Title: Skeletal muscle enhancer programming of cardiorespiratory fitness

Authors: Anne M. Weitzel[1], Peter Orchard[2], Charles R. Evans[3], Nandini Manickam[2], Mary K. Treutelaar[3], Steven L. Britton[1,4], Lauren G. Koch[5], Jun Z. Li[6], Stephen C.J. Parker[2,7,8]†, Charles F. Burant[1,2,3]†*

Affiliations:

[1]Molecular & Integrative Physiology, University of Michigan; Ann Arbor, MI, USA.

[2]Computational Medicine & Bioinformatics, University of Michigan; Ann Arbor, MI, USA.

[3]Internal Medicine, University of Michigan; Ann Arbor, MI, USA.

[4]Anesthesiology, University of Michigan; Ann Arbor, MI, USA.

[5]Physiology & Pharmacology, The University of Toledo; Toledo, Ohio, USA.

[6]Molecular Genetics and Genome Sciences, The University of Oklahoma Health Campus; Oklahoma City, OK, USA.

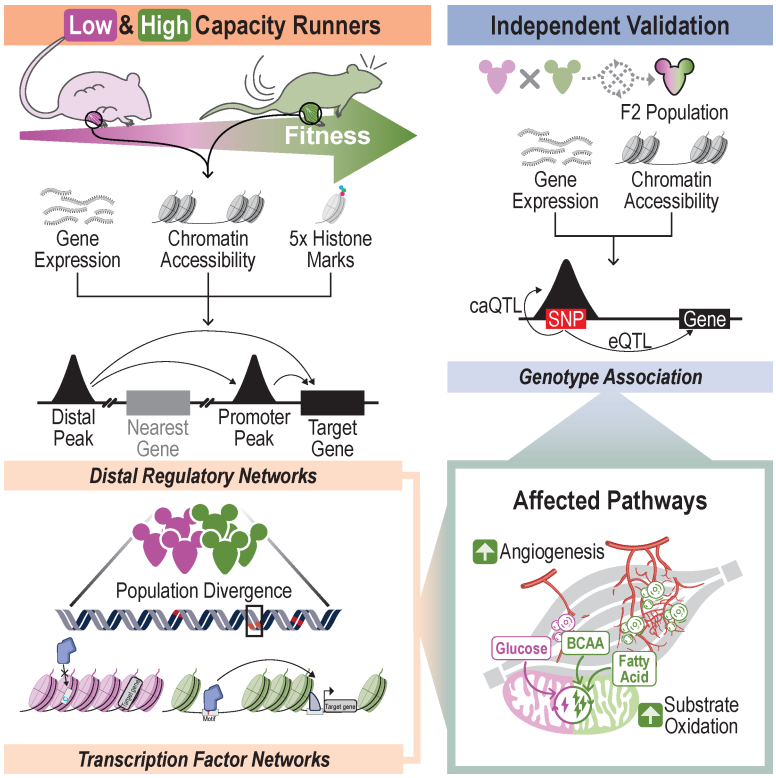
[7]Human Genetics, University of Michigan; Ann Arbor, MI, USA.

[8]Biostatistics, University of Michigan; Ann Arbor, MI, USA.

†Co-Senior Author

*Corresponding author. Email: burantc@umich.edu

Abstract: Cardiorespiratory fitness (CRF) is a heritable trait associated with improved metabolic health and longevity. To identify regulatory mechanisms underlying CRF, we integrated 546 transcriptomic and epigenomic profiles from skeletal muscle of 128 genetically heterogeneous rats selectively bred for high and low running capacity, a model that mirrors CRF-associated traits in humans. Selection drove genetic convergence in coordinated skeletal muscle enhancer networks linked to lipid metabolism and angiogenesis genes. We validated thousands of these genetic effects through integration of 426 genotype, gene expression, and chromatin accessibility profiles in an independent HCR×LCR F2 population (n=147). These 972 multi-omics profiles show that CRF-associated genetic variation reshapes the chromatin landscape to support energy metabolism and oxygen delivery, offering a molecular framework for identifying targets to reduce cardiometabolic disease risk.



Keywords: Cardiorespiratory fitness, Exercise, Metabolism, Fatty acid oxidation, Oxidative phosphorylation, Genetics, Epigenomics, Transcriptomics, Chromatin state, Histone modification

Abbreviations:

- ACAD = Acyl-CoA dehydrogenase
- AME = analysis of motif enrichment
- BCAA = branched-chain amino acid
- caQTL = chromatin accessibility quantitative trait locus
- CRF = cardiorespiratory fitness
- ETC = electron transport chain
- ETF = electron-transfer flavoprotein
- ETS = E26 transformation-specific
- eQTL = expression quantitative trait locus
- FDR = false discovery rate
- F_{ST} = fixation index
- HCR = high capacity runners
- LCR = low capacity runners
- LD = linkage disequilibrium
- PCA = principal component analysis
- RPM = reads per million
- SNP = single nucleotide polymorphism
- TCA = tricarboxylic Acid
- TF = transcription factor
- TFP = trifunctional protein
- TSS = transcription start site
- VO_2max = maximal oxygen uptake

Glossary of Terms

- Modality/modalities: essentially a dataset. A modality can refer to one or multiple of ATAC-Seq, H3K27ac & additional histone modifications, chromatin states such as Strong Enhancer, and/or RNA-Seq
- Peaks: only refers to ATAC-Seq and CUT&Tag data
- Features: can collectively refer to genomic regions of one or more modality classes (genes, peaks, and/or chromatin state regions)
- Variant(s): SNP(s)

Physical activity is recognized for its positive effect on human health and longevity (*1*, *2*). Conversely, sedentary time independently increases risk of cardiometabolic disease and mortality (*1–3*). Cardiorespiratory fitness (CRF), measured by the maximal oxygen uptake (VO_2max) during a graded exercise test, is a strong predictor of cardiometabolic risk and mortality (*4*, *5*) even when accounting for preexisting cardiovascular disease risk factors (*6*, *7*). CRF is a heritable trait with 50-75% of its variation being genetically determined (*intrinsic* CRF) (*8–10*). Regular exercise can increase VO_2max by 8-15% and has shown particularly strong fitness and longevity effects in individuals with low intrinsic CRF (*11–13*), though these individuals still face higher risks of cardiometabolic disease and premature death than their non-exercising, high CRF counterparts (*12*). While large-scale human population studies have identified genetic variants associated with CRF (*14*, *15*), which are distinct from those associated with exercise training (*16*), the regulatory mechanisms by which these variants shape skeletal muscle biology remain poorly defined.

To study intrinsic CRF, we use a genetically heterogeneous rat model selectively bred for running capacity, generating two divergent populations of High Capacity Runner (HCR) and Low Capacity Runner (LCR) lines (*17*) from the N:NIH diversity outbred rat (*18*). This well-characterized model shows persistent and heritable phenotypes that recapitulate human traits associated with divergent CRF, including sex-independent differences in physiologic measures of cardiometabolic health and longevity (*19*, *20*). As seen in humans with high CRF, the HCR line has an increased capacity for fatty acid oxidation and shows upregulation of genes involved in fatty acid and branched chain amino acid metabolism in skeletal muscle (*21*, *22*) and heart (*23*). The genetic variants fixed through selection provide a unique and tractable system to study the skeletal muscle's regulatory architecture and its contribution to intrinsic CRF.

Approximately 90% of trait-associated common genetic variants are concentrated in noncoding regions of the genome (*24*) where they can modulate transcription factor (TF) binding and epigenomic activity (*25*) in cell-specific and trait-relevant contexts. We hypothesized that selection-associated transcriptional divergence between HCR and LCR is reflected in line-specific epigenomic regulatory architecture shaped by noncoding genetic variation. To test this, we generated the first comprehensive skeletal muscle epigenomic map of CRF using 128 HCR and LCR rats, profiling chromatin accessibility via assay for transposase-accessible chromatin (ATAC-Seq) (*26*) and five gene expression-relevant histone modifications (H3K27ac, H3K4me1, H3K4me3, H3K36me3, H3K27me3) using Cleavage Under Targets and Tagmentation (CUT&Tag) (*27*). We integrated these data with transcriptome profiles and genetic variation, identifying widespread distal enhancer networks enriched for line-differentiated alleles that converge on mitochondrial inner membrane and angiogenesis gene programs.

Results

Divergent skeletal muscle epigenomic signatures associate with intrinsic cardiorespiratory fitness

To understand the previously described divergent transcriptomic patterns associated with CRF and mitochondrial function (*21*, *22*), we profiled lateral gastrocnemius muscle from 128 HCR/LCR rats balanced for sex and acute exercise conditions (Fig. S1A). We generated 546 omic profiles spanning seven modalities including chromatin accessibility (ATAC-Seq, n = 123), five histone marks (CUT&Tag, n = 59 per mark), and gene expression (RNA-Seq, n = 128) (Fig. 1A). The histone marks (H3K27me3, H3K27ac, H3K4me1, H3K4me3, and H3K36me3) alongside chromatin accessibility enable discovery of active, poised, and repressed chromatin states (Fig. 1B). As an example, we show the epigenomic signal from an individual sample surrounding *Gapdh*, a housekeeping gene expressed in skeletal muscle (Fig. 1C). H3K4me3, ATAC-Seq, and H3K27ac signals are observed at the transcription start site (TSS), indicative of an open and active promoter region. Moderate levels of H3K36me3 signal, along with low levels of the repressive H3K27me3, reflects active transcription of the gene. Study-wide consistency of our data is exemplified by signal concordance at representative loci across modalities (Fig. 1D) and within-modality clustering from principal component analysis (PCA) of genome-wide fragment counts (Fig. 1E). Notably, PC1 exhibited separation between repressive and active marks, while PC2 separated different classes of active marks, demonstrating the reproducibility and biological specificity of the diverse epigenomic profiles.

To capture the combinatorial logic of chromatin accessibility and histone modifications, we integrated all epigenomic modalities into a unified ChromHMM (*28*) model deriving 12 chromatin states (Fig. 2A). A representative chr14 locus illustrates HCR/LCR chromatin state divergence where increased ATAC-Seq and H3K4me1 signal in HCR support a Strong Enhancer state, whereas LCR exhibit a Quiescent/Low state (Fig. 2B). ChromHMM assignment of state probabilities across genome-wide loci enables systematic quantification of regulatory divergence. Processing data for gene expression, chromatin accessibility, histone modification peak abundance, and probabilistic chromatin state assignment resulted in 14,781 **genes**, 101,638 chromatin accessible **peaks**, 45k-91k histone modification **peaks**, and 97k-124k chromatin state **regions**, respectively (Fig. 2C). Each of these datasets is treated as a distinct **modality** represented by a collection of genomic loci with sample-level values, collectively referred to as **features**.

To assess epigenomic differences in parallel with transcriptional differences between HCR and LCR lines, all analyses were adjusted for sex, training, and acute exercise to isolate **selection-associated CRF effects** (Fig. S4E-F). Features increased in HCR or LCR were interpreted as candidate regulatory features associated with high or low CRF, respectively. Using modality-specific methods (Fig. 2C) that normalized for technical and experimental variables (Fig. S4B), we identified thousands of differences in gene expression, chromatin accessibility, histone abundance, and chromatin state probabilities between HCR and LCR (Fig. 2C). For example, Fig. 2D shows the magnitude of differences in H3K4me1 peak abundance with 3,707 HCR up and 2,560 LCR up peaks (FDR < 0.1 & $|\log_2 FC| > 0.5$), indicating selection-driven alterations in enhancer elements. These results demonstrate widespread CRF-associated epigenomic divergence across active, poised, and repressed chromatin states (Fig. 2C, Fig. S6A-F).

To test whether this epigenomic divergence reflects underlying genetic selection for running capacity, we imputed genotypes from RNA-Seq derived single nucleotide polymorphisms (SNPs) (637,713 variants) and quantified allele frequency divergence between lines by calculating per-variant fixation indices (F_{ST}). We then assessed whether genetic differentiation is more pronounced in differential epigenomic peaks by performing linear regression between peak-level F_{ST} and HCR/LCR peak significance ($-\log_{10}$(*p*-value); Fig. 2E). F_{ST} scaled with the magnitude of epigenomic differences, particularly for ATAC-Seq and H3K4me1 peaks (Fig. 2F). These findings implicate selection on regulatory variants in chromatin remodeling at enhancers, though the downstream transcriptional targets remain undefined.

RNA-Seq analysis defines the transcriptional landscape central to intrinsic cardiorespiratory fitness

A transcriptional reference point for the CRF phenotype in these rats was established through pathway enrichment analysis of the HCR/LCR differential gene expression. As described previously (*21*, *22*), genes upregulated in HCR were predominantly enriched for mitochondrial-related processes including aerobic respiration, oxidative phosphorylation, fatty acid oxidation, and BCAA degradation (FDRs < 4.26×10^{-9}) (Fig. S7A-B), consistent with previous studies (*21*, *22*). A distinct signal related to angiogenesis was also identified (23 gene sets, FDRs = 6.9×10^{-5} to 3.6×10^{-2}, highlighted in Fig. S7A). These pathways coincide with known CRF phenotypes of enhanced mitochondrial function (*29*) and capillary density (*30–33*) in HCR, mirroring that observed in humans with high intrinsic CRF (*21*, *29*, *34–36*). These results define distinct but related biological processes that provide a framework for interpreting how selection of distal regulatory elements shapes gene expression in skeletal muscle.

Multimodal correlation analysis identifies distal epigenomic and promoter networks linked to the expression of genes involved in mitochondrial metabolism

The concentration of genetic variation within CRF-associated epigenomic peaks indicates that selection targets multiple regulatory regions scattered across the genome, rather than at a single locus. On average, differential epigenomic regions between HCR and LCR lines (Fig. 2C) reside 143.4 kb from the nearest TSS, consistent with widespread distal enhancer activity. We linked regulatory elements to target genes by applying a peak-to-gene (Pk2G) correlation framework (Fig. 3A), annotating 88,440 peaks to 10,786 genes (FDR < 0.05, Fig. 3B). Peaks were linked to a median of 2 genes (Fig. 3C), reflecting the selectivity of enhancer-gene relationships (*37*, *38*). Notably, only 17.5% of the involved peaks most strongly correlated to the nearest gene (Fig. 3C), consistent with regulatory elements bypassing nearby genes to modulate distal targets (*39–41*). This map of gene regulation is illustrated at the *Decr1* locus, a mitochondrial fatty acid beta-oxidation gene (*42*), where multiple regulatory elements show both positive and negative correlation with expression (Fig. 3D) (*43*).

Ranked pathway enrichment analysis on the Pk2G target genes linked to differential HCR/LCR peaks (FDR threshold in Fig. 2C), stratified by positive and negative regulatory relationships (Fig. 3E), identified biological processes most directly associated with CRF-divergent epigenomic regulation. Genes positively linked to HCR up peaks and negatively linked to LCR up peaks (Fig. 3F) are enriched for aerobic respiration, fatty acid and BCAA metabolism, as well as endothelial angiogenesis (Fig. 3G), mirroring RNA-Seq enrichment patterns (Fig. S7A). Genes positively linked to LCR up peaks are enriched only for contractile fiber gene sets, driven by glycolytic type II fiber genes (*Actn3*, *Myl1*, *Casq1*; boxed in Fig. 3G), consistent with greater reliance on glycolytic metabolism during acute exercise in LCR (*29*). Line-specific analyses preserved both the strength and directionality of the raw Pk2G correlations driving enrichment (Fig. 3H-I), indicating that selection amplifies epigenomic regulation of energy metabolism and vascularization which is central to high CRF.

Among the enriched clusters with epigenomic and transcriptional concordance, we highlight the "small molecule catabolic process" cluster as it represents a set of biologically relevant genes that function in oxidative fatty acid metabolism. In this cluster, 123 HCR up peaks were positively linked to 129 target genes. Of these, 123 genes overlapped with the 426 genes that enriched this cluster from the RNA-Seq differential analysis. Comparison of the differential gene expression and Pk2G significance values highlights *Cyp2j4*, *Acadsb*, *Acad11*, *Mtm1*, and *Eci2* as top candidates under strong selection-driven epigenomic regulation (Fig. 3J). These findings demonstrate how selection within the epigenome results in coordinated transcriptional control of mitochondrial efficiency through preferential fatty acid and BCAA utilization, core metabolic processes associated with CRF.

Distinct transcription factor networks in cardiorespiratory fitness associated enhancer regions regulate energy metabolism and vascularization

We next asked whether distinct TF binding sites underlie the CRF-divergent enhancers linked to mitochondrial and angiogenic gene programs by comparing motif enrichment between HCR up regions and LCR up regions (Fig. 4A). We identified 146 differential motifs, represented by 78 TFs (Fig. 4B, Fig. S9B), independent of changes in TF abundance (Fig. S10B). We performed over-representation analysis of differential motif associated Pk2G target genes stratified by modality, motif, peak direction, and Pk2G link directionality and identified 905 significant gene sets (FDR < 0.05) predominantly driven by motifs localized to differential ATAC-Seq and H3K4me1 peaks (Fig. 4C). HCR up peaks enriched for E26 transformation-specific (ETS) family TF motifs and LCR up peaks enriched for Mef2 motifs converge on mitochondrial oxidative phosphorylation, fatty acid oxidation, BCAA oxidation, mitochondrial ribosomes, and vascular development gene targets, but with positive and negative Pk2G links, respectively (Fig. 4D). Selection appears to have acted cumulatively across enhancers of multiple genes, favoring allelic variants that establish a hierarchical regulatory architecture in which ETS-mediated activation overrides Mef2-associated constraint to enhance expression of genes supporting mitochondrial oxidative capacity and capillary density that underlie high CRF.

Line-divergent alleles in cardiorespiratory fitness associated enhancers reshape transcription factor binding potential to influence metabolic gene expression

The enrichment of line-divergent variants in CRF-associated enhancers raised the question of whether line-specific alleles can directly influence TF binding affinity. To test this, we compared line-specific sequences within the same epigenomic regions by substituting HCR- and LCR-preferred alleles ($F_{ST} > 0.3$), stratified by direction of change (HCR up, LCR up) and modality (Fig. 5A). We identified 246 differentially enriched motifs representing 121 TFs (Fig. 5B, Fig. S9A), predominantly in the Cys2-His2 zinc finger (zf-C2H2), ETS, Forkhead, and basic helix-loop-helix (bHLH) TF families, all with established roles in skeletal muscle identity (*44–50*) and mitochondrial function (*23*, *51–56*). Importantly, 91.9% of the differentially enriched motifs correspond to TFs that are not differentially expressed between HCR and LCR (Fig. S10A), reinforcing the finding that selection acts primarily through altered TF binding potential rather than TF abundance.

We identified 799 disrupted motifs linking specific alleles to gene expression through predicted changes in TF binding affinity. The Sp1/Acad11 locus illustrates this regulatory logic (Fig. 5C-E): a variant at Sp1 motif position 10 (chr8:104691692, $F_{ST} = 0.54$) strengthens potential binding of Sp1 in HCR (G allele frequency = 0.80) relative to LCR (C allele frequency = 0.92) despite no differential expression of *Sp1* between lines ($\log_2FC = -0.15$, FDR $= 6.64\times10^{-10}$; Fig. 5E). Rather, the HCR G allele coincides with elevated signal in ATAC-Seq, H3K27ac, H3K4me1, and H3K4me3 peaks (Fig. 5D) that are all positively linked to expression of *Acad11* (Fig. 5C), a fatty

acid dehydrogenase (*57*) upregulated in HCR ($\log_2$FC = 1.04, FDR = 3.59×10^{-66}). These observations suggest a regulatory chain in which a variant alters motif affinity, reshapes local chromatin, and ultimately influences gene expression, independent of TF abundance.

To generalize our biological understanding beyond the highlighted Sp1 motif, we performed over-representation analysis of Pk2G target genes linked to all disrupted motifs, stratified by peak direction (HCR up, LCR up) and Pk2G link direction (positive, negative). Genes positively linked to HCR preferred motifs are enriched for fatty acid metabolism and aerobic respiration (FDR < 0.1; Fig. 5F), extending the pathway convergence observed in the initial motif analysis to specific allelic variants (Fig. 5G). To establish this as a molecular mechanism by which genetic divergence shapes CRF physiology, we next tested whether these allele-dependent regulatory effects are preserved in an independent population carrying the same variants.

Genetic variation in an independent population validates regulatory mechanisms underlying cardiorespiratory fitness

The preceding analyses established that CRF-divergent enhancers contain line-differentiated alleles predicted to alter TF binding potential. To test whether the same genetic effects inferred from the HCR/LCR model extend beyond the original selection for these lines, we performed a large-scale validation study using an HCR×LCR (F2) cross population we created 11 generations prior, whose segregating genotypes produced heritable variation in running capacity and muscle phenotypes (*54*) (Fig. 6A). We genotyped the F2 population and generated RNA-Seq (n = 147) and ATAC-Seq (n = 132) from gastrocnemius muscle to conduct expression quantitative trait locus (eQTL) and chromatin quantitative trait locus (caQTL) analyses (Fig. 6B). Accounting for linkage disequilibrium (LD), we identified 4,747 genes with a significant eQTL (eGenes, FDR < 0.05) and 10,099 peaks with a significant caQTL (caPeaks, FDR < 0.05; Fig. 6B), providing a framework to evaluate how genetic variation contributes to transcriptional and epigenomic regulation underlying intrinsic CRF.

Next, we tested whether the variants associated with gene expression or chromatin accessibility in F2 show directional concordance with the gene expression and chromatin activity differences observed between HCR and LCR. For example, if an HCR-predominant allele positively associates with ATAC-Seq signal in F2, the corresponding peak should be elevated in HCR relative to LCR (Fig. 6C). We assessed concordance across three comparisons: F2 caQTL effects versus HCR/LCR epigenomic differences (94% concordance; Fig. 6D), F2 eQTL effects versus HCR/LCR gene expression differences (85% concordance; Fig. 6E), and F2 eQTL effects versus HCR/LCR Pk2G links (69% concordance, Fig. 6F). All three showed significant directional concordance (binomial *p*-value $\leq 3.15\times10^{-17}$ for all), indicating that genetic variants exert consistent regulatory effects across populations. The Pk2G concordance further validates that the correlation-based peak-gene links capture genuine regulatory relationships. These associations include genes involved in mitochondrial metabolism and vascular development (colored in Fig. 6E-F) and disrupted motifs from the HCR/LCR analyses (outlined in Fig. 6D,F).

The Sp1/Acad11 locus exemplifies this validation. The variant chr8:104691692 within the Sp1 motif, which strengthens binding potential in HCR (Fig. 5C-E), exhibits the same allelic effects on chromatin accessibility and gene expression in both populations. The ATAC-Seq peak is positively correlated with *Acad11* expression in HCR and LCR (Fig. 6G) and F2 (Fig. 6J). In the HCR/LCR model, the presence of the C allele is positively associated with the intronic-residing ATAC-Seq peak (Fig. 6H) and *Acad11* expression (Fig. 6I). The F2 genotypes at variant chr8:104691692 are also significantly associated with chromatin accessibility of the overlapping ATAC-Seq peak from the HCR/LCR dataset (caQTL $\beta = 0.47$, *p*-value = 5.1×10^{-4}; Fig. 6K) and expression of *Acad11* (eQTL $\beta = 1.11$, *p*-value = 4×10^{-37}; Fig. 6L). This illustrates how noncoding variation propagates through the regulatory cascade, where a single variant can enhance *Sp1* binding potential with paralleled increases in chromatin accessibility and expression of *Acad11*. This is one of the 293 validated variant-peak-gene examples that likely shape the regulatory landscape underlying CRF.

Ultimately, the concordance between F2 genetic effects and the HCR/LCR epigenomic and transcriptomic differences demonstrates that genetic divergence drives a substantial fraction of the regulatory differences between lines. It further indicates that the broader epigenomic and transcriptomic differences observed between HCR and LCR reflect heritable regulatory divergence rather than line-specific artifacts. Because a large portion of genetic variants associated with complex traits likely exert their effects through gene regulation (*24*, *58*), the

regulatory variants identified here should be of particular interest as entry points for dissecting the regulatory logic linking noncoding variation to CRF.

Discussion

Selection for CRF is reflected in the regulatory architecture connected to genes involved in mitochondrial oxidative metabolism and vascular development (Fig. 4). By integrating 546 transcriptomic and epigenomic profiles from N:NIH diversity outbred rats (*18*) selectively bred for high and low running capacity (*17*), we demonstrate that 39 generations of selection has concentrated pre-existing allelic variants from the founder population at distal enhancers, particularly those marked by H3K4me1, reshaping the skeletal muscle chromatin landscape. These CRF-associated epigenomic regions are characterized by high densities of ETS-family TF motifs in HCR and Mef2 TF motifs in LCR, suggesting that selection acted on clustered noncoding variants where multiple discrete regulatory elements cumulatively (*59*, *60*) shape divergent transcriptional profiles, as proposed for complex trait architecture (*61*). The convergence of multiple regulatory elements on the same target genes (Fig. 3D) is reminiscent of the "shadow" enhancer architecture described for developmental loci (*62*), in which distributed elements buffer expression against individual mutations while providing substrates for cumulative selection observed here. Concordant allelic effects on chromatin accessibility and gene expression in the independent F2 cross population confirmed regulatory effects established in the HCR/LCR study (Fig. 6).

Our findings extend the prevailing model of mitochondrial gene regulation in skeletal muscle, which centers on PGC-1α coactivation of Nrf1 and Nrf2/Gabpa at proximal promoters (*63*, *64*), to distal enhancer networks shaped by CRF-associated noncoding genetic variation. Expression of these trans-acting components is largely conserved between lines (*Ppargc1a* $\log_2$FC = 0.48, FDR = 1×10^{-14}; *Nrf1* $\log_2$FC = -0.03, FDR = 0.23; *Gabpa* $\log_2$FC = -0.14, FDR = 3.5×10^{-11}) and does not account for gene-specific regulatory divergence observed across thousands of distal enhancers linked to oxidative metabolism and angiogenesis targets (Fig. 3, Fig. 4). We identify this complementary cis-regulatory layer as a network of distal ETS-family enriched enhancers (including Gabpa; Fig. 4D) in which divergent TF binding opportunity tunes which genes are brought under PGC-1α mediated control in a given regulatory context.

While allelic effects on TF binding affinity through motif disruption were identified (Fig. 5), widespread CRF-associated epigenomic divergence in the absence of direct motif disruption (Fig. 4) may reflect proximal and distal variant effects on TF binding, chromatin accessibility, and enhancer activity (*65–70*). Distinct TF networks underlie CRF-associated regulatory control of genes involved in mitochondrial oxidative metabolism through ETS-enriched enhancers positively correlated with gene expression in HCR and Mef2-enriched enhancers negatively correlated with expression in LCR. This supports a hierarchical regulatory architecture where the incremental accumulation of variants favoring ETS binding potential supersedes basal Mef2-associated constraint. This ETS-mediated control extends to angiogenesis (Fig. 4D, Fig. S7A), suggesting that selection for running capacity targeted a shared regulatory node of metabolic-vascular coupling. Skeletal muscle and endothelial cells derive from common bipotent progenitors in which ETS-family TFs, including Fli1 and Gabpa, play lineage-determining roles (*71–74*), and the prominence of CRF-divergent H3K4me1-marked enhancers raises the possibility that some regulatory elements were established during this shared developmental window and subsequently co-opted for tissue-specific gene regulation. This would explain why selection on ETS binding potential coordinately elevates oxidative capacity and vascular supply in HCR, reflected in phenotypes of greater skeletal muscle mass (*31*), increased skeletal muscle mitochondria (*75–77*) and capillary density (*30–33*), and higher type I and IIa fiber content (*30*, *78*) compared to LCR, supporting metabolic efficiency and delayed depletion of muscle glycogen during intense exercise (*29*).

The biological consequence of this selection-driven enhancer remodeling is a coordinated upregulation of genes that enhance electron flux through the mitochondrial inner membrane at multiple entry points (Fig. S12). While previous work identified ETS-family TF binding at individual mitochondrial gene promoters, including the ATP synthase beta-subunit (*51*) and cytochrome c oxidase subunits (*79*), our results demonstrate broader involvement of ETS-enriched distal enhancers linked to fatty acid oxidation, TCA cycle, and oxidative phosphorylation genes in skeletal muscle. Acyl-CoA dehydrogenases (ACADs) are prominent ETS-enriched enhancer targets that play

a central role in β-oxidation with direct routing of electrons to Co-Enzyme Q (CoQ) via the electron-transfer flavoprotein (ETF) and ETF-QO (*Etfdh*, $\log_2$FC = 0.54, FDR = 5.5×10^{-83}) (*80*), along with the downstream delivery of NADH to Complex I by trifunctional protein generation of Acetyl-CoA. ETS-enriched regulatory domains are further linked to the upregulation of ETC complex I-V subunits, as well as fumarate hydratase, which sustains TCA flux by reducing product inhibition of Complex II through conversion of fumarate to malate. In conjunction with acyl-CoA dehydrogenase isoforms mitigating oxidative damage through peroxidized fatty acid metabolism (*81*, *82*), this optimized metabolic flux is also protective against mitochondrial dysfunction observed in individuals predisposed to type 2 diabetes (*83–85*) through the reduction of lipotoxic intermediate accumulation, (*86*), short chain acyl-CoA accumulation (*87*), and reactive oxygen species by reducing mitochondrial membrane potential (*88*). Notably, many of the fatty acid and BCAA degrading dehydrogenases upregulated in HCR are evolutionarily ancient genes derived from α-proteobacteria with orthologs across Archaea, Bacteria, and Eukaryota (*89*, *90*), suggesting that coding sequences and enzymatic function are conserved while regulatory adaptation fine-tunes expression based on energy availability and demands.

Intrinsic CRF is a complex trait that serves as a determinant of cardiometabolic health and a reproducible predictor of longevity (*19*). Selective breeding of HCR and LCR rats gradually amplified phenotypes directly relevant to human cardiometabolic health (*20*), with the greatest impact on skeletal muscle as observed by fewer mitochondrial gene expression differences in heart (*91*), liver (*92*), and adipose tissue (*93*) between the lines. We identified coordinated epigenomic tuning of skeletal muscle fatty acid metabolism, BCAA catabolism, and vascular remodeling. From a modern health perspective, regulatory variants associated with high CRF may counteract the accumulation of fatty acids, reducing mitochondrial damage in skeletal muscle associated with sarcopenia, aging (*94*), obesity, and insulin resistance (*95*, *96*). Beyond CRF, our framework consisting of multi-omic profiling and integration to identify peak-gene regulatory networks, nomination of TFs through motif enrichment, and cross-population validation of allelic effects provides a template for dissecting complex trait architecture (Fig. S11).

References

1. I.-M. Lee, E. J. Shiroma, F. Lobelo, P. Puska, S. N. Blair, P. T. Katzmarzyk, Lancet Physical Activity Series Working Group, Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet Lond. Engl.* **380**, 219–229 (2012).

2. Physical Activity Guidelines for Americans (2023). https://health.gov/paguidelines.

3. C. Bouchard, S. N. Blair, P. T. Katzmarzyk, Less Sitting, More Physical Activity, or Higher Fitness? *Mayo Clin. Proc.* **90**, 1533–1540 (2015).

4. J. J. Lang, S. A. Prince, K. Merucci, C. Cadenas-Sanchez, J.-P. Chaput, B. J. Fraser, T. Manyanga, R. McGrath, F. B. Ortega, B. Singh, G. R. Tomkinson, Cardiorespiratory fitness is a strong and consistent predictor of morbidity and mortality among adults: an overview of meta-analyses representing over 20.9 million observations from 199 unique cohort studies. *Br. J. Sports Med.* **58**, 556–566 (2024).

5. S. N. Blair, H. W. Kohl, R. S. Paffenbarger, D. G. Clark, K. H. Cooper, L. W. Gibbons, Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA* **262**, 2395–2401 (1989).

6. Ross Robert, Blair Steven N., Arena Ross, Church Timothy S., Després Jean-Pierre, Franklin Barry A., Haskell William L., Kaminsky Leonard A., Levine Benjamin D., Lavie Carl J., Myers Jonathan, Niebauer Josef, Sallis Robert, Sawada Susumu S., Sui Xuemei, Wisløff Ulrik, Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. *Circulation* **134**, e653–e699 (2016).

7. B. A. Franklin, I. J. Wedig, R. E. Sallis, C. J. Lavie, S. J. Elmer, Physical Activity and Cardiorespiratory Fitness as Modulators of Health Outcomes. *Mayo Clin. Proc.* **98**, 316–331 (2023).

8. N. M. Schutte, I. Nederend, J. J. Hudziak, M. Bartels, E. J. C. De Geus, Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption. *Physiol. Genomics* **48**, 210–219 (2016).

9. M. A. Sarzynski, T. K. Rice, J.-P. Després, L. Pérusse, A. Tremblay, P. R. Stanforth, A. Tchernof, J. L. Barber, F. Falciani, C. Clish, J. M. Robbins, S. Ghosh, R. