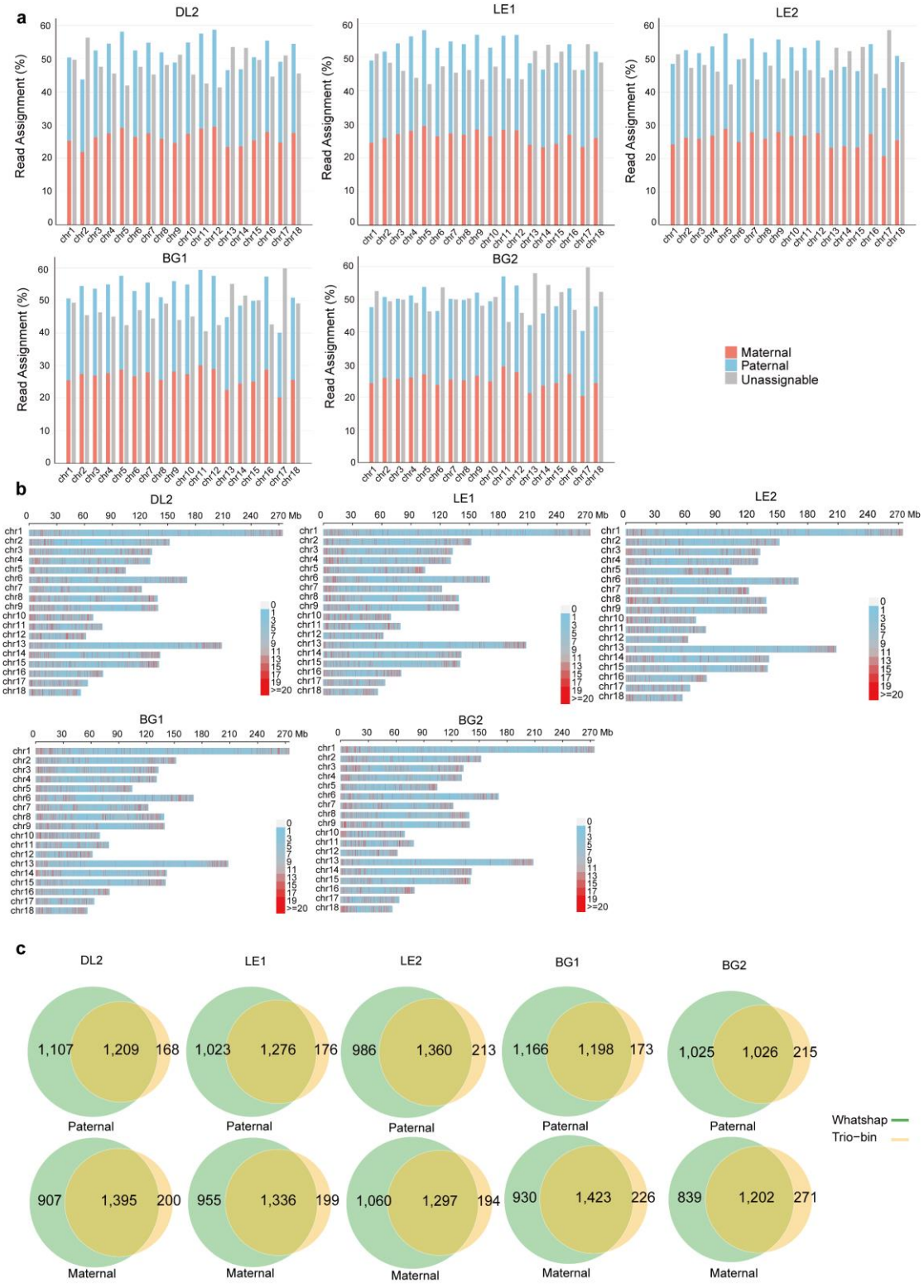
Supplementary Figures and Legends




k
Valid read pairs after processing
1.0e+10
7.5e+09
5.0e+09
2.5e+09
0.0e+00
DL1 DL2 LE1 LE2 BG1
l
% of valid read pairs after processing
60
40
20
0
DL1 DL2 LE1 LE2 BG1
m
Long Range(>20Kb)
Short Range(<20Kb)
Below MAPQ Threshold
PCR Duplicates
Inter chromosomal
Intra fragment Reads
Unmapped
Chimeric Ambiguous
Percentage (%)
100
75
50
25
0
DL1 DL2 LE1 LE2 BG1
n
6.8e-16
GC content (%)
0.7
0.6
0.5
0.4
4.2e-03
TSS density
(average TSS per Mb)
400
300
200
100
0
A compartment B compartment
o
DL1
DL2
LE1
LE2
BG1
Density
0.003
0.002
0.001
0.000
0 500 1,000 1,500 2,000
TAD size (kb)
p
TSS counts
500
400
300
200
100
-500kb Boundary 500kb
Housekeeping genes counts
200
150
100
50
-500kb Boundary 500kb
DRC×LG
LW×EHL
Berk×GX
q
<1,000
1,000−2,000 (Kb)
DL1
DL2
LE1
LE2
BG1
Density
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
0 500 1,000 1,500 2,000
Length (kb)
r
normalized contact frequency
loess
Normalized Contact Frequency
150
100
50
0
200
160
120
80
40
0 500 1,000 1,500 2,000
Length (kb)

