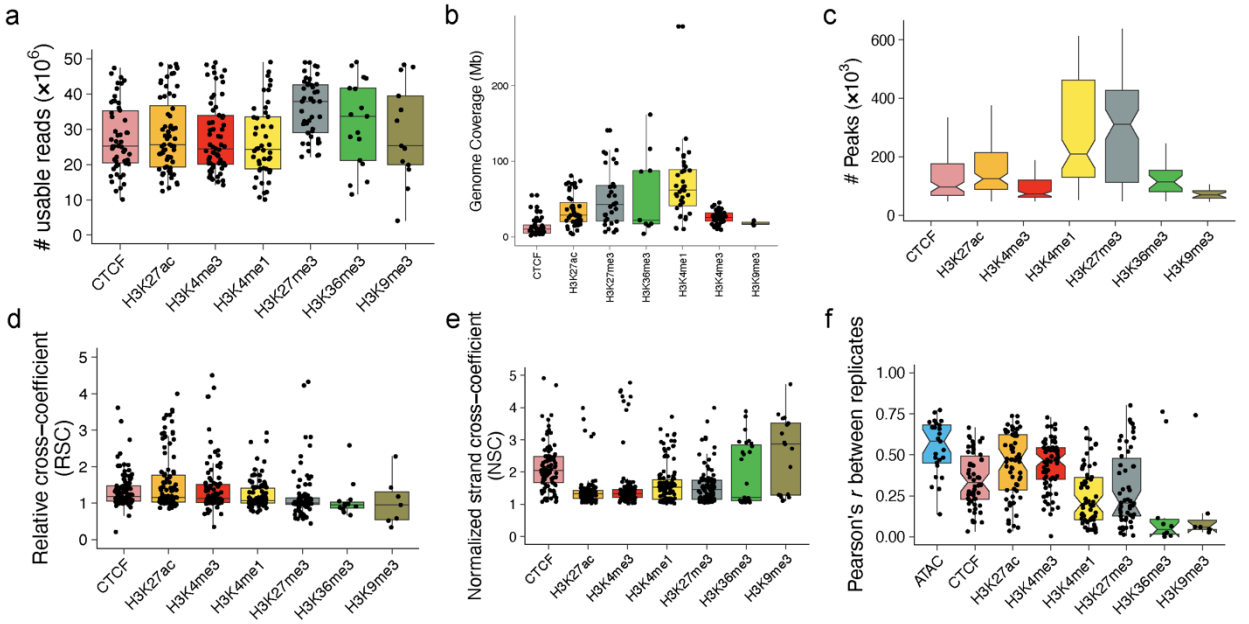
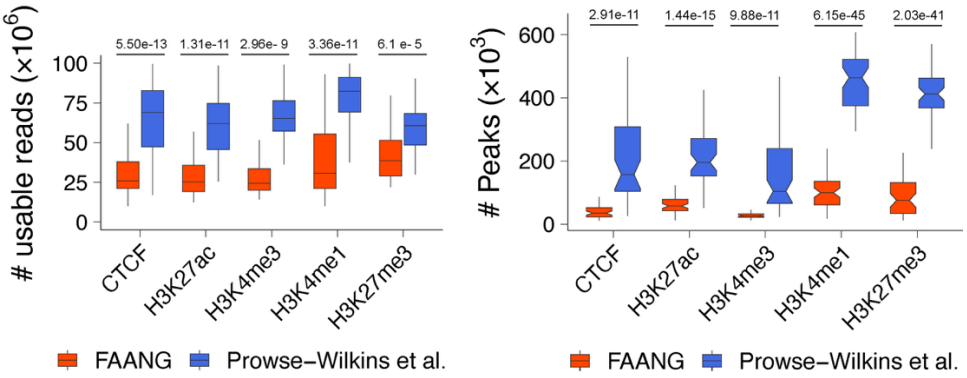
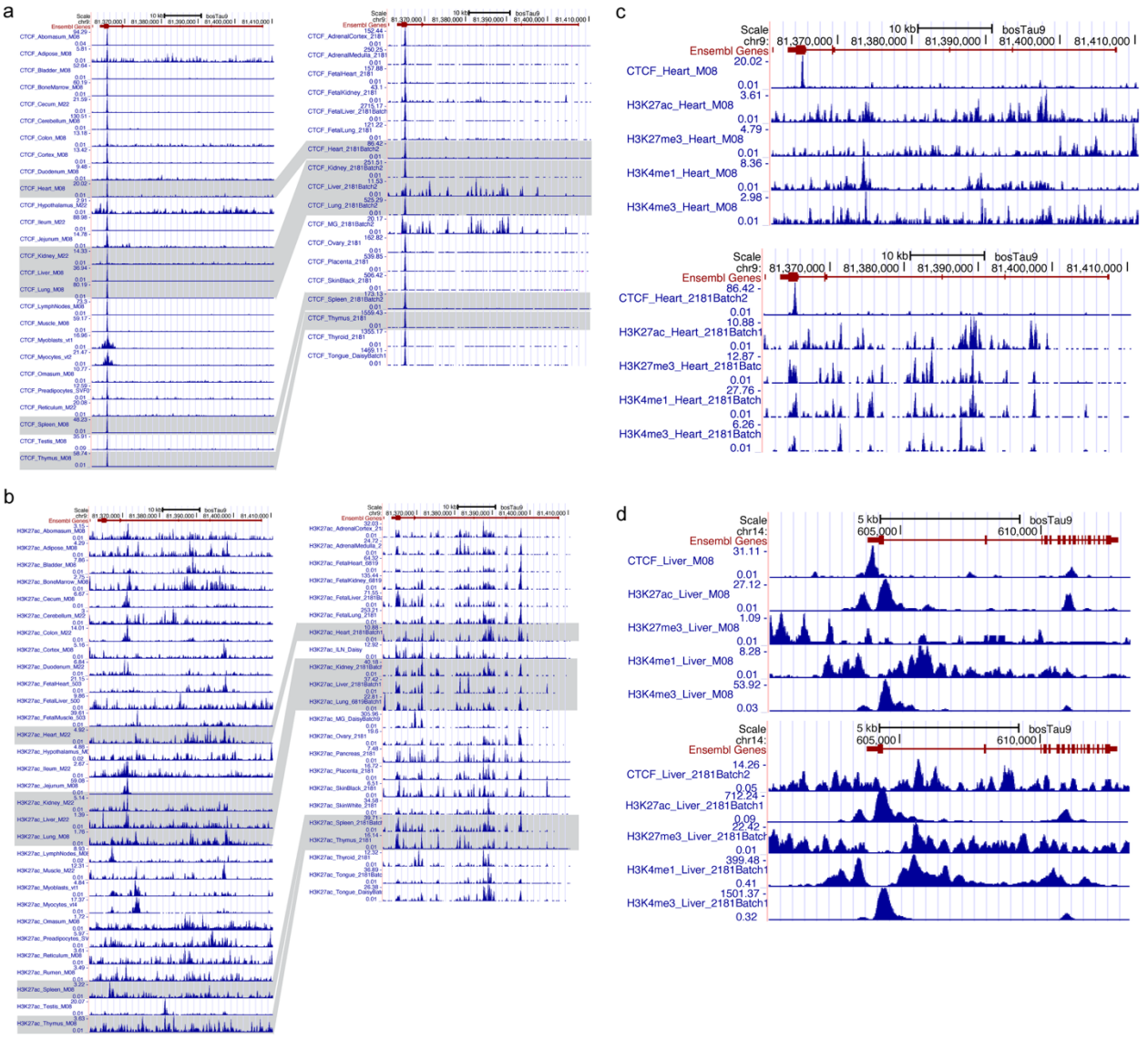


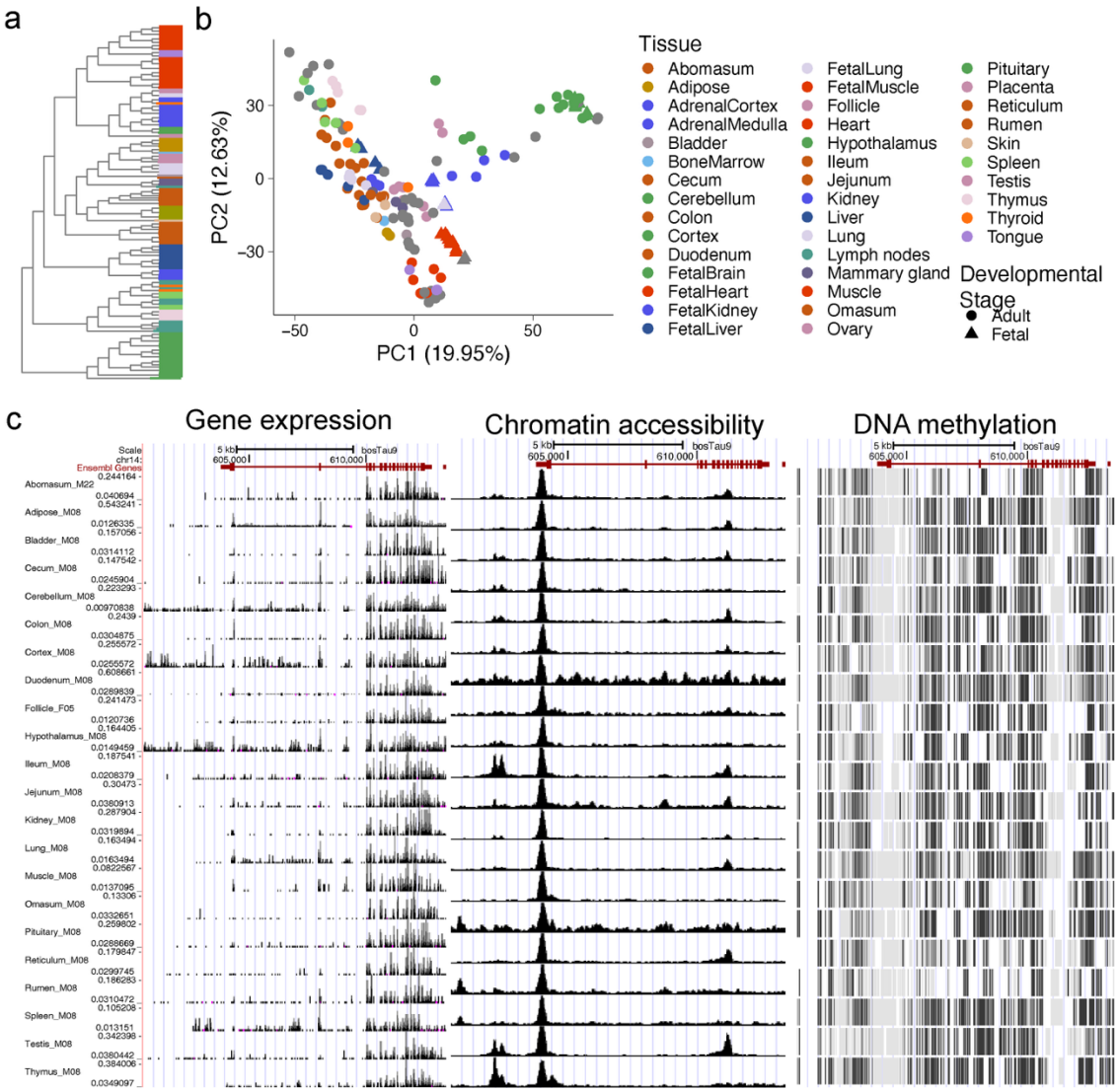
Supplementary Figure 1. Quality control metrics for epigenomic datasets. (a) Number of usable reads (y-axis) across different ChIP-seq histone and CTCF marks. (b) Genome coverage (in megabases, Mb) of ChIP-seq reads across marks. (c) Number of peaks called per ChIP-seq sample. (d–e) Quality metrics for ChIP-seq datasets: Relative Strand Cross-Correlation (RSC, d) and Normalized Strand Cross-Correlation Coefficient (NSC, e) across marks. (f) Pearson's correlation coefficients between biological replicates. In these boxplots, the central line represents the median; box limits indicate the 25th and 75th percentiles; whiskers extend to 1.5× the interquartile range; individual black dots represent samples.



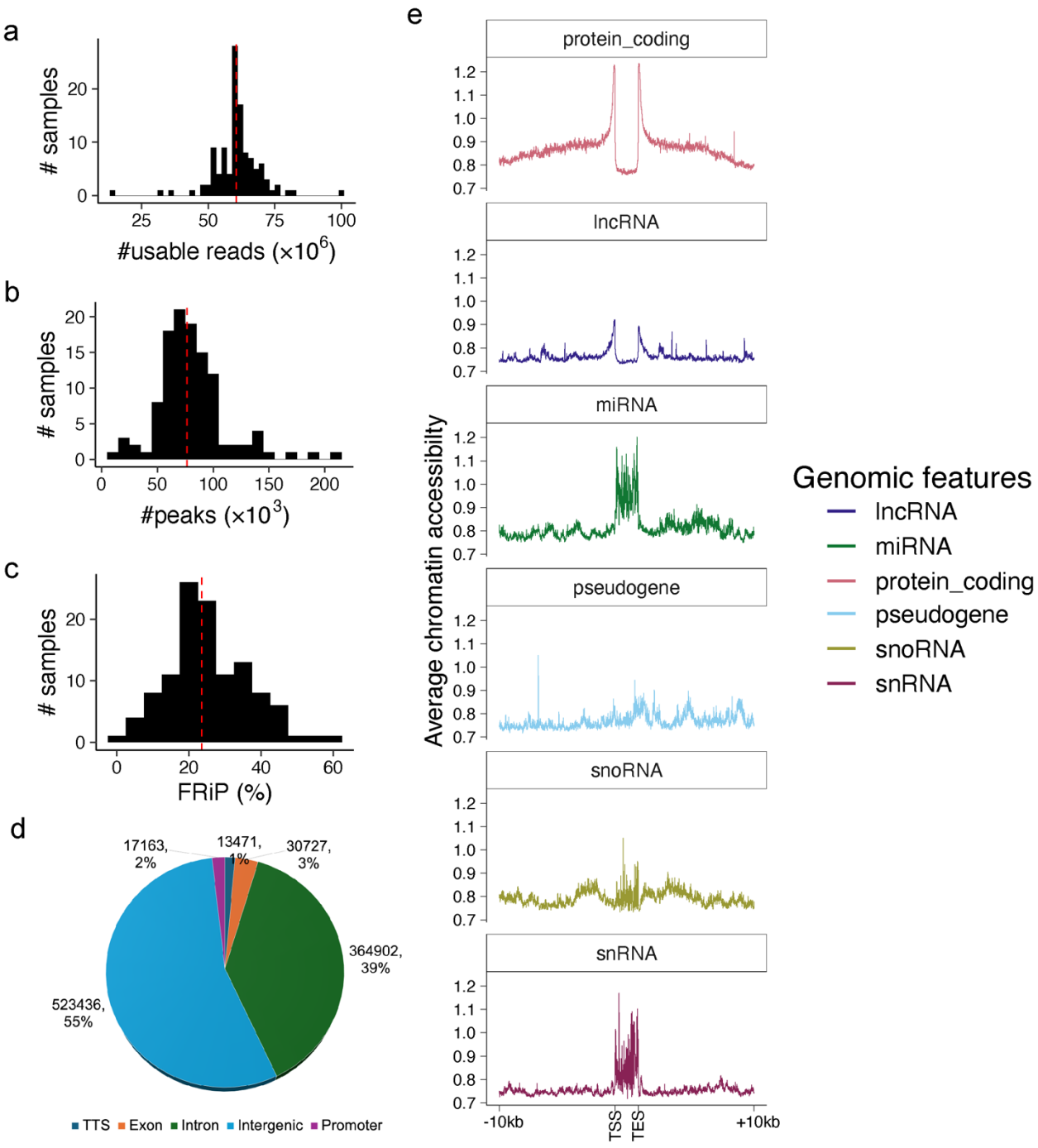
Supplementary Figure 2. Comparison of data quality between bovine epigenomic datasets. Number of usable reads (left) and number of called peaks (right) for datasets generated by the bovine Functional Annotation of Animal Genomes (FAANG) consortium (this study) and by Prowse-Wilkins et al.[1]



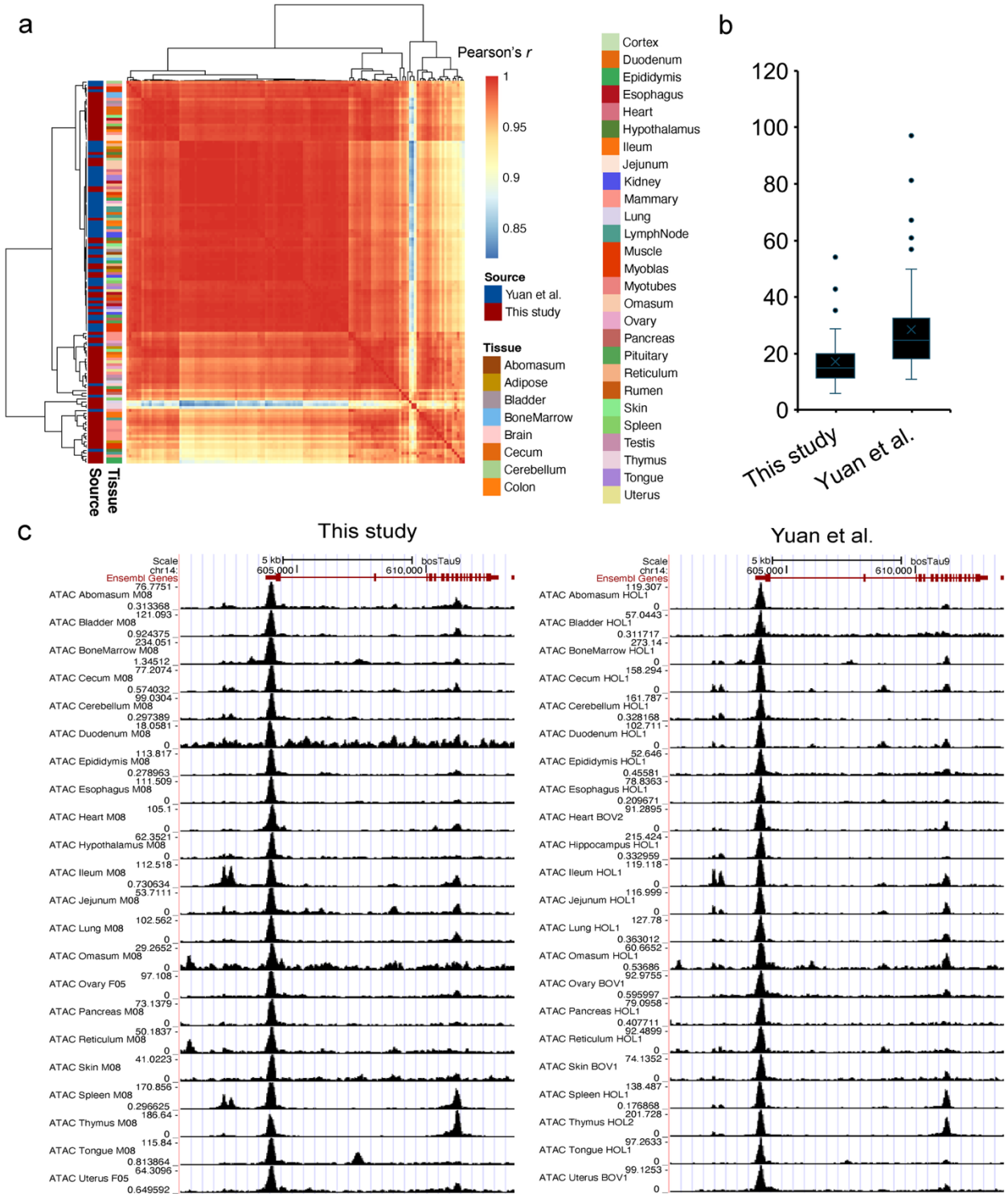
Supplementary Figure 3. Genome browser snapshots of ChIP-seq datasets. (a) Genome browser snapshots of CTCF ChIP-seq datasets generated in this study (left) and by Prowse-Wilkins et al.[1] (right). Gray shading highlights matched tissue samples across datasets. (b) Genome browser snapshots of H3K27ac ChIP-seq datasets generated in this study (left) and by Prowse-Wilkins et al.[1] (right), with matched tissues shaded in gray. (c) Genome browser snapshots of CTCF and histone ChIP-seq marks in muscle tissue, comparing datasets generated in this study (top) with those from Prowse-Wilkins et al.[1] (bottom). The genomic region shown in (a-c) corresponds to the *PLAGL1* locus. (d) Genome browser snapshots of CTCF and histone ChIP-seq marks in liver tissue, from this study (top) and from Prowse-Wilkins et al.[1] (bottom) at the *DGAT1* locus.



Supplementary Figure 4. RNA-seq data analysis. (a) Hierarchical clustering of RNA-seq samples. Colors indicate tissue types and are consistent with panel (b). (b) Principal component (PC) analysis of RNA-seq samples. (c) UCSC Genome Browser snapshot of the *DGAT1* locus, showing RNA-seq read coverage (left), corresponding chromatin accessibility tracks (middle), and DNA methylation profiles (right). Sample IDs are shown at the left side.

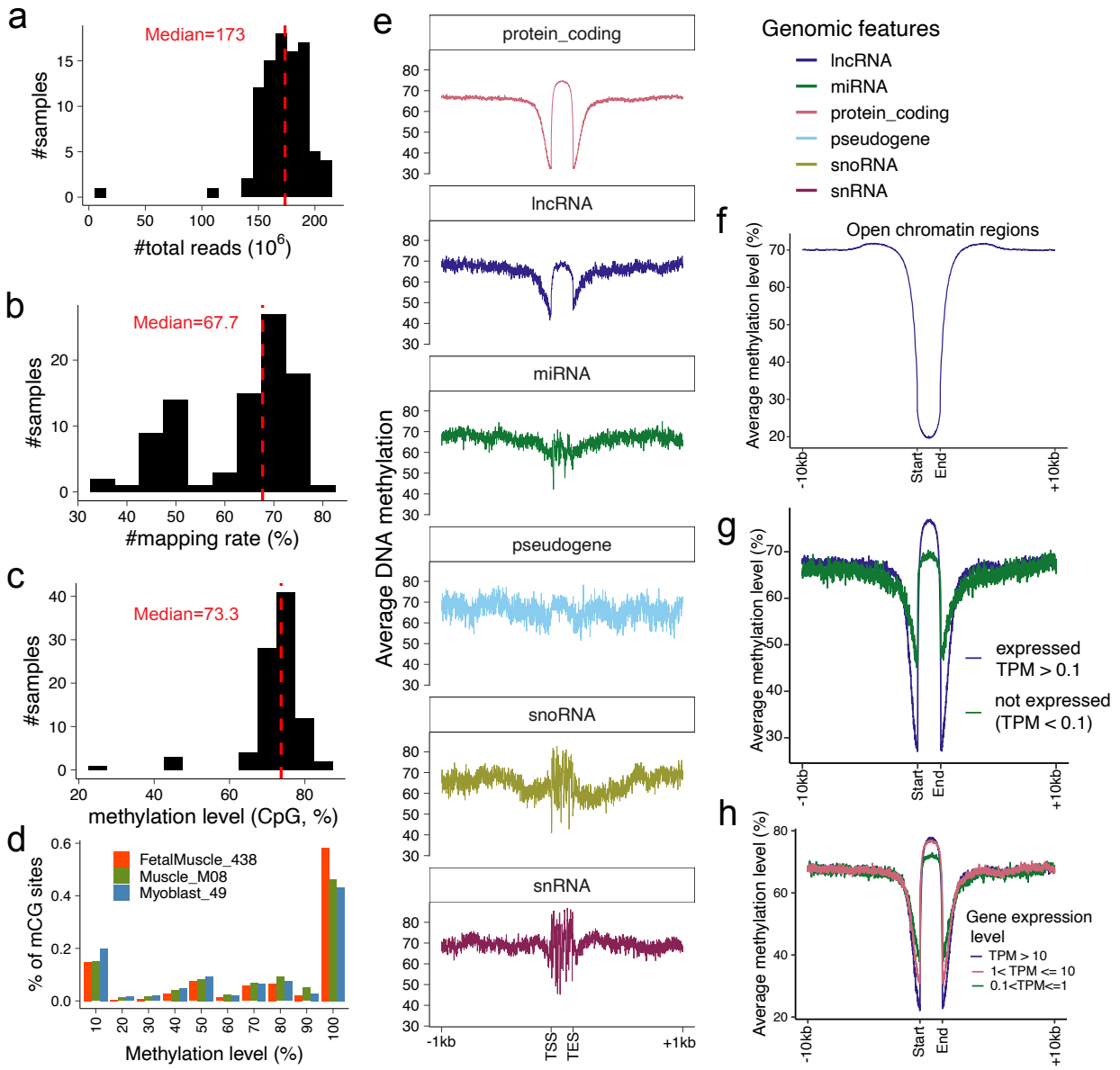


Supplementary Figure 5. ATAC-seq quality control metrics. (a) Histogram showing the distribution of usable reads per sample. The vertical red dashed line marks the median number of usable reads. (b) Number of peaks detected per sample (x-axis). The vertical red dashed line indicates the median peak count. (c) Fraction of reads in peaks (FRiP) scores across samples. FRiP is defined as the proportion of usable reads that fall within significantly enriched peak regions. (d) Pie chart illustrating the distribution of ATAC-seq peaks overlapping various genomic features. (e) Average chromatin accessibility (y-axis) across regions surrounding transcription start sites (TSS) and transcription end sites (TES) of annotated genomic features (Ensembl v105).

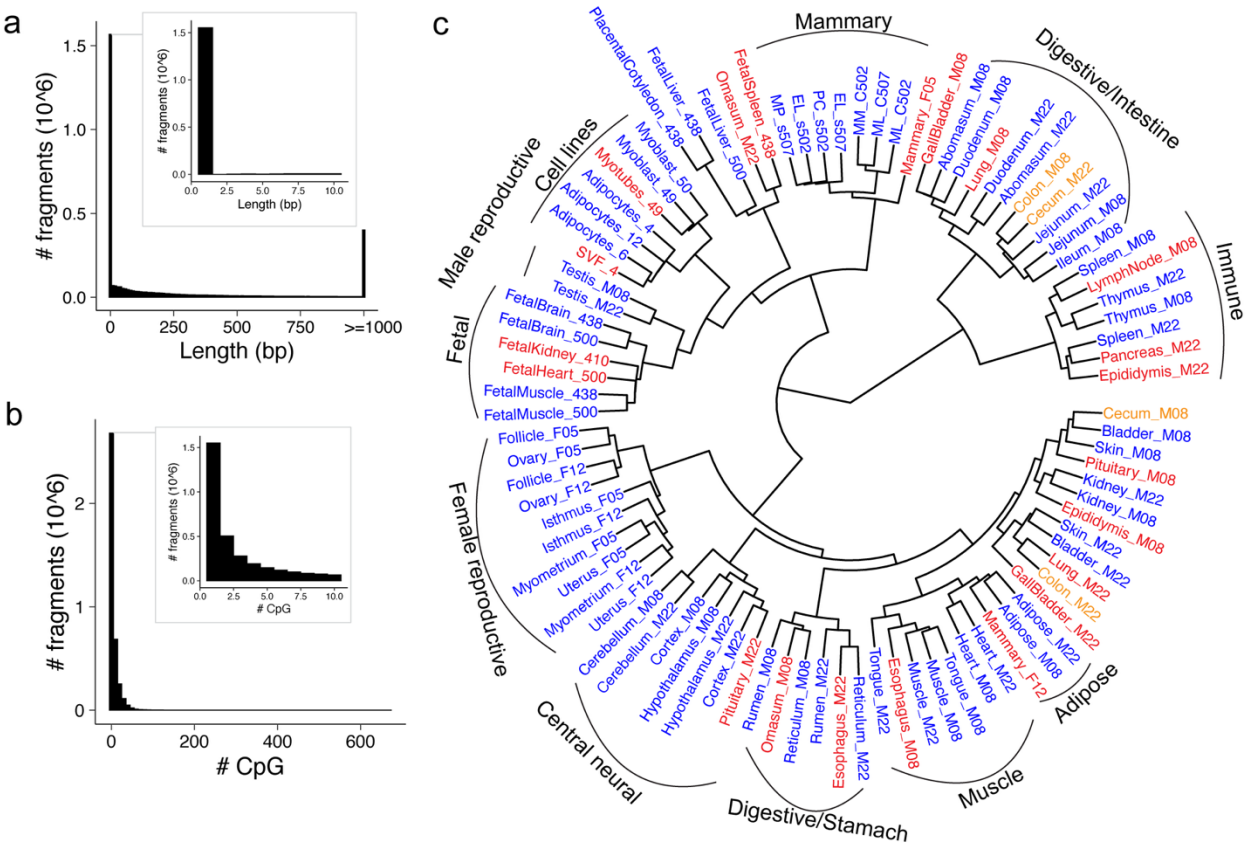


Supplementary Figure 6. ATAC-seq data analysis. (a) Heatmap showing Pearson's correlation among ATAC-seq samples. Samples were clustered based on dissimilarity (1 – d), where d is the Euclidean distance computed from normalized ATAC-seq read counts, using complete linkage clustering. (b) Enrichment of transcription start sites (TSS, y-axis) across ATAC-seq samples

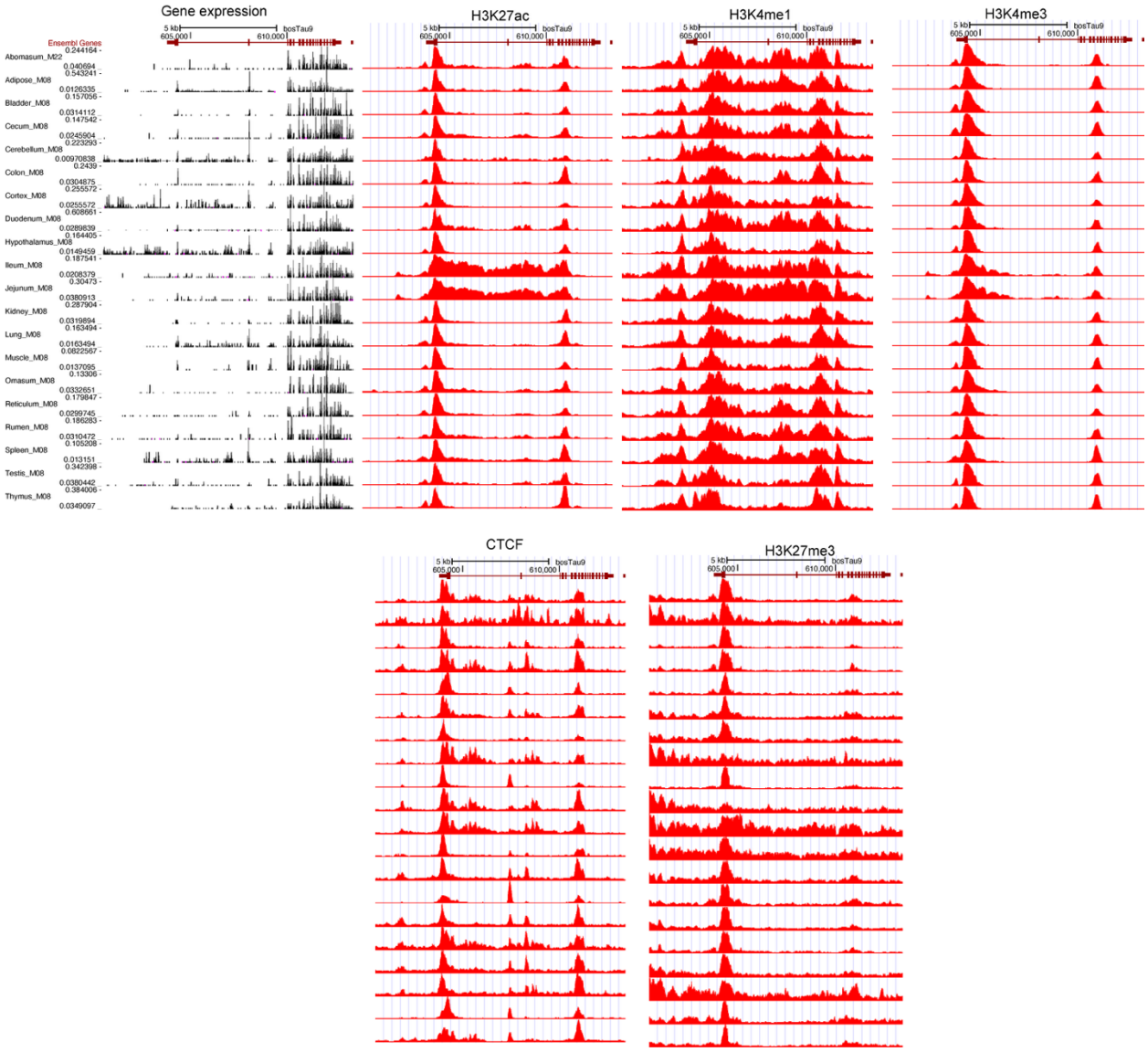
generated in this study and those from Yuan et al.[2]. (c) Genome browser snapshots displaying normalized chromatin accessibility signals at the *DGAT1* locus, comparing data generated in this study (left) and by Yuan et al.[2] (right).



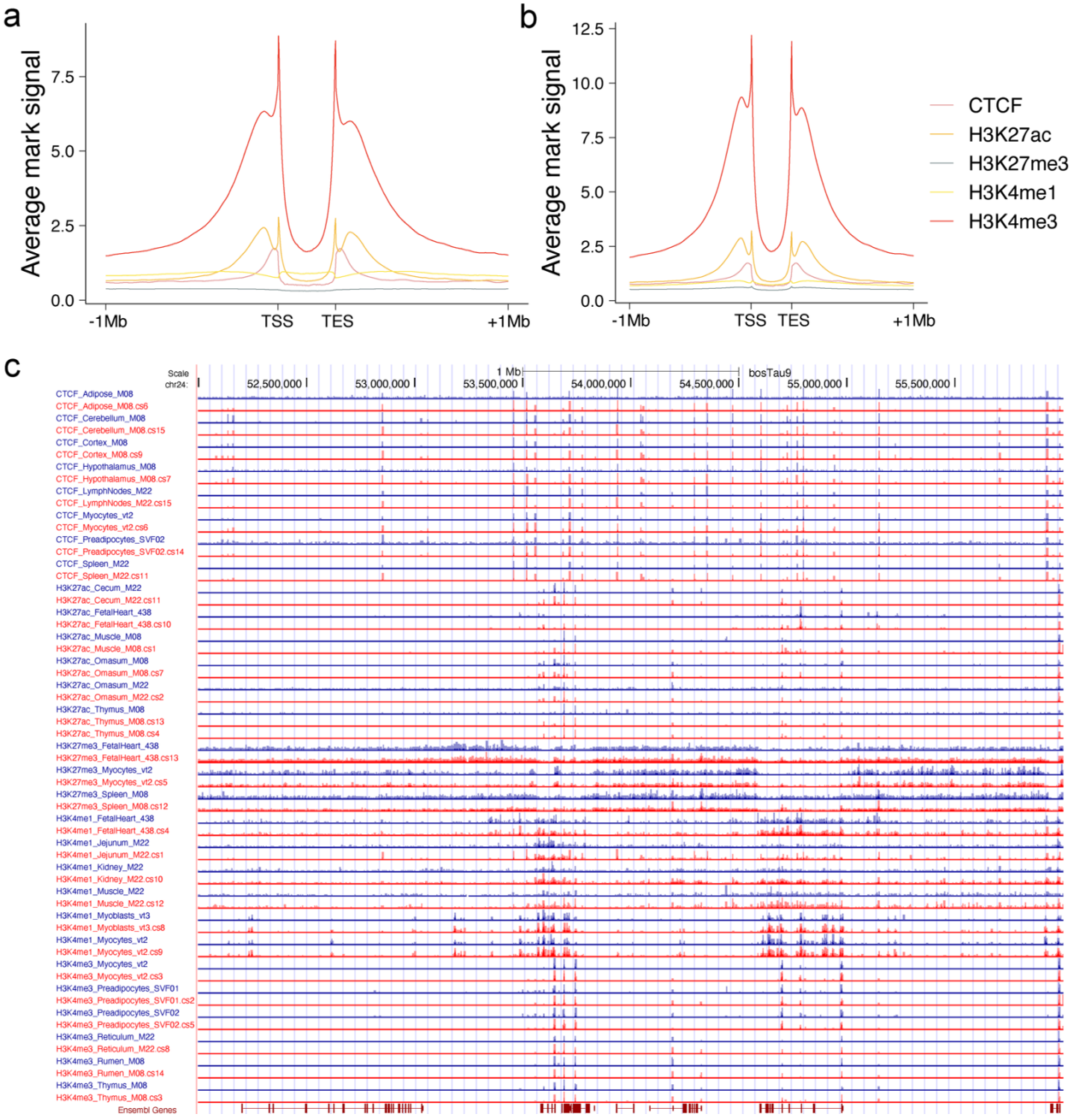
Supplementary Figure 7. Quality control of DNA methylation data. (a) Total number of clean reads per sample (x-axis). The median value (173 million reads) is indicated by a red dashed line. (b) Mapping rates (%) across samples, with the median mapping rate to the autosomes (67.7%) highlighted by a red dashed line. (c) CpG methylation levels (%) across samples, with the median level (73.3%) shown by a red dashed line. (d) Example plot comparing the proportion of methylated CpG sites (y-axis) across myoblast, fetal, and adult bulk muscle samples as a function of methylation levels (x-axis). (e) Average DNA methylation levels surrounding transcription start sites (TSS) and transcription end sites (TES) of annotated genomic features (Ensembl v105).



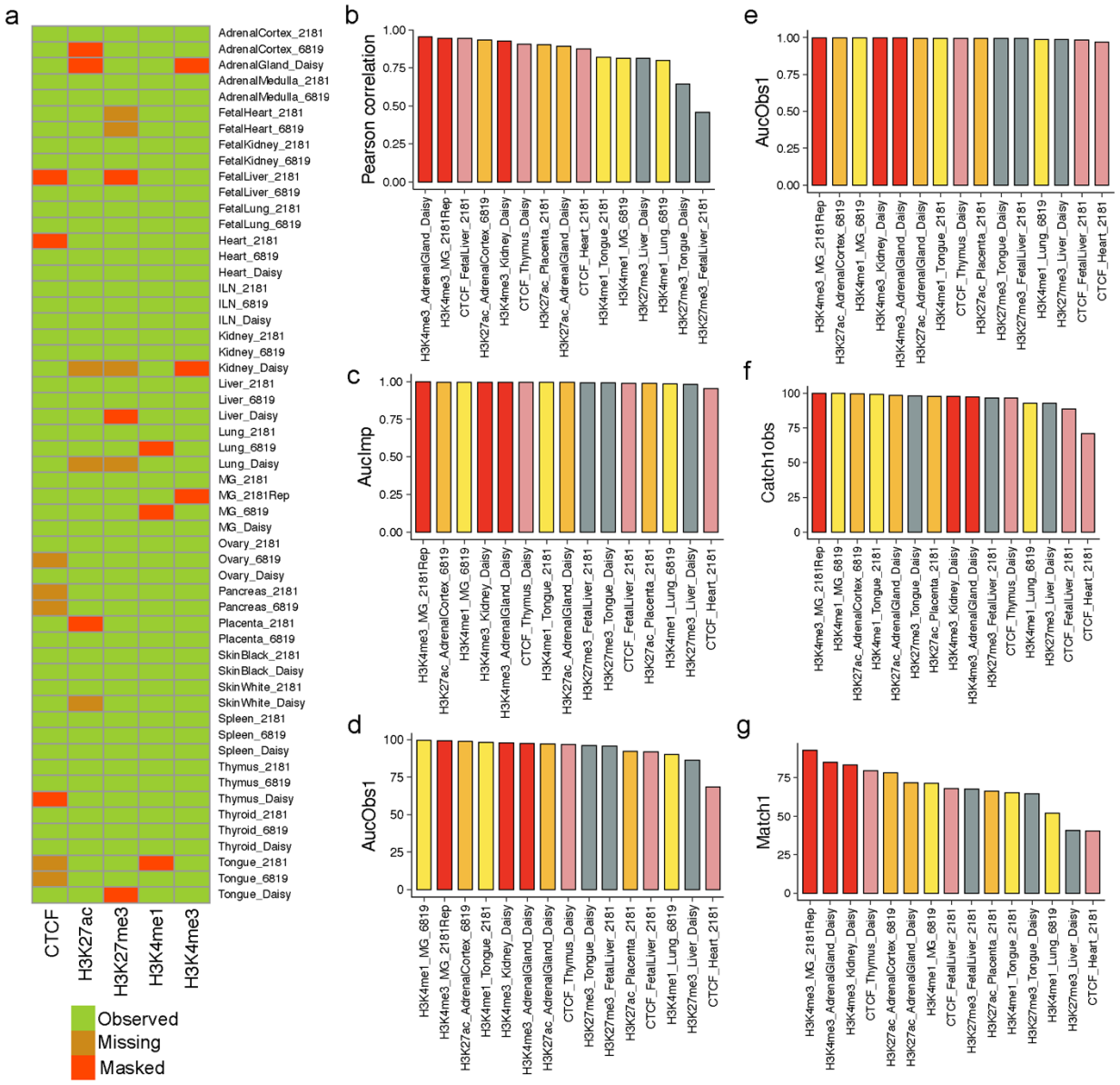
Supplementary Figure 8. DNA methylation data analysis. (a) Length distribution of homogeneously methylated blocks (defined as contiguous CpG sites with similar methylation levels, in base pair, bp). (b) Distribution of the number CpG sites per homogeneously methylated blocks. (c) Hierarchical clustering of WGBS samples. We highlighted some samples that two replicated are not clustered together, potentially due to low quality or cell type heterogeneity. WGBS: whole genome bisulfite sequencing.



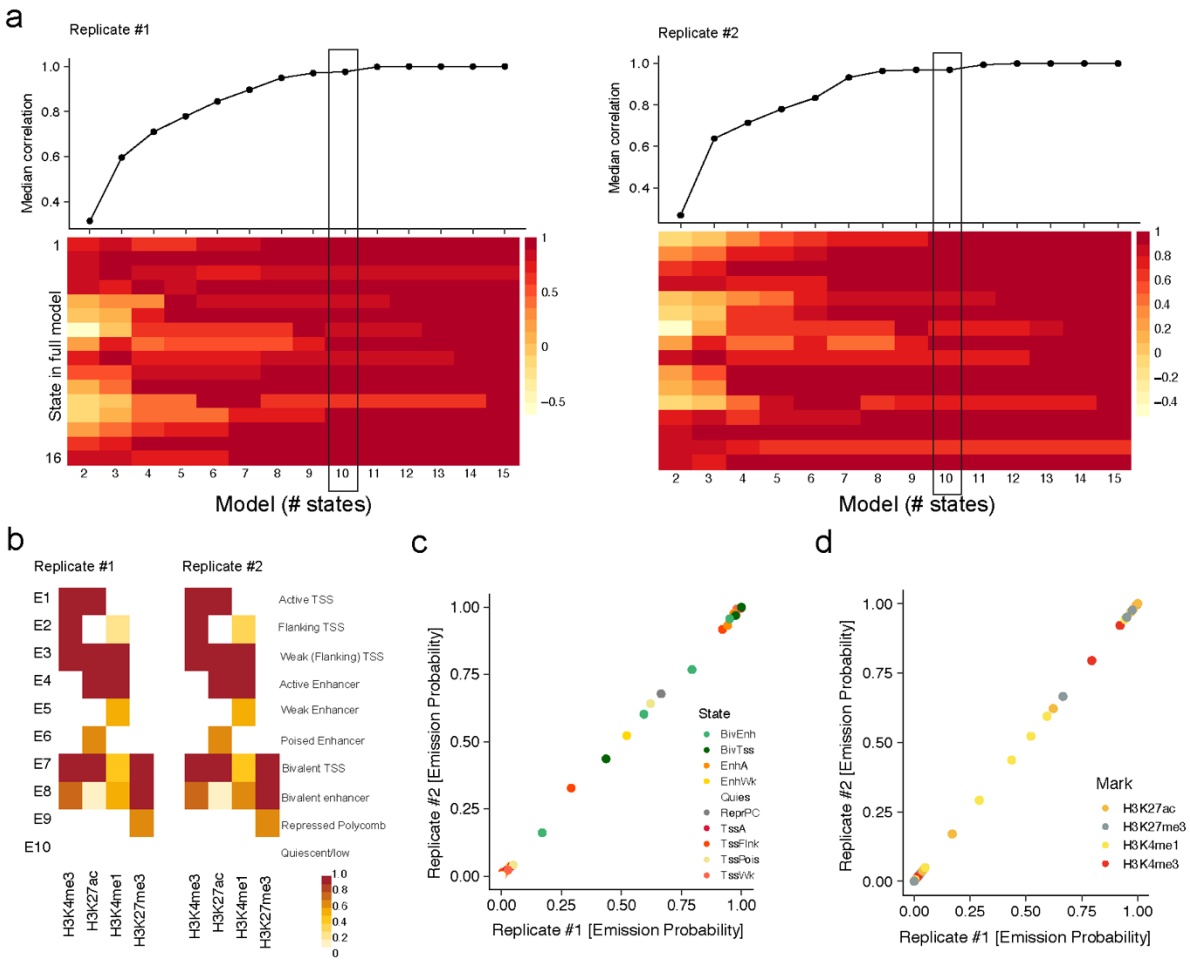
Supplementary Figure 9. UCSC Genome Browser snapshot of the *DGAT1* locus. Tracks shown include RNA-seq gene expression (top left) and ChIP-seq profiles for H3K27ac, H3K4me1, H3K4me3, CTCF, and H3K27me3 across selected samples.



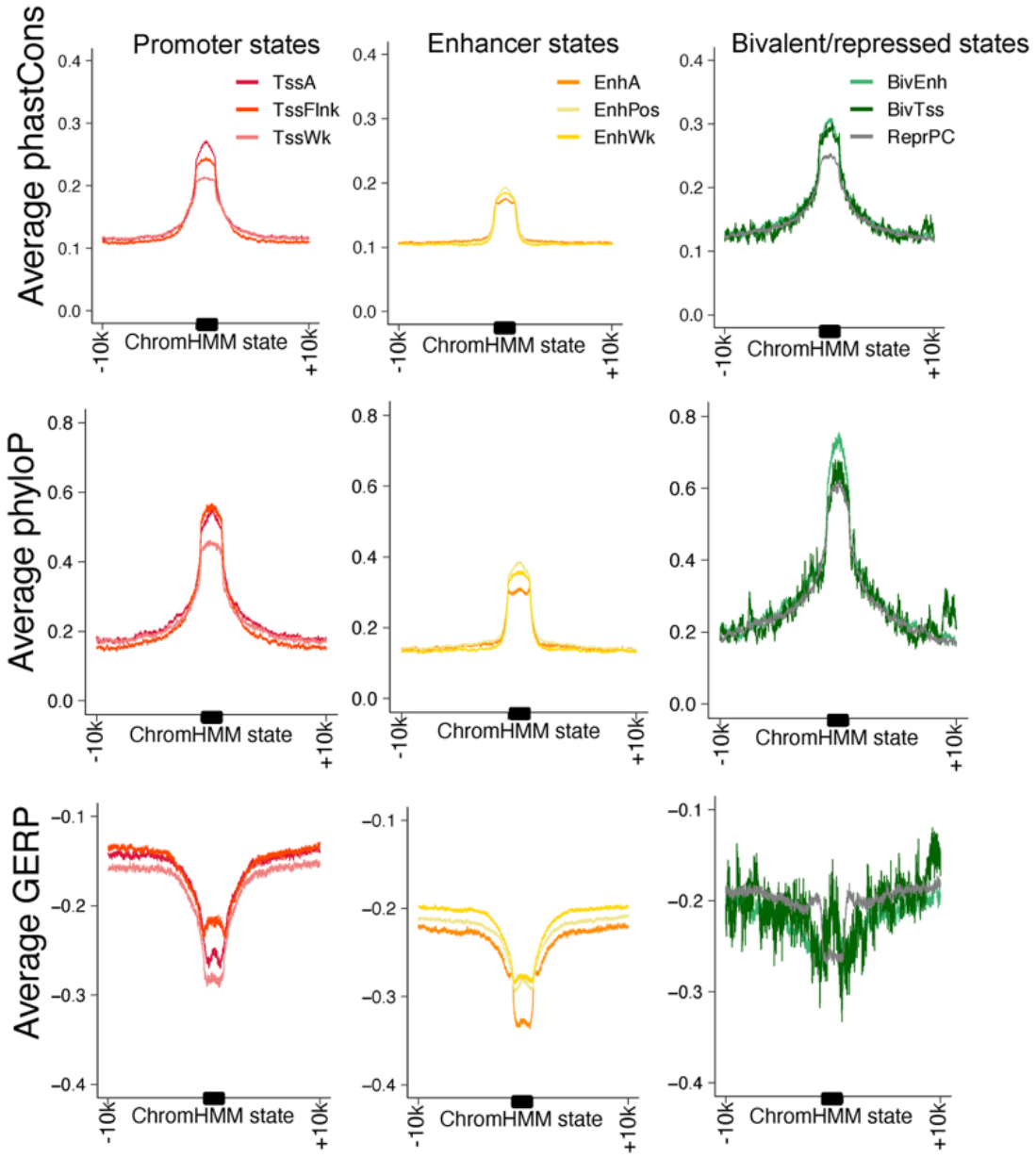
Supplementary Figure 10. Quality control of missing and low-quality epigenomic imputation. (a–b) Average ChIP-seq signal profiles surrounding transcription start sites (TSS) and transcription end sites (TES) for observed (a) and imputed (b) data. (c) UCSC Genome Browser snapshot illustrating observed (blue) and imputed (red) signals across a randomly selected 2-Mb genomic region (chr24:52-56 Mb).



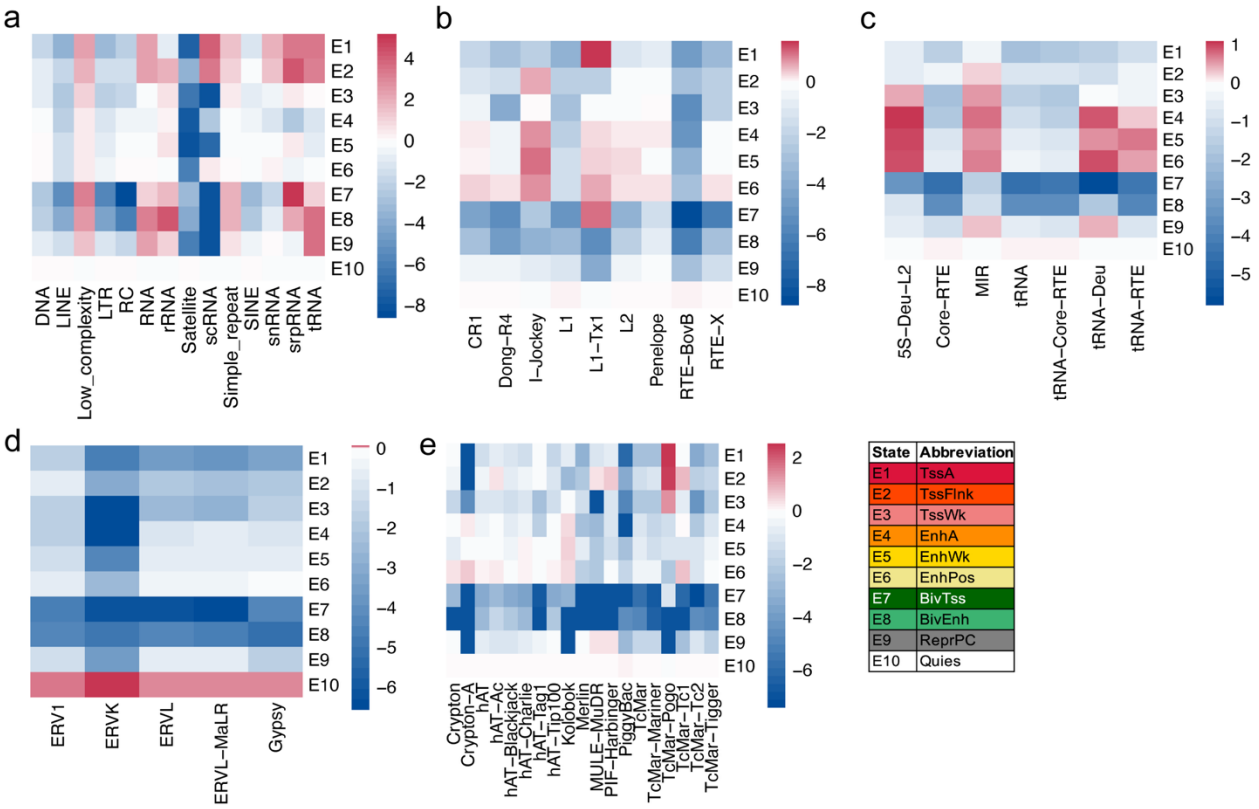
Supplementary Figure 11. Validation of epigenome imputation. (a) Data presentation used in epigenome validation and evaluation, retrieved from Prowse-Wilkins et al.[1] (b-g) Evaluation of imputation performance using various quality metrics, including Pearson correlation between the observed and imputed data (b), the area under the receiver operating characteristic (ROC) curve for predicting the top 1% imputed signal with the full range of observed signal (c), the area under the ROC for predicting the top 1% observed signal with the full range of imputed signal (d), the fraction of the observed top 1% locations in the imputed top 5% locations (e), the fraction of the observed top 1% in the imputed top 5%, and the fraction of the observed top 1% locations in the imputed top 1% locations (g).



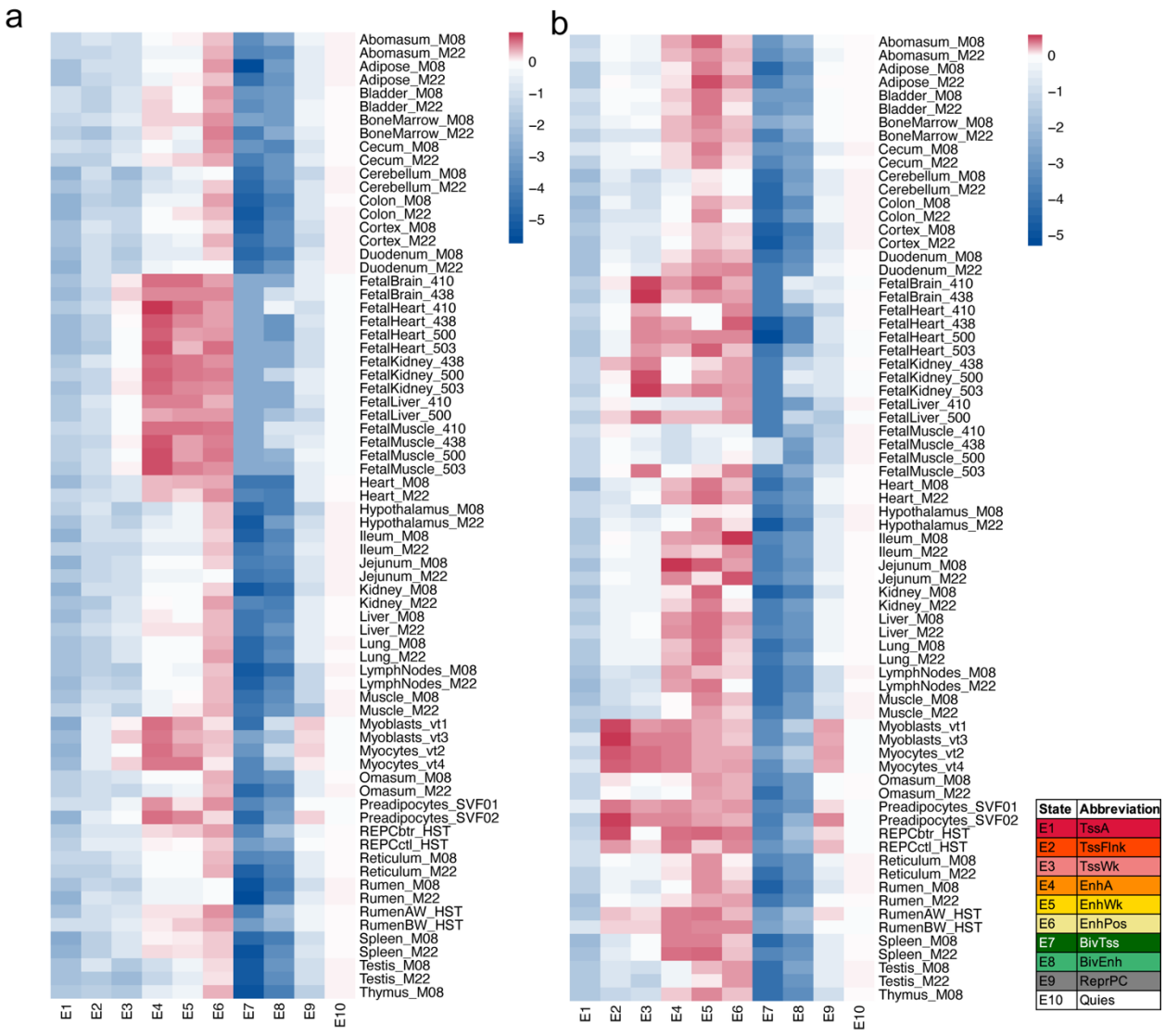
Supplementary Figure 12. Learning chromatin state using ChromHMM. (a–b) Heatmaps showing the maximum Pearson correlation between each chromatin state from the full model (y-axis) and its best-matching state in simplified models (x-axis) for replicate 1 (left) and replicate 2 (right). The median correlation across all 14 states is indicated above each heatmap. (b) Emission probability heatmaps for each chromatin mark across all learned chromatin states, as defined by ChromHMM, for both biological replicates. (c) Spearman correlation of emission probabilities between replicates, color-coded by chromatin state (left) or histone mark (right), demonstrating reproducibility of state definitions and mark-specific patterns.



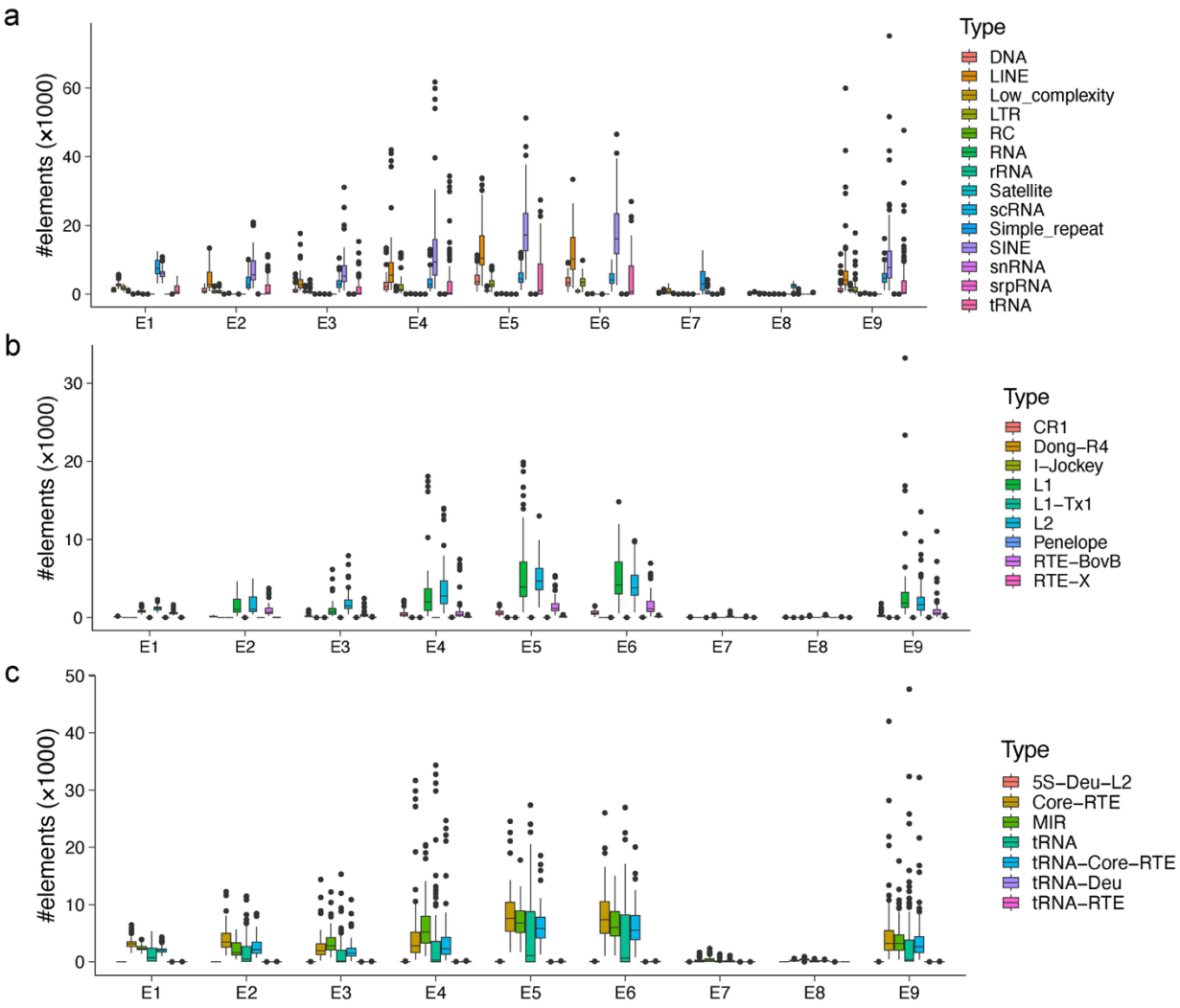
Supplementary Figure 13. Evolutionary constraint around annotated chromatin states. Line plots showing average conservation scores across chromatin states as measured by phastCons (top), phyloP (middle), and GERP (bottom), highlighting the evolutionary constraints associated with different regulatory annotations. GERP: Genomic Evolutionary Rate Profiling.



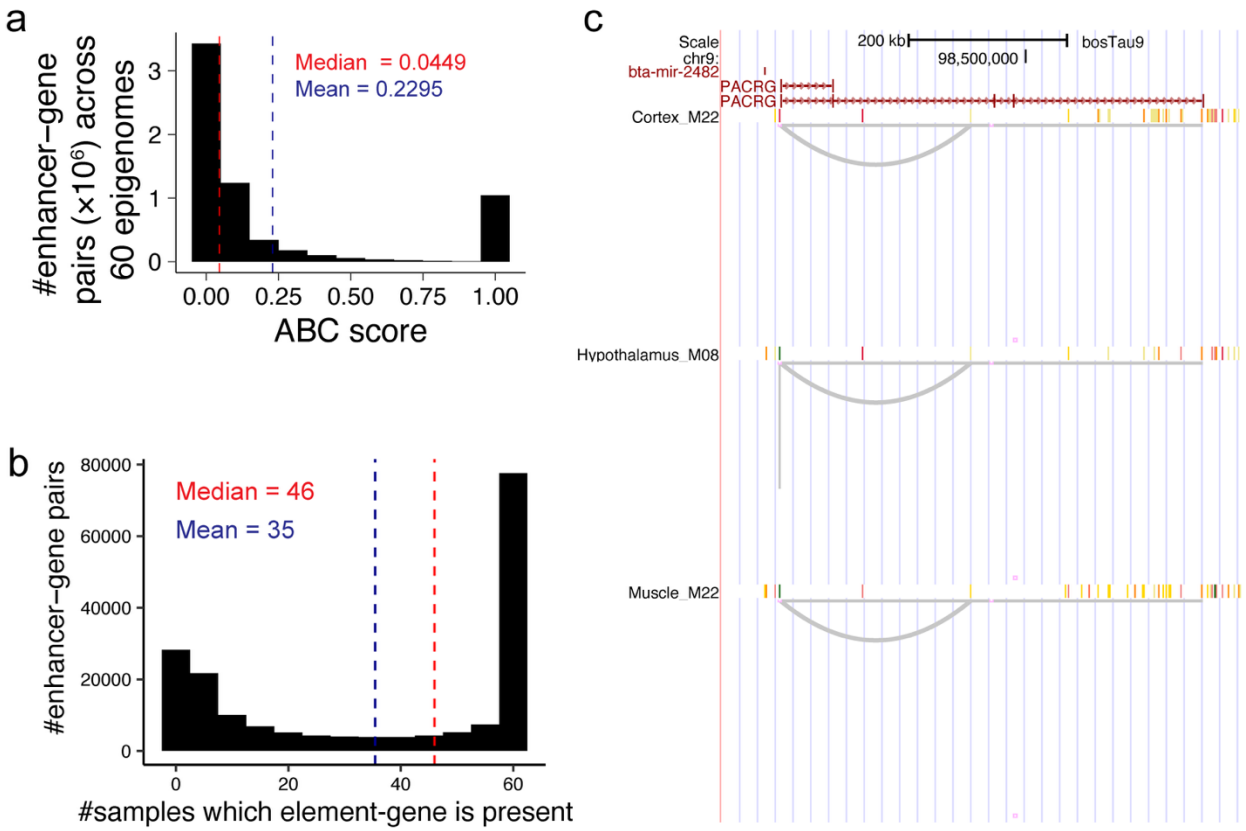
Supplementary Figure 14. Enrichment of chromatin states across transposable elements. (a) Average enrichment of chromatin states across all transposable elements (TEs). (b–e) Enrichment of chromatin states across specific TE classes, including LINE subfamilies (b), SINEs (c), long terminal repeats (LTRs) (d), and DNA elements (e). Color scale indicates $\log_2$-transformed fold enrichment, with negative values indicating depletion and positive values indicating enrichment.



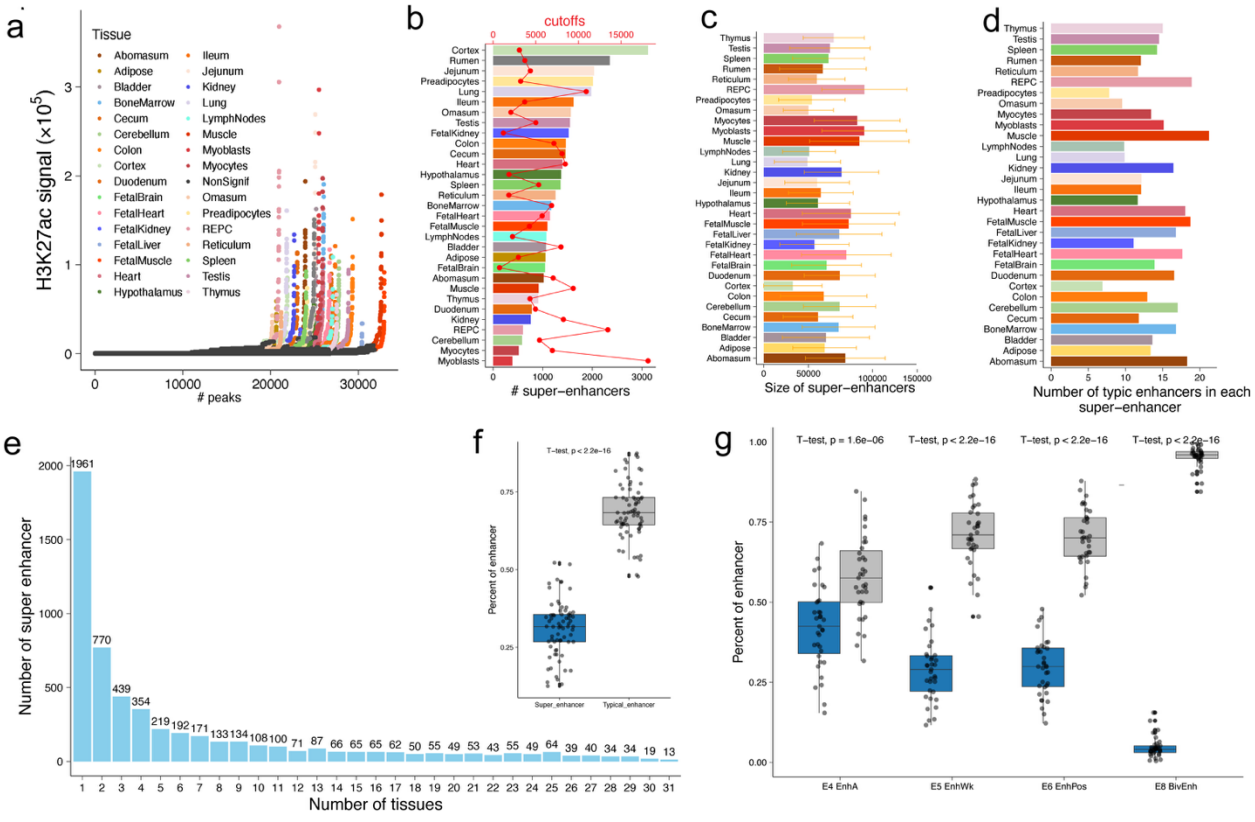
Supplementary Figure 15. Enrichment of chromatin states across transposable element subfamily stratified by individual epigenomes. Enrichment of chromatin states within CR1 (a), and L2 subfamily (b). Enhancer-like states (E4, E5, and E6) show strong enrichment in fetal tissues and primary cell lines. Color scale indicates $\log_2$-transformed fold enrichment, with negative values indicating depletion and positive values indicating enrichment.



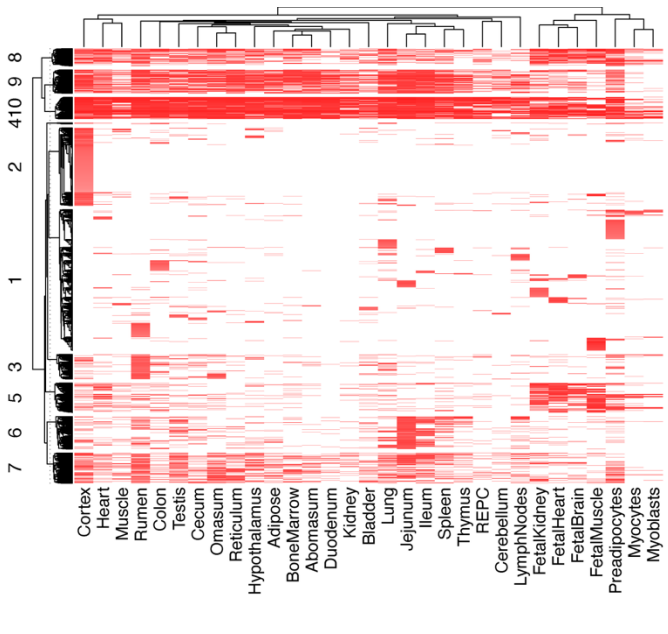
Supplementary Figure 16. Overlap of chromatin states with transposable elements. (a) Number of annotated chromatin states overlapping various classes of transposable elements (TE). (b) Overlap between chromatin states and LINE subfamilies. (c) Overlap between chromatin states and SINE subfamilies.



Supplementary Figure 17. Activity-by-Contact (ABC) linking enhancers to targets. (a) Histogram showing the distribution of ABC scores across 60 epigenomes with available ATAC-seq and H3K27ac ChIP-seq data. (b) Histogram displaying the number of epigenomes (x-axis) in which unique enhancer–gene pairs are identified. (c) Genome browser snapshot illustrating an example of enhancer–promoter linkage, with an enhancer located in the intron of PACRG linked to its promoter.

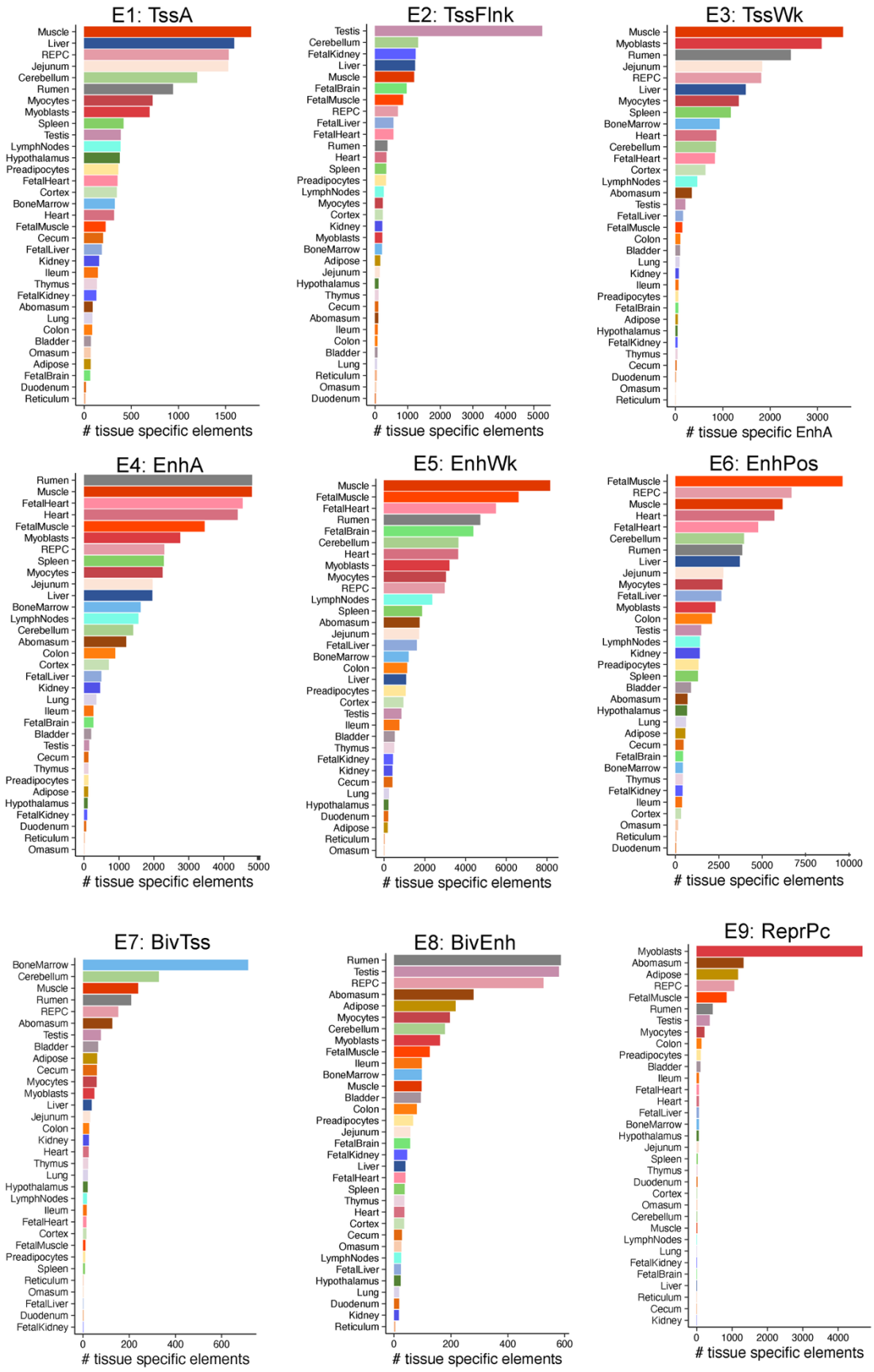


Supplementary Figure 18. Identification and characterization of super-enhancers across tissues. (a) Enhancer ranking (x-axis) based on H3K27ac ChIP-seq signal intensity (y-axis), with super-enhancers shown as colored dots (representing different tissues) and typical enhancers in grey. (b) Number of super-enhancers identified (bottom x-axis) and corresponding signal thresholds (top x-axis) across tissues (y-axis). (c) Distribution of super-enhancer sizes (x-axis, in base pairs [bp]) across tissues. Bars represent standard deviation within each tissue. (d) Average number of constituent typical enhancers within each super-enhancer across tissues. (e) Tissue specificity of super-enhancers, with the x-axis representing the number of tissues in which a super-enhancer is observed and the y-axis indicating the count. 1,961 super-enhancers were found to be tissue-specific. (f) Proportion of enhancers classified as super-enhancers versus typical enhancers across all tissues (represented by a dot). (g) Proportion of enhancers assigned to super-enhancers versus typical enhancers stratified by four enhancer-like chromatin states (x-axis) across all tissues (represented by a dot).

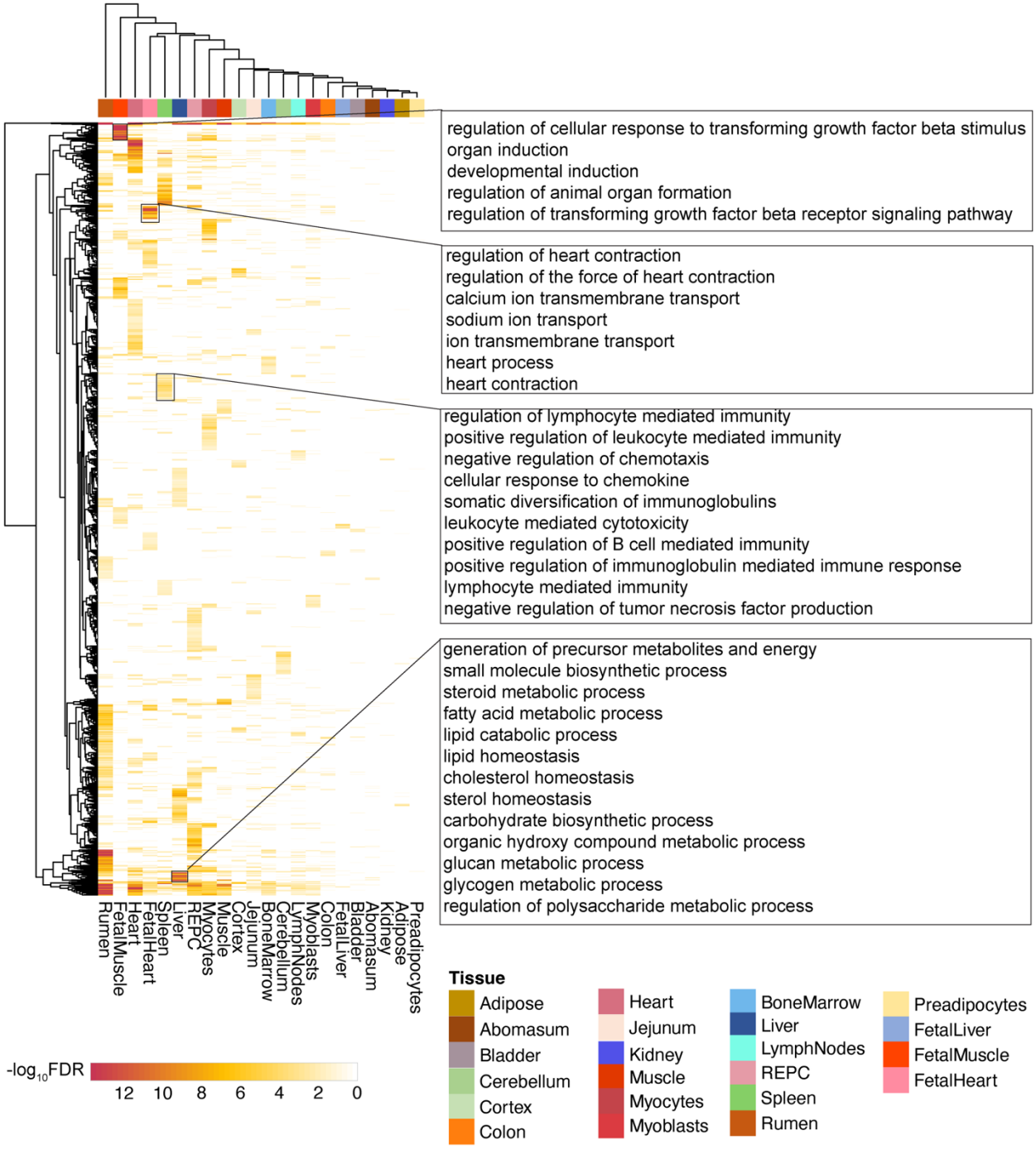


Cluster	# SE	GO Term (-log10P)
C8	220	transcription by RNA polymerase III (3.43)
C9	325	regulation of biosynthetic process (6.11)
C10	302	organic substance metabolic process (4.25)
C4	315	negative regulation of transforming growth factor beta receptor signaling pathway (6.53)
C2	1144	response to external stimulus (3.34)
C1	1495	macromolecule metabolic process (5.49)
C3	552	cellular nitrogen compound metabolic process (5.49)
C5	355	muscle organ development (2.96)
C6	461	lipid homeostasis (4.14)
C7	425	biological_process (6.01)

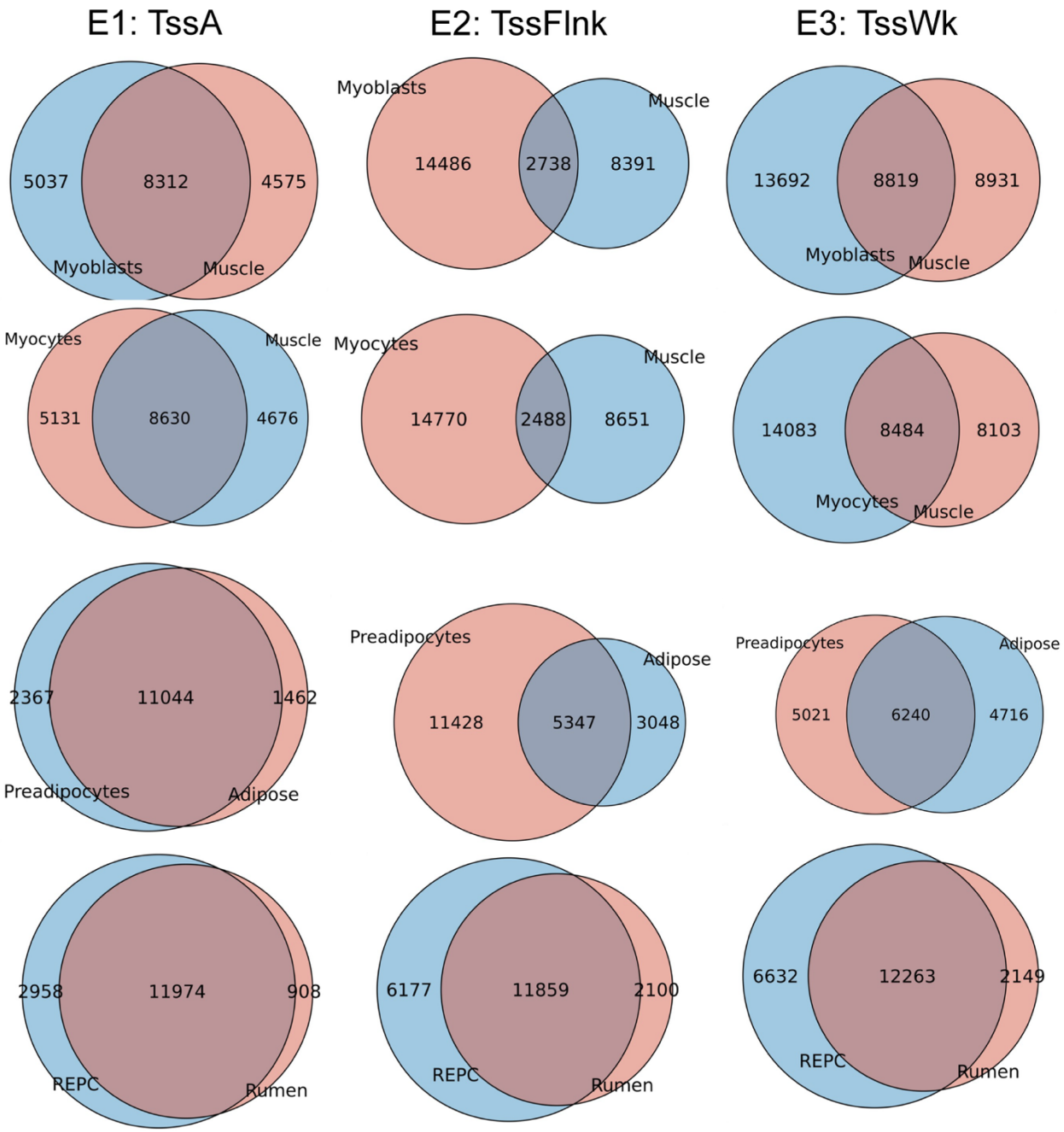
Supplementary Figure 19. Clustering of 5,594 nonredundant super-enhancers based on tissue activity. Nonredundant super-enhancers were clustered into 10 groups according to their activity profiles, as measured by H3K27ac read counts across tissues. The heatmap displays the activity patterns of super-enhancers within each cluster, with representative functional enrichment results shown to the right of the corresponding cluster. Tissue relevant functions were identified, such as muscular tissues (including fetal muscle, myocytes, myoblasts) were enriched for muscle organ development.


E1: TssA
Muscle
Liver
REPC
Jejunum
Cerebellum
Rumen
Myocytes
Myoblasts
Spleen
Testis
LymphNodes
Hypothalamus
Preadipocytes
FetalHeart
Cortex
BoneMarrow
Heart
FetalMuscle
Cecum
FetalLiver
Kidney
Ileum
Thymus
FetalKidney
Abomasum
Lung
Colon
Bladder
Omasum
Adipose
FetalBrain
Duodenum
Reticulum
0 500 1000 1500
tissue specific elements
E2: TssFlnk
Testis
Cerebellum
FetalKidney
Liver
Muscle
FetalBrain
FetalMuscle
REPC
FetalLiver
FetalHeart
Rumen
Heart
Spleen
Preadipocytes
LymphNodes
Myocytes
Cortex
Kidney
Myoblasts
BoneMarrow
Adipose
Jejunum
Hypothalamus
Thymus
Cecum
Abomasum
Ileum
Colon
Bladder
Lung
Reticulum
Omasum
Duodenum
0 1000 2000 3000 4000 5000
tissue specific elements
E3: TssWk
Muscle
Myoblasts
Rumen
Jejunum
REPC
Liver
Myocytes
Spleen
BoneMarrow
Heart
Cerebellum
FetalHeart
Cortex
LymphNodes
Abomasum
Testis
FetalLiver
FetalMuscle
Colon
Bladder
Lung
Kidney
Ileum
Preadipocytes
FetalBrain
Adipose
Hypothalamus
FetalKidney
Thymus
Cecum
Duodenum
Omasum
Reticulum
0 1000 2000 3000
tissue specific EnhA
E4: EnhA
Rumen
Muscle
FetalHeart
Heart
FetalMuscle
Myoblasts
REPC
Spleen
Myocytes
Jejunum
Liver
BoneMarrow
LymphNodes
Cerebellum
Abomasum
Colon
Cortex
FetalLiver
Kidney
Lung
Ileum
FetalBrain
Bladder
Testis
Cecum
Thymus
Preadipocytes
Adipose
Hypothalamus
FetalKidney
Duodenum
Reticulum
Omasum
0 1000 2000 3000 4000 5000
tissue specific elements
E5: EnhWk
Muscle
FetalMuscle
FetalHeart
Rumen
FetalBrain
Cerebellum
Heart
Myoblasts
Myocytes
REPC
LymphNodes
Spleen
Abomasum
Jejunum
FetalLiver
BoneMarrow
Colon
Liver
Preadipocytes
Cortex
Testis
Ileum
Bladder
Thymus
FetalKidney
Kidney
Cecum
Lung
Hypothalamus
Duodenum
Adipose
Reticulum
Omasum
0 2000 4000 6000 8000
tissue specific elements
E6: EnhPos
FetalMuscle
REPC
Muscle
Heart
FetalHeart
Cerebellum
Rumen
Liver
Jejunum
Myocytes
FetalLiver
Myoblasts
Colon
Testis
LymphNodes
Kidney
Preadipocytes
Spleen
Bladder
Abomasum
Hypothalamus
Lung
Adipose
Cecum
FetalBrain
BoneMarrow
Thymus
FetalKidney
Ileum
Cortex
Omasum
Reticulum
Duodenum
0 2500 5000 7500 10000
tissue specific elements
E7: BivTss
BoneMarrow
Cerebellum
Muscle
Rumen
REPC
Abomasum
Testis
Bladder
Adipose
Cecum
Myocytes
Myoblasts
Liver
Jejunum
Colon
Kidney
Heart
Thymus
Lung
Hypothalamus
LymphNodes
Ileum
FetalHeart
Cortex
FetalMuscle
Preadipocytes
Spleen
Reticulum
Omasum
FetalLiver
Duodenum
FetalKidney
0 200 400 600
tissue specific elements
E8: BivEnh
Rumen
Testis
REPC
Abomasum
Adipose
Myocytes
Cerebellum
Myoblasts
FetalMuscle
Ileum
BoneMarrow
Muscle
Bladder
Colon
Preadipocytes
Jejunum
FetalBrain
FetalKidney
Liver
FetalHeart
Spleen
Thymus
Heart
Cortex
Cecum
Omasum
LymphNodes
FetalLiver
Hypothalamus
Lung
Duodenum
Kidney
Reticulum
0 200 400 600
tissue specific elements
E9: ReprPc
Myoblasts
Abomasum
Adipose
REPC
FetalMuscle
Rumen
Testis
Myocytes
Colon
Preadipocytes
Bladder
Ileum
FetalHeart
Heart
FetalLiver
BoneMarrow
Hypothalamus
Jejunum
Spleen
Thymus
Duodenum
Cortex
Omasum
Cerebellum
Muscle
LymphNodes
Lung
FetalKidney
FetalBrain
Liver
Reticulum
Cecum
Kidney
0 1000 2000 3000 4000
tissue specific elements

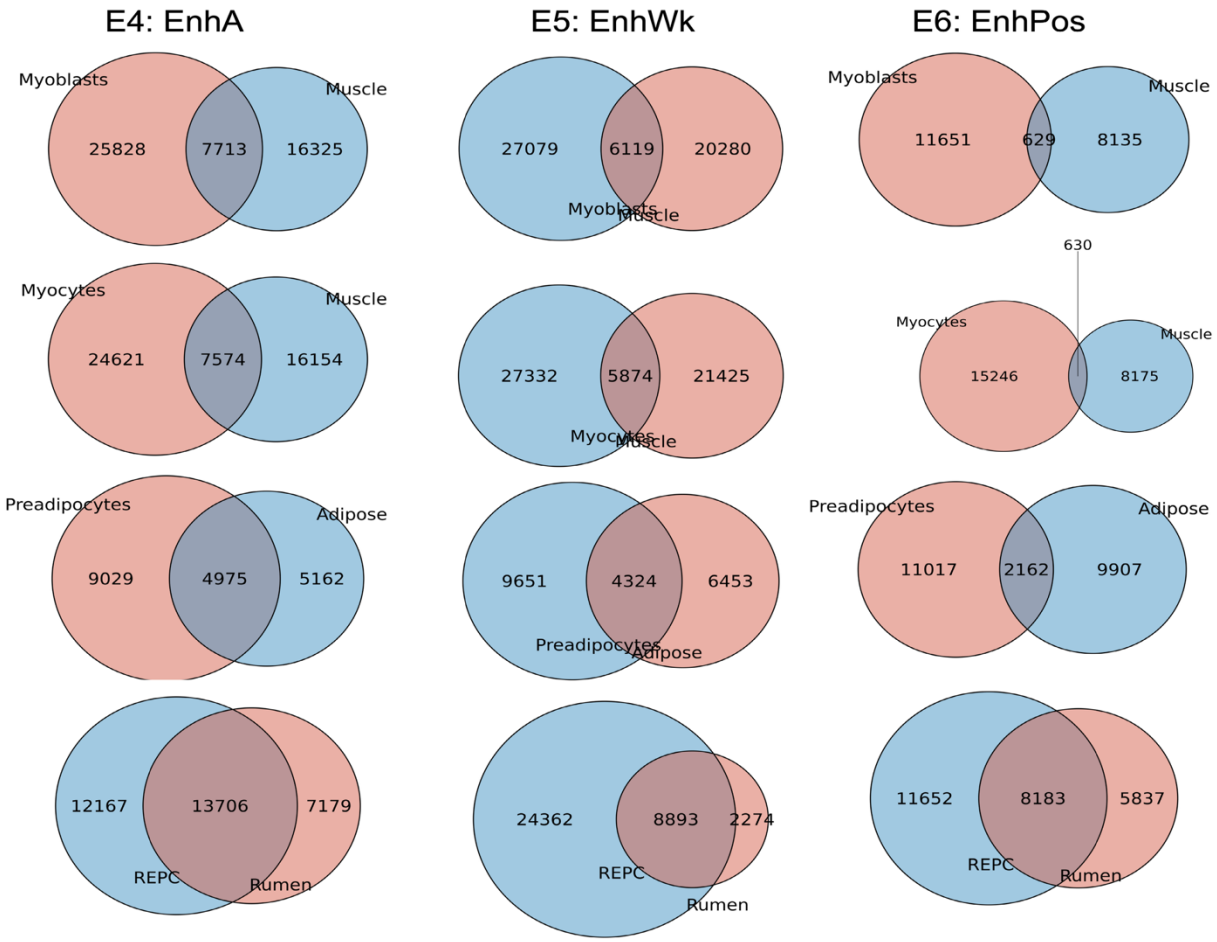
Supplementary Figure 20. Identification of tissue-specific chromatin states. Bar plots display the number of tissue-specific chromatin states (x-axis) across tissues (y-axis) for TssA (active transcription start site), TssFlnk (flanking TSS), and TssWk (weak flanking TSS); EnhA (active enhancer), EnhWk (weak enhancer), and EnhPos (poised enhancer); BivTss (bivalent TSS) and BivEnh (bivalent enhancer); and ReprPC (Polycomb-repressed regions).



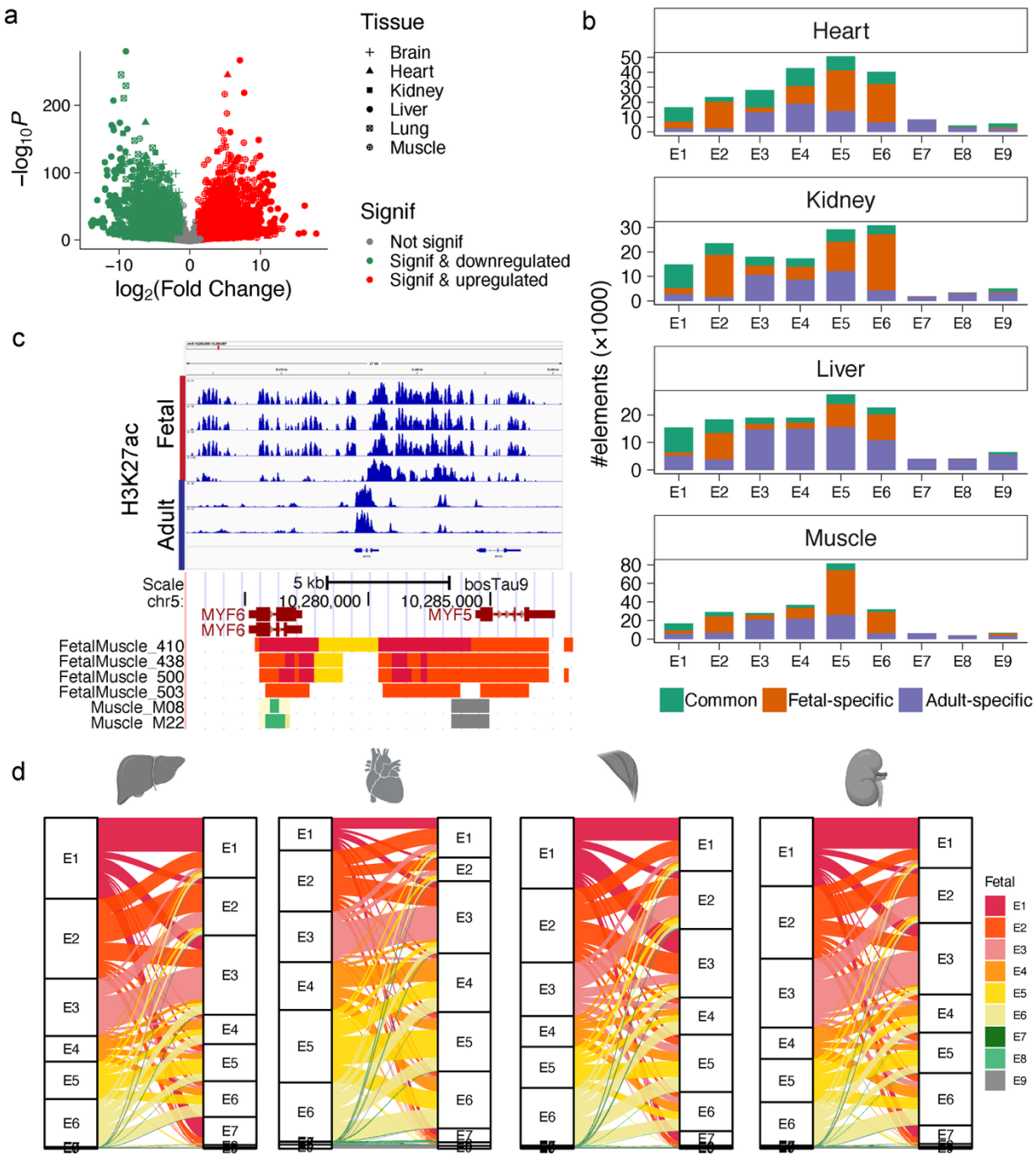
Supplementary Figure 21. Functional enrichment of tissue-specific EnhA.



Supplementary Figure 22. Comparison of promoter-like states between primary cell lime and corresponding bulk tissue.

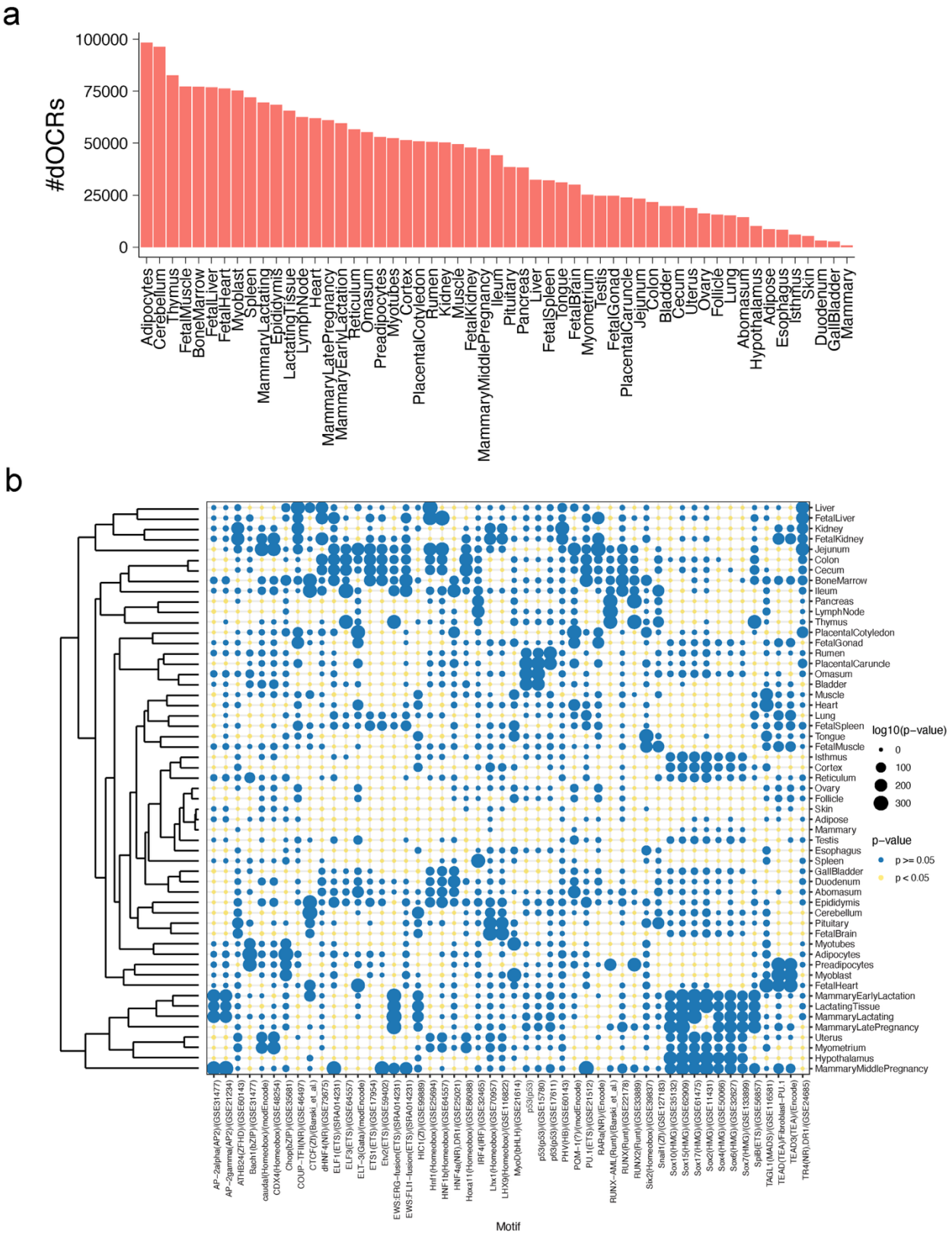


Supplementary Figure 23. Comparison of enhancer-like states between primary cell lime and corresponding bulk tissue.

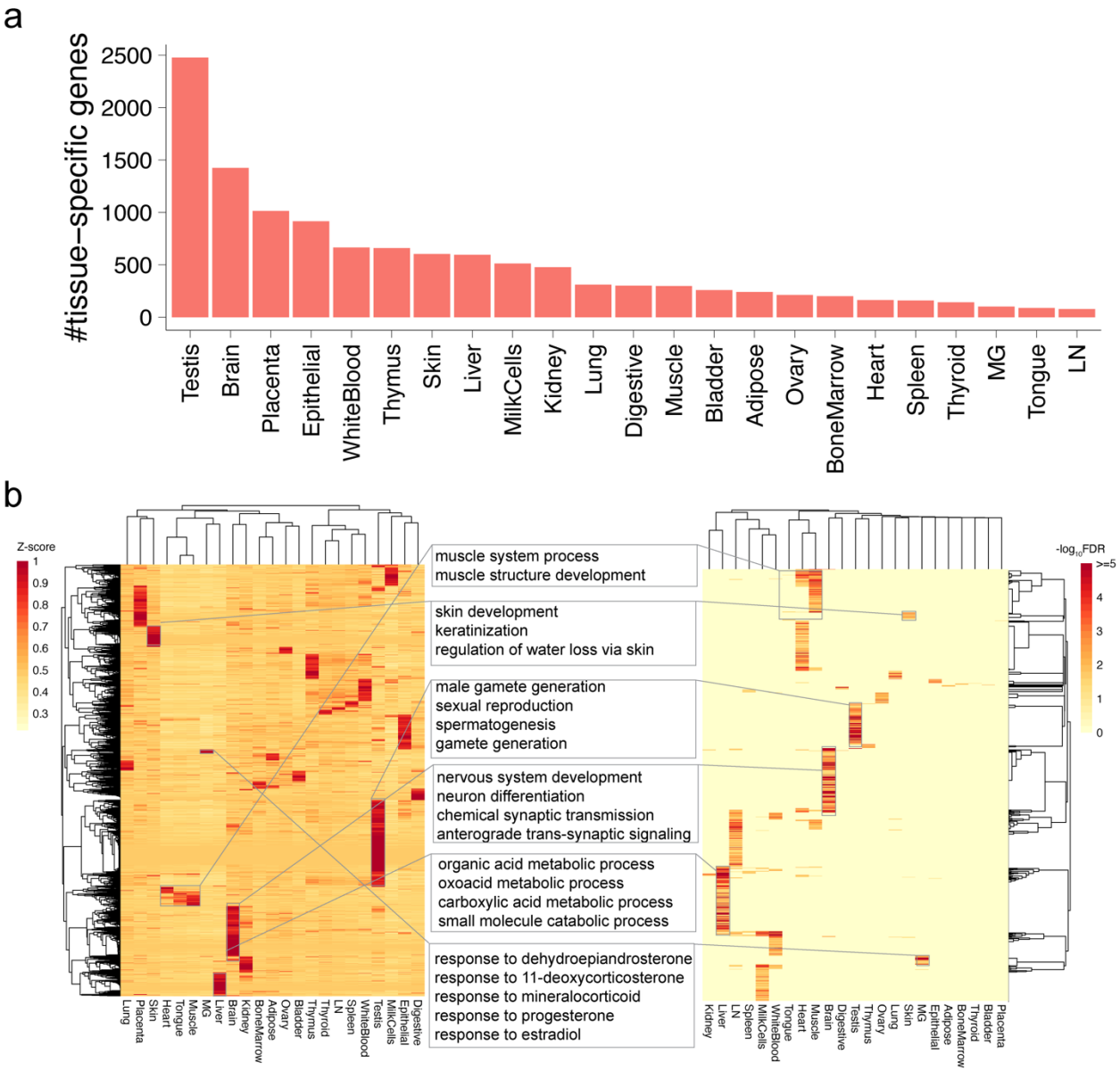


Supplementary Figure 24. Comparison of chromatin states between fetal and adult tissues. (a) Volcano plot displaying gene expression changes between fetal and adult tissues across six tissue comparisons. The x-axis shows $\log_2$ fold changes, and the y-axis represents statistical significance as $-\log_{10}(P)$; each shape denotes a different tissue comparison. (b) Bar plots showing the number of chromatin state elements categorized as common, fetal-specific, or adult-specific in heart, kidney, liver, and muscle (top to bottom). (c) Genome browser snapshots illustrating H3K27ac signal intensity and chromatin state annotations near the *MYH5* and *MYH6* loci. (d)

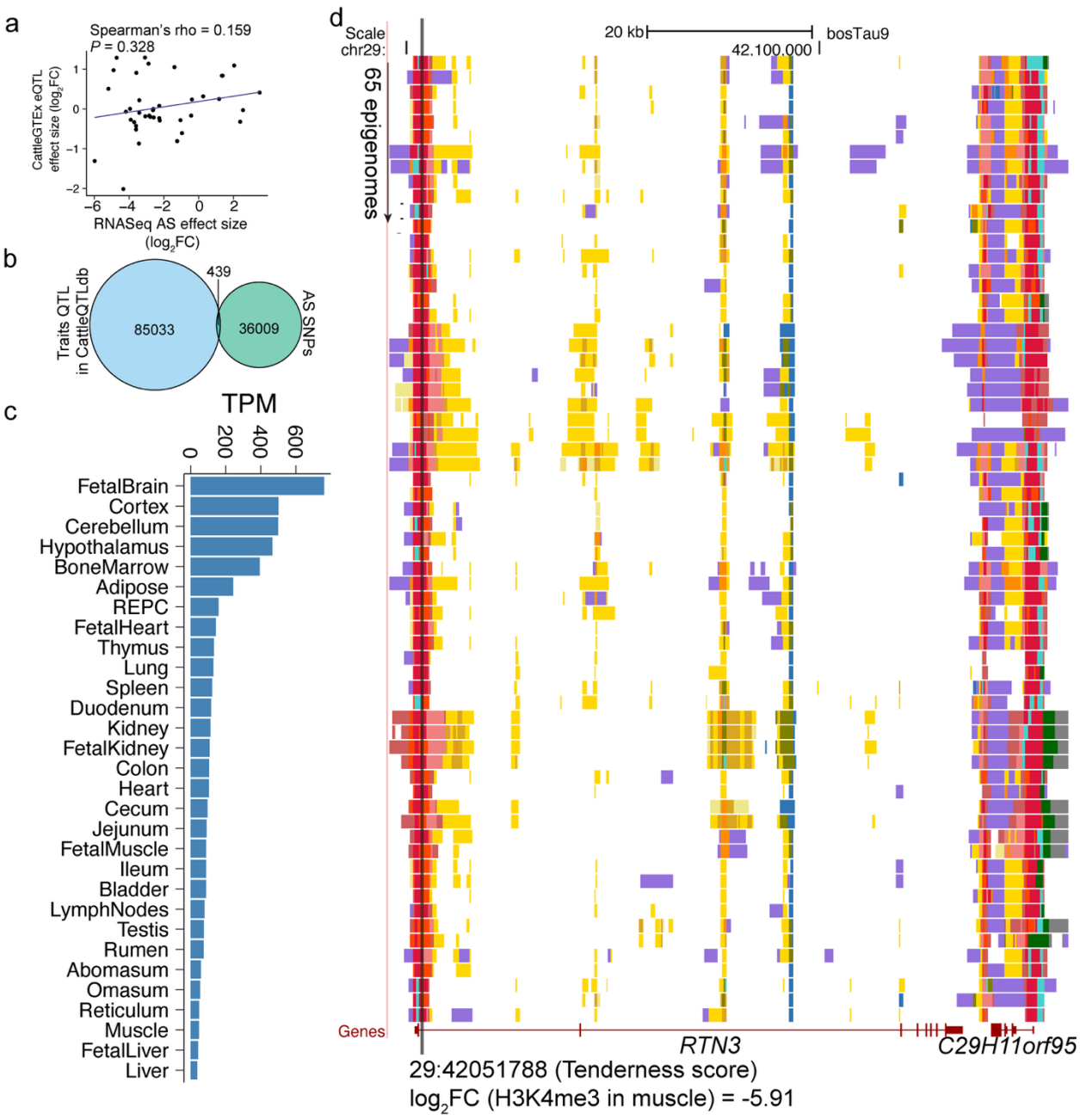
Sankey-like plots depicting chromatin state transitions from fetal (left) to adult (right) tissues for liver, heart, muscle, and kidney (left to right).


a
#dOCRs
100000
75000
50000
25000
0
Adipocytes
Cerebellum
Thymus
FetalMuscle
BoneMarrow
FetalLiver
FetalHeart
Myoblast
Spleen
MammaryLactating
Epididymis
LactatingTissue
LymphNode
Heart
MammaryLatePregnancy
MammaryEarlyLactation
Reticulum
Omasum
Preadipocytes
Myotubes
Cortex
PlacentalCotyledon
Rumen
Kidney
Muscle
FetalKidney
MammaryMiddlePregnancy
Ileum
Pituitary
Pancreas
Liver
FetalSpleen
Tongue
FetalBrain
Myometrium
Testis
FetalGonad
PlacentalCaruncle
Jejunum
Colon
Bladder
Cecum
Uterus
Ovary
Follicle
Lung
Abomasum
Hypothalamus
Adipose
Esophagus
Isthmus
Skin
Duodenum
GallBladder
Mammary
b
Liver
FetalLiver
Kidney
FetalKidney
Jejunum
Colon
Cecum
BoneMarrow
Ileum
Pancreas
LymphNode
Thymus
PlacentalCotyledon
FetalGonad
Rumen
PlacentalCaruncle
Omasum
Bladder
Muscle
Heart
Lung
FetalSpleen
Tongue
FetalMuscle
Isthmus
Cortex
Reticulum
Ovary
Follicle
Skin
Adipose
Mammary
Testis
Esophagus
Spleen
GallBladder
Duodenum
Abomasum
Epididymis
Cerebellum
Pituitary
FetalBrain
Myotubes
Adipocytes
Preadipocytes
Myoblast
FetalHeart
MammaryEarlyLactation
LactatingTissue
MammaryLactating
MammaryLatePregnancy
Uterus
Myometrium
Hypothalamus
MammaryMiddlePregnancy
log10(p−value)
0
100
200
300
p−value
p >= 0.05
p < 0.05
AP−2alpha(AP2)/(GSE31477)
AP−2gamma(AP2)/(GSE21234)
ATHB24(ZF-HD)/(GSE60143)
Bach1(bZIP)/(GSE31477)
caudal(Homeobox)/(modEncode)
CDX4(Homeobox)/(GSE48254)
Chop(bZIP)/(GSE35881)
COUP−TFII(NR)/(GSE46497)
CTCF(Zf)/(Barski_et_al.)
dHNF4(NR)/(GSE73675)
ELF1(ETS)/(SRA014231)
ELF3(ETS)/(GSE64557)
ELT−3(Gata)/(modEncode)
ETS1(ETS)/(GSE17954)
Etv2(ETS)/(GSE59402)
EWS:ERG−fusion(ETS)/(SRA014231)
EWS:FLI1−fusion(ETS)/(SRA014231)
HIC1(Zf)/(GSE59889)
Hnf1(Homeobox)/(GSE25694)
HNF1b(Homeobox)/(GSE64557)
HNF4a(NR),DR1/(GSE25021)
Hoxa11(Homeobox)/(GSE86088)
IRF4(IRF)/(GSE32465)
Lhx1(Homeobox)/(GSE70957)
LHX9(Homeobox)/(GSE116822)
MyoD(bHLH)/(GSE21614)
p53(p53)
p53(p53)/(GSE15780)
p63(p53)/(GSE17611)
PHV(HB)/(GSE60143)
POU−1(?)/(modEncode)
PU.1(ETS)/(GSE21512)
RARa(NR)/(Encode)
RUNX−AML(Runt)/(Barski_et_al.)
RUNX(Runt)/(GSE22178)
RUNX2(Runt)/(GSE33889)
Six2(Homeobox)/(GSE89837)
Snail1(Zf)/(GSE127183)
Sox10(HMG)/(GSE35132)
Sox15(HMG)/(GSE62909)
Sox17(HMG)/(GSE61475)
Sox2(HMG)/(GSE11431)
Sox4(HMG)/(GSE50066)
Sox6(HMG)/(GSE32627)
Sox7(HMG)/(GSE133899)
Spi8(ETS)/(GSE56857)
TAGL1(MADS)/(GSE116581)
TEAD(TEA)/Fibroblast−PU.1
TEAD3(TEA)/(Encode)
TR4(NR),DR1/(GSE24685)
Motif

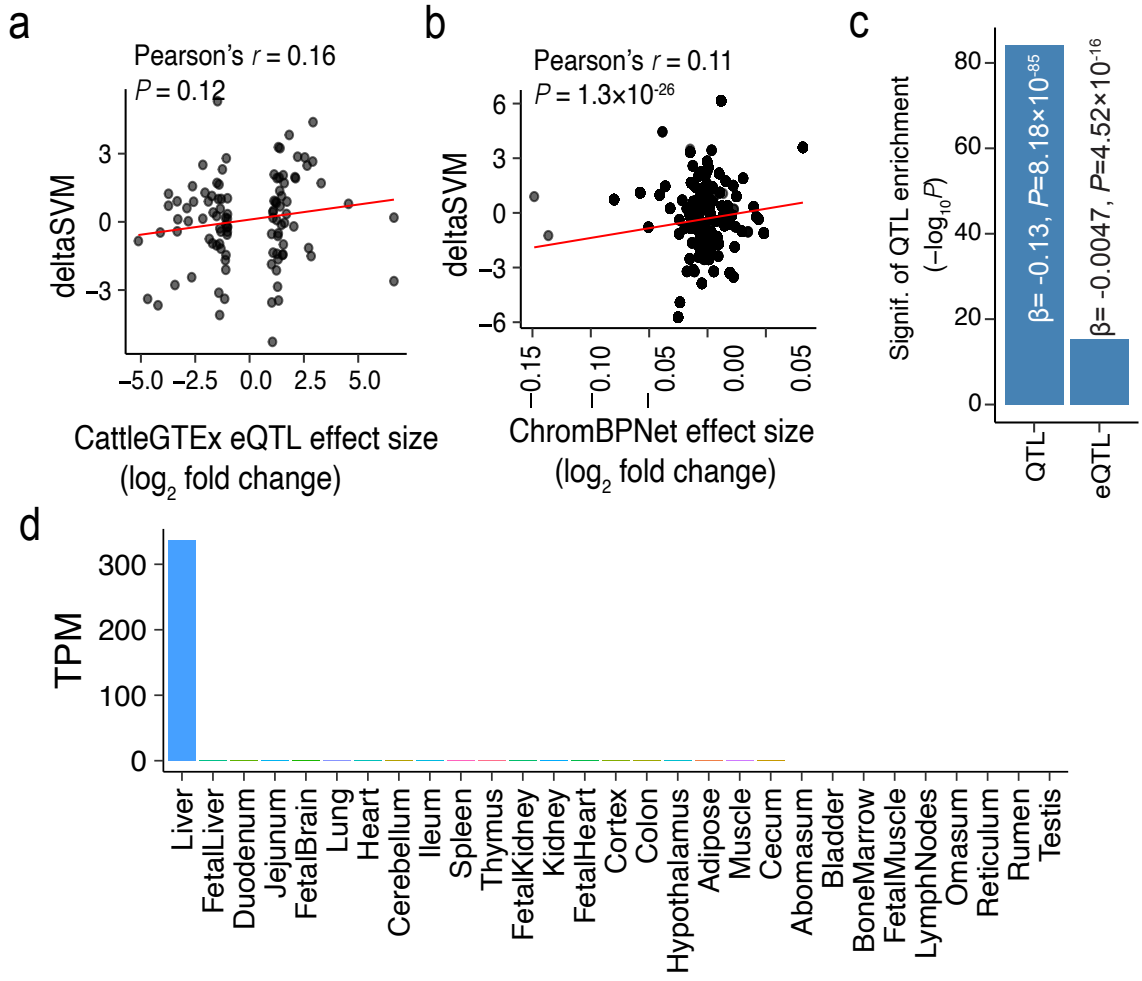
Supplementary Figure 25. Tissue specificity of chromatin accessibility. (a) Number of differential open chromatin regions (dOCRs, y-axis) across tissues being analyzed (x-axis). (b) Heatmap showing the significance of enriched motifs by dOCRs using HOMER software.



Supplementary Figure 26. Tissue specificity of gene expression. (a) Bar plot showing the number of tissue-specific genes (y-axis) identified in each tissue (x-axis). (b) Left: heatmap displaying Z-scores of gene expression across tissues. Middle: selected enriched biological processes from Gene Ontology analysis. Right: heatmap showing the significance of full functional enrichment results. All heatmaps are clustered using a distance metric of 1 – r, where r denotes Pearson's correlation coefficient. Epithelial: rumen epithelial primary cells; MG: mammary gland; LN: Lymph Node.



Supplementary Figure 27. Appendix of allele-specific (AS) regulatory variation identification. (a) Scatter plot showing the effect size (log_2 scaled) of RNA-seq AS SNPs identified in this study (x-axis) compared with the corresponding effect sizes of eQTL SNPs identified by CattleGTEx (y-axis). (b) Venn diagram depicting the overlap between RNA-seq AS SNPs identified in this study and trait-associated QTL SNPs from CattleQTLdb. (c) Gene expression profile (measured in TPM) of the *RTN3* gene across tissues. (d) Genome browser snapshots displaying chromatin states around the *RTN3* locus, highlighting the promoter region targeted by the AS SNP 29:42051788, which shows allele-specific H3K4me3 signal in muscle tissue. SNP: single nucleotide polymorphism, QTL: quantitative trait locus, eQTL: expression QTL, TPM: transcripts per million.



Supplementary Figure 28. Assessment of deltaSVM scores. (a) deltaSVM scores as a function of effect size (log_2 fold change) of eQTLs retrieved from CattleGTEx. (b) deltaSVM scores as a function of effect size (log_2 fold change) estimated by ChromBPNet. (c) Enrichment significance (y-axis) of of eQTLs (from CattleGTEx18) and trait-associated QTLs (from CattleQTLdb5) with respect to regulatory magnitude. (d) Expression measured by Transcripts Per Million (TPM) of the *MGC152010* gene.

References:

1. Prowse-Wilkins, C. P. *et al.* Putative Causal Variants Are Enriched in Annotated Functional Regions From Six Bovine Tissues. *Front. Genet.* **12**, (2021).
2. Yuan, C. *et al.* An organism-wide ATAC-seq peak catalog for the bovine and its use to identify regulatory variants. *Genome Res.* (2023) doi:10.1101/gr.277947.123.
3. Shen, X. *et al.* Single-cell transcriptome atlas revealed bronchoalveolar immune features related to disease severity in pediatric Mycoplasma pneumoniae pneumonia. *MedComm* **5**, e748 (2024).