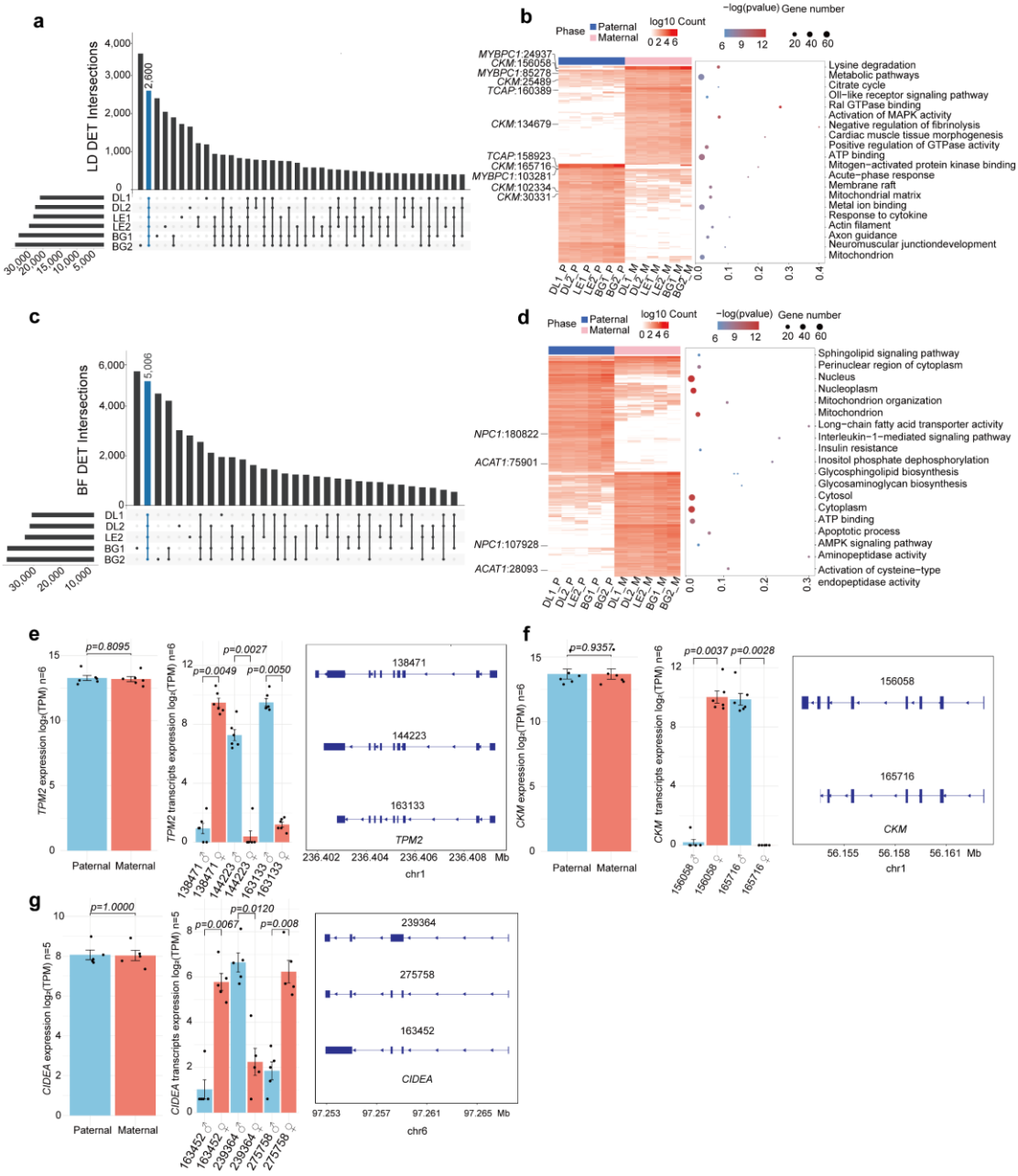
Supplementary Fig.1 | Samples composition and multi-omics analysis performed in this study. a Information on the trio family composition and WGS of parents and offspring. **b** The display of transcriptomic and epigenomic sequencing information for hybrid pigs in BF and LD. **c** The average DNA methylation levels in offspring individuals and the significant difference between BF and LD (Wilcox test, p=0.0022). **d** The frequency distribution of genome-wide CpG site methylation levels in BF. **e** The frequency distribution of genome-wide CpG site methylation levels in LD. **f** The distribution of average DNA methylation levels in the BF and LD of all hybrid pigs and the correlation of average methylation levels between tissues. **g** The fragment distribution for each CUT&Tag sample. **h** The number of consensus peaks for histone modifications and CTCF identified in different tissues in this study (bottom) and their reproducibility in the corresponding assays from the previous study by Zhou et al. in pigs (top). **i** The sequence and functional conservation of regions characterized by histone modifications and CTCF in this study were compared with those in humans. **j** The sequence and functional conservation of regions characterized by histone modifications and CTCF in this study were compared with those in mice. **k** The number of valid read pairs that were mapped to the reference genome in each sample after processing using the Juicer. **l** The percentage of valid read pairs that were mapped to the reference genome in each sample after processing. **m** Percentage of mapped reads with high mapping quality (such as long range, short range, inter-chromosomal) which were used to generate the Hi-C map, as well as the read pairs classes which

did not included in the Hi-C contact map (such as unmapped reads, read-pairs with low mapping quality, intra-fragment reads and duplicates). **n** Box plot of the GC content of the A compartment and B compartment of samples in LD (upper). Box plot of the TSS densities in the A compartment and B compartment of samples in LD (bottom). Both significant differences identified using the Wilcoxon test. **o** The length distribution of TADs in LD. **p** Enrichment of protein-coding genes of TSS near TAD boundaries, and the proportion of these genes that are housekeeping genes. **q** For each sample, the loop length distribution obtained at 1 kb resolution was analyzed, along with the percentages of loops above and below 1 Mb in length. **r** The relationship between loop length and interaction strength in each sample, with the red line representing the linear fit.



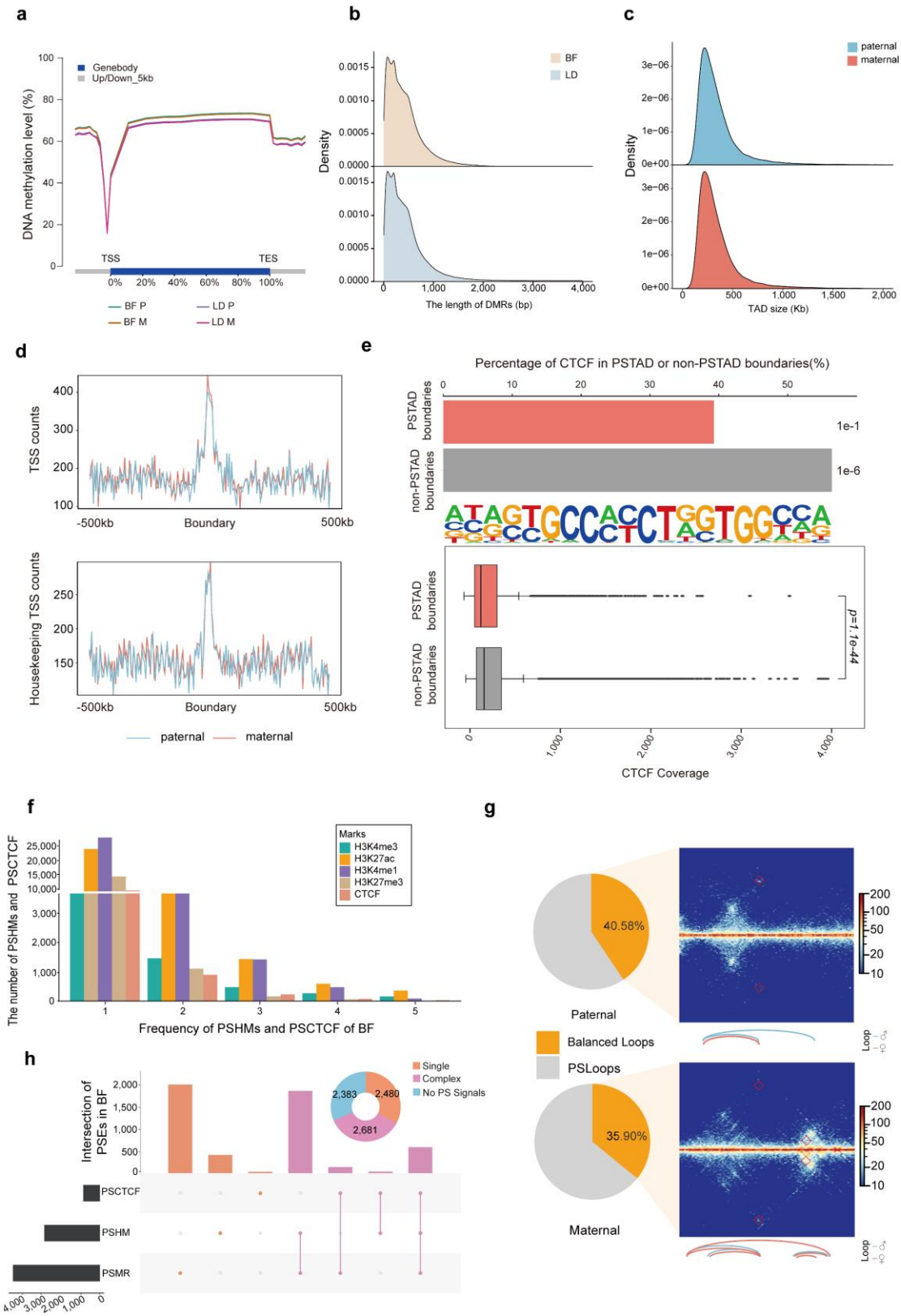
Supplementary Fig.2 | Phasing information of hybrid pigs. a The phasing efficiency among all autosomes of the rest samples. **b** The SNP density among all autosomes of sample of the rest samples. **c** Comparing the shared and differential k-mers (Million)

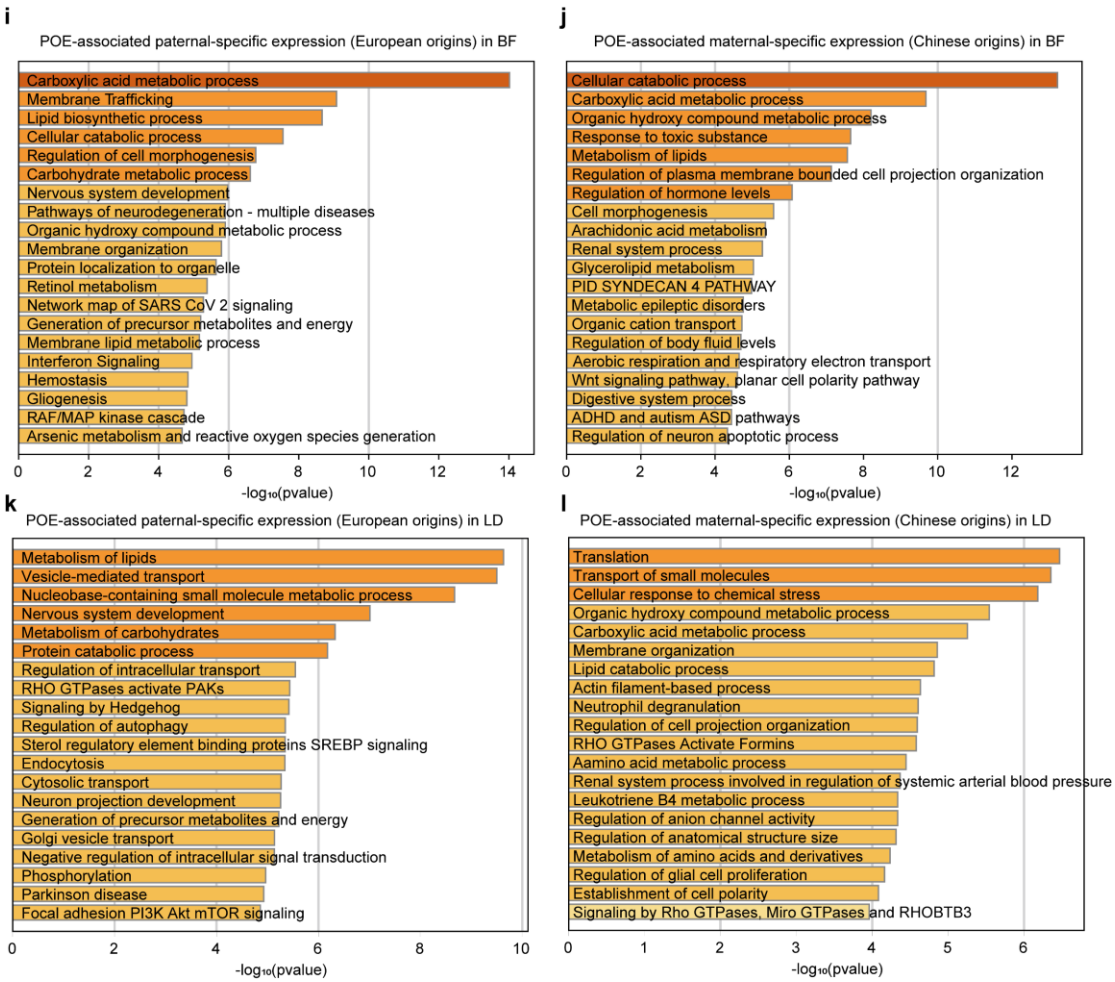
obtained through Phase-Tag and Trio-bin phasing strategies in rest of samples.



Supplementary Fig.3 | Differentia expression at the transcript levels between parental phases. a Differential expressed transcripts (DETs) identified between different phases in LD of each sample. Blue bar represents the sample-shared DETs. **b** In LD**,** the sample-shared DETs between phases and genes belong (left), the pathway enrichment analysis (GO & KEGG) of genes containing sample-shared DETs (right). **c** Differential expressed transcripts (DETs) identified between different phases in BF of

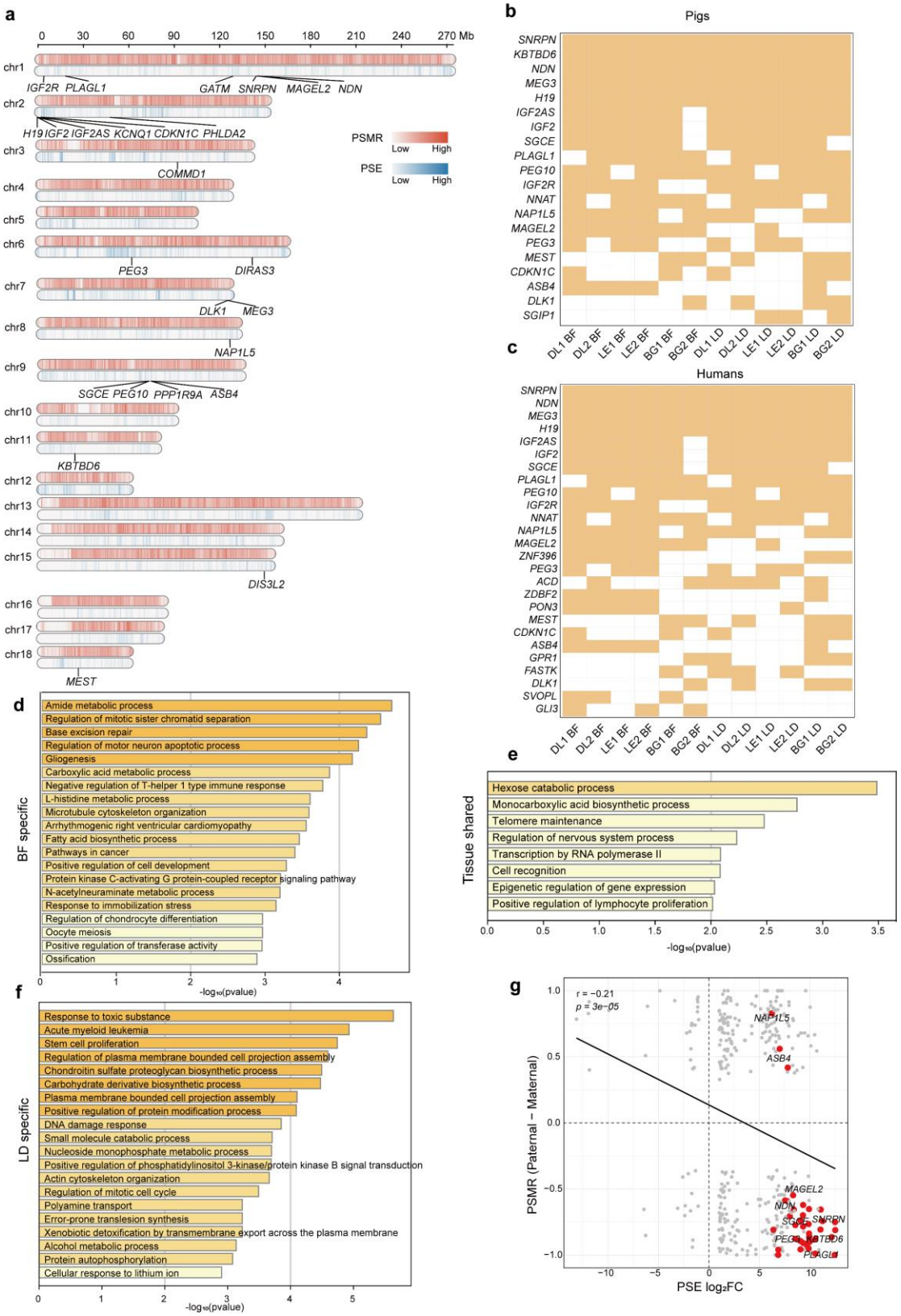
each sample. Blue bar represents the sample-shared DETs. **d** In BF**,** the sample-shared DETs between phases and genes belong (left), the pathway enrichment analysis (GO & KEGG) of genes containing sample-shared DETs (right). **e-g** depict examples of genes (*CIDEA* in BF and *TPM2*, *CKM* in LD) that did not exhibit PSE at the gene level, but exhibited DET between phases. In each case (left panel), there is no significant difference in the overall expression level of the gene between phases as determined by the Wilcoxon test. However, phase-specific transcripts (middle panel) show distinct expression patterns between phases, with significant differences identified using the Wilcoxon test. The right panel illustrates the structures of these phase-specific transcripts for each gene (*CIDEA*, *TPM2*, and *CKM*).

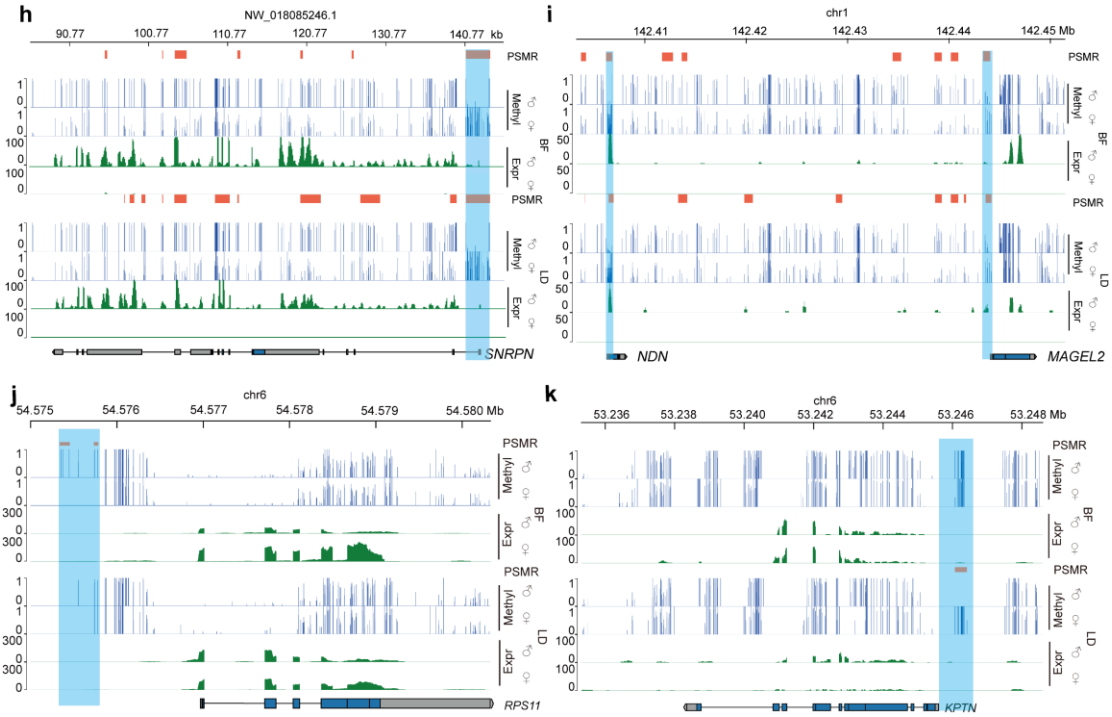

a
Genebody
Up/Down_5kb
DNA methylation level (%)
TSS
TES
0% 20% 40% 60% 80% 100%
BF P
BF M
LD P
LD M
b
BF
LD
Density
The length of DMRs (bp)
c
paternal
maternal
Density
TAD size (Kb)
d
TSS counts
-500kb
Boundary
500kb
Housekeeping TSS counts
paternal
maternal
e
Percentage of CTCF in PSTAD or non-PSTAD boundaries(%)
PSTAD boundaries
non-PSTAD boundaries
1e-1
1e-6
p=1.1e-44
CTCF Coverage
f
Marks
H3K4me3
H3K27ac
H3K4me1
H3K27me3
CTCF
The number of PSHMs and PSCTCF
Frequency of PSHMs and PSCTCF of BF
g
40.58%
Paternal
Balanced Loops
PSLoops
35.90%
Maternal
Loop
h
Single
Complex
No PS Signals
Intersection of PSEs in BF
2,383
2,480
2,681
PSCTCF
PSHM
PSMR



Supplementary Fig.4 | The distribution and evaluation of phase-specific multi-omics signal. a The average DNA methylation levels near protein-coding genes, including the transcription start site (TSS) and the transcription end site (TES), between different phases in BF and LD. **b** The average length distribution of PSMRs between parental phases in BF and LD. **c** The average length distribution of TAD identified of parental phases in LD. d Enrichment of protein-coding genes of TSS near phased TAD boundaries, and the proportion of these genes that are housekeeping genes. **e** Proportions of CTCF binding at PSTAD and non-PSTAD boundaries, and significance of canonical CTCF motif prediction (top); Nineteen core motifs of canonical CTCF (middle); CTCF binding intensity at PSTAD and non-PSTAD boundaries (bottom). **f**

The frequency of identified PSHMs and PSCTCF across all samples in BF. The paired t-test was used to calculate the P-value. g The percentage of phased-specific loops among the phased loops (left), and the pattens of a typical phase-specific loop (right). **h** An overview of the number of PSEs associated with single or complex phase-specific epigenomic marks in BF. **i-l** The biological functions and pathways enriched in genes with POE-associated parent-specific expression in BF and LD. (i) POE-associated paternal-specific expression (European origins) in BF (j) POE-associated maternal-specific expression (Chinese origins) in BF (k) POE-associated paternal-specific expression (European origins) in LD (l) POE-associated maternal-specific expression (Chinese origins) in LD.

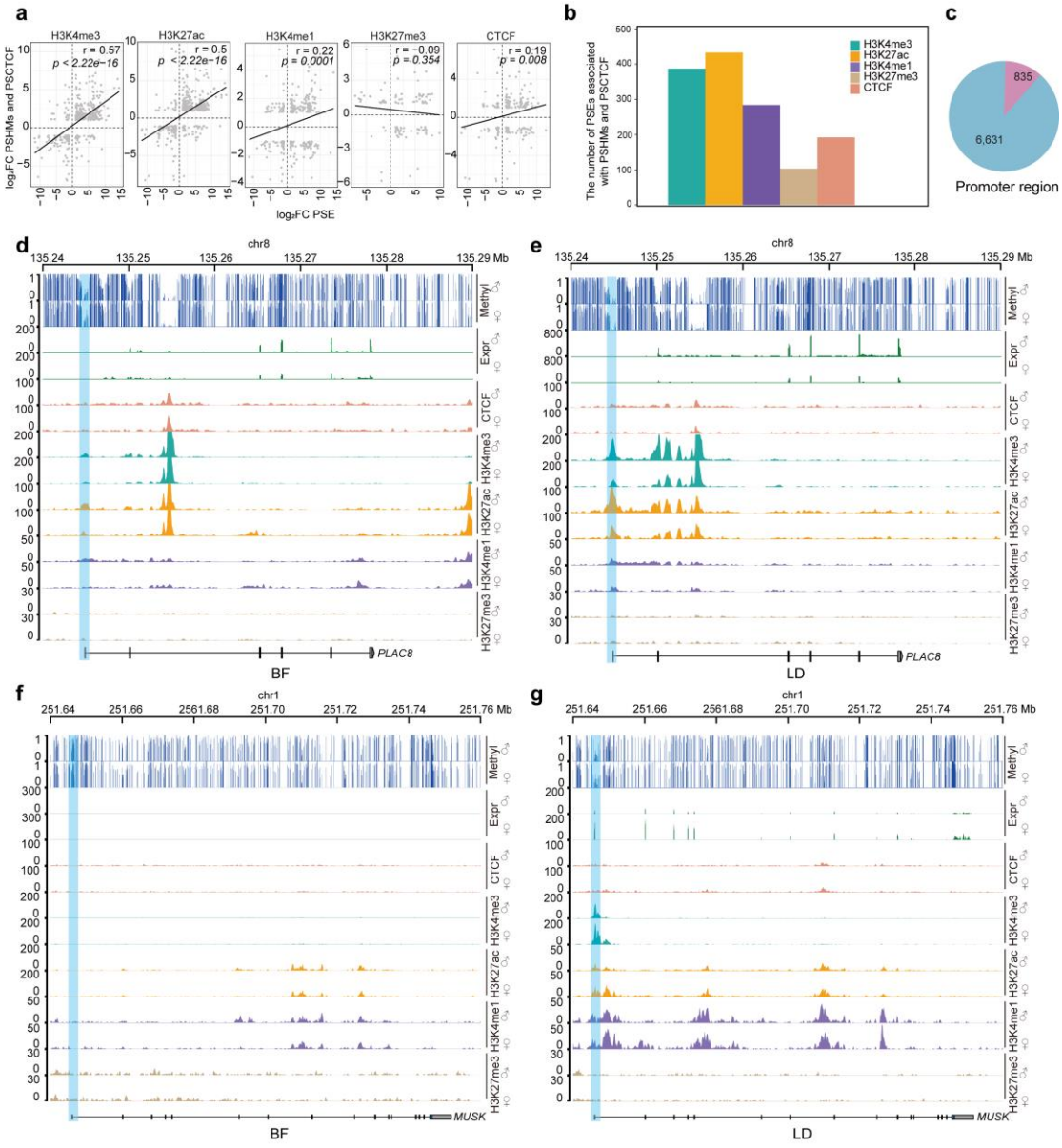

a
0 30 60 90 120 150 180 210 240 270 Mb
chr1
IGF2R PLAGL1
GATM SNRPN MAGEL2 NDN
chr2
H19 IGF2 IGF2AS KCNQ1 CDKN1C PHLDA2
chr3
COMMD1
chr4
chr5
chr6
PEG3
DIRAS3
chr7
DLK1 MEG3
chr8
NAP1L5
chr9
SGCE PEG10 PPP1R9A ASB4
chr10
chr11
KBTBD6
chr12
chr13
chr14
chr15
DIS3L2
chr16
chr17
chr18
MEST
PSMR
Low High
PSE
Low High
b
Pigs
SNRPN
KBTBD6
NDN
MEG3
H19
IGF2AS
IGF2
SGCE
PLAGL1
PEG10
IGF2R
NNAT
NAP1L5
MAGEL2
PEG3
MEST
CDKN1C
ASB4
DLK1
SGIP1
DL1 BF DL2 BF LE1 BF LE2 BF BG1 BF BG2 BF DL1 LD DL2 LD LE1 LD LE2 LD BG1 LD BG2 LD
c
Humans
SNRPN
NDN
MEG3
H19
IGF2AS
IGF2
SGCE
PLAGL1
PEG10
IGF2R
NNAT
NAP1L5
MAGEL2
ZNF396
PEG3
ACD
ZDBF2
PON3
MEST
CDKN1C
ASB4
GPR1
FASTK
DLK1
SVOPL
GLI3
DL1 BF DL2 BF LE1 BF LE2 BF BG1 BF BG2 BF DL1 LD DL2 LD LE1 LD LE2 LD BG1 LD BG2 LD
d
BF specific
Amide metabolic process
Regulation of mitotic sister chromatid separation
Base excision repair
Regulation of motor neuron apoptotic process
Gliogenesis
Carboxylic acid metabolic process
Negative regulation of T-helper 1 type immune response
L-histidine metabolic process
Microtubule cytoskeleton organization
Arrhythmogenic right ventricular cardiomyopathy
Fatty acid biosynthetic process
Pathways in cancer
Positive regulation of cell development
Protein kinase C-activating G protein-coupled receptor signaling pathway
N-acetylneuraminate metabolic process
Response to immobilization stress
Regulation of chondrocyte differentiation
Oocyte meiosis
Positive regulation of transferase activity
Ossification
0 1 2 3 4
-log10(pvalue)
e
Tissue shared
Hexose catabolic process
Monocarboxylic acid biosynthetic process
Telomere maintenance
Regulation of nervous system process
Transcription by RNA polymerase II
Cell recognition
Epigenetic regulation of gene expression
Positive regulation of lymphocyte proliferation
0.0 0.5 1.0 1.5 2.0 2.5 3.0 3.5
-log10(pvalue)
f
LD specific
Response to toxic substance
Acute myeloid leukemia
Stem cell proliferation
Regulation of plasma membrane bounded cell projection assembly
Chondroitin sulfate proteoglycan biosynthetic process
Carbohydrate derivative biosynthetic process
Plasma membrane bounded cell projection assembly
Positive regulation of protein modification process
DNA damage response
Small molecule catabolic process
Nucleoside monophosphate metabolic process
Positive regulation of phosphatidylinositol 3-kinase/protein kinase B signal transduction
Actin cytoskeleton organization
Regulation of mitotic cell cycle
Polyamine transport
Error-prone translesion synthesis
Xenobiotic detoxification by transmembrane export across the plasma membrane
Alcohol metabolic process
Protein autophosphorylation
Cellular response to lithium ion
0 1 2 3 4 5
-log10(pvalue)
g
r = −0.21
p = 3e−05
NAP1L5
ASB4
MAGEL2
NDN
SGCE
SNRPN
PEG3
KBTBD6
PLAGL1
1.0 0.5 0.0 −0.5 −1.0
PSMR (Paternal − Maternal)
−10 −5 0 5 10
PSE log2FC

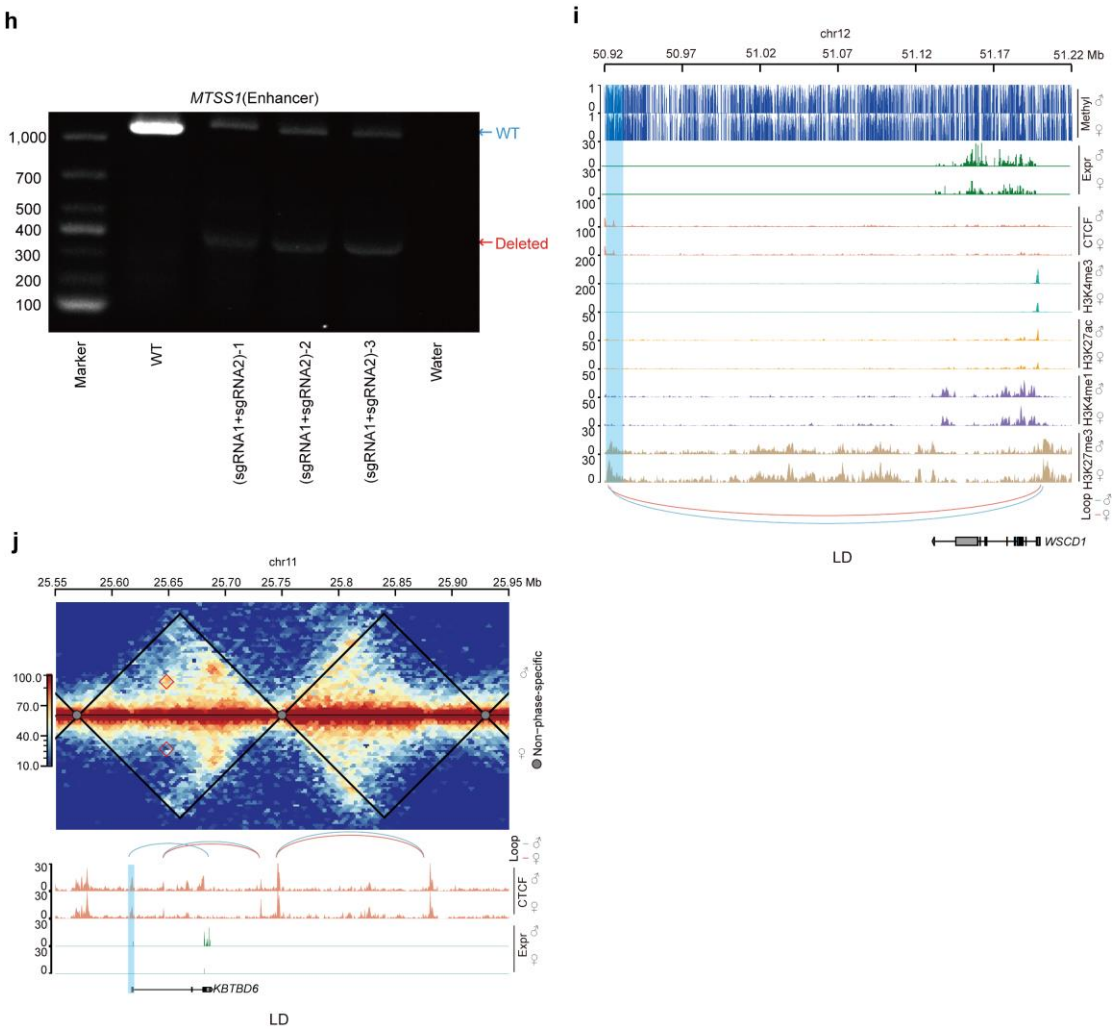


Supplementary Fig.5 | PSEs and imprinted genes regulated by DNA methylation

a The genome-wide (autosomal) distribution of PSMRs and PSEs identified in BF. The imprinted genes with significant expression differences between phases were labeled in corresponding genomic regions. **b-c** Known imprinted genes present in tissue-conserved PSE and duplicates among samples were identified referring to the imprinted gene repositories of pigs and humans. **d-f** Functional enrichment of imprinted-like genes: BF-specific (d), tissue-shared (e), and LD-specific (f). **g** The correlation between differential gene expression and promoter methylation levels between phases in BF, with red scatter points indicating imprinted genes across various samples. The cor.test function was used to calculate the Pearson correlation coefficient (r) and the corresponding p-values from the two-sided t-test. **h-i** The tracks for PSMR, DNA methylation level, and gene expression level of *SNRPN*, *NDN* and *MAGEL2* between the paternal phase and the maternal phase of BF and LD. **j** The example of BF-specific

PSE of *RPS11* regulated by PSMR within promoter. **k** The example of LD-specific PSE of *KPTN* regulated by PSMR within promoter.

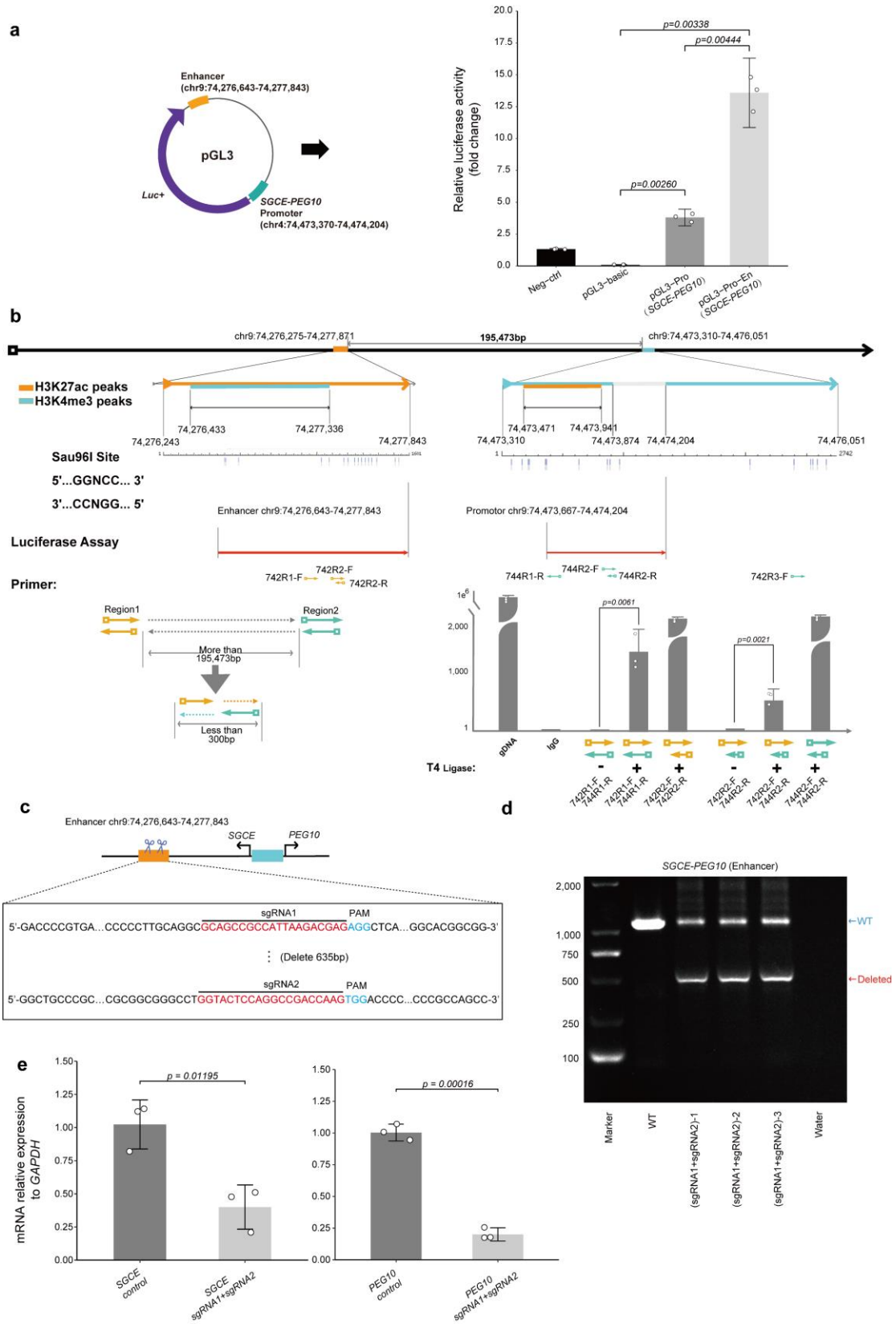




Supplementary Fig.6 | The POE form PSHM on PSEs and the complex POE model

a In BF, the correlation between PSEs and intensities of different PSHMs within promoter regions. The cor.test function was used to calculate the Pearson correlation coefficient (r) and the corresponding p-values from the two-sided t-test. **b** The number of PSEs associated with PSHMs and PSCTCF in promoter regions in BF. **c** In BF, the proportion of PSEs with PSHMs or PSCTCFs in the promoter region in BF. Pink indicates PSEs with one or more PSHMs or PSCTCFs colocalized; Blue indicates PSEs without any PSHMs or PSCTCFs colocalized. **d-e** An example of a conserved PSE, *PLACB*, regulated by similar POE patterns in both BF (d) and LD (e). **f-g** An example of a tissue-specific PSE is *MUSK*, which lacks the promoter in BF (f) and regulated by

PSHM in LD (g). **h** The fragment results of CRlSPR/Cas9-mediated knockout：Deleted: PCR for detection of enhancer deletion; WT: PCR for detection of the control (wild-type) region. **i** The paternal highly expressed genes *WSCD1* are regulated by the PSHM (the higher H3K27me3 signals in the maternal phase) in the loop-mediated regulatory region. **j** The promoter of *KBTBD6* was connected to a distal region through a paternal-specific loop, consistent with its paternal-specific expression pattern.



Supplementary Fig.7 | Experimental validation of the regulatory mechanism of *SGCE-PEG10* locus a Plasmid construction for fluorescence reporter assay (left). The

fluorescence intensity of reporter genes was used to examine the activity of regulatory element. Bar plots, from left to right, represent the following groups: no vector, control plasmid, cells transfected with a recombinant plasmid containing the promoter sequence, and cells transfected with a recombinant plasmid containing both the enhancer and promoter sequences (right). b The 3C-qPCR assay for *SGCE-PEG10* locus was designed to include genomic regions predicted by Hi-C to interact, restriction enzyme sites, histone modification peaks identified by Cut&Tag, and the positions of dual-luciferase reporter insertion sequences (top). The PCR detection model is illustrated in the lower-left panel. Through ligation reactions, crosslinking, and subsequent reverse crosslinking steps, the two distal genomic regions were brought into proximity, reducing the intervening sequence to less than 300 bp, thereby enabling detection by qPCR (lower-left). The PCR results are presented as a bar chart in the lower-right panel. gDNA input and IgG-enriched DNA served as positive and negative controls, respectively, using primers targeting the *GAPDH* promoter region. To further validate the experimental results, the proximal flanking sequence of the detected region was amplified as an internal control. The detection of ligation products in the +ligase group confirmed the successful ligation of distal regions following reverse crosslinking (lower-right). **c** CRISPR/Cas9-mediated deletion strategy and sgRNA design targeting the enhancer of *SGCE–PEG10* locus. **d** The fragment results of CRlSPR/Cas9-mediated knockout：Deleted: PCR for detection of enhancer deletion; WT: PCR for detection of the control (wild-type) region. **e** Relative expression of *SGCE* (left) and *PEG10* (right) following disruption of enhancer separately by sgRNAs. In all bar

graphs related to functional validation, each data point represents technical replicates, with a consistent sample size (n=3). Two-tailed t-tests were performed to calculate the corresponding P-values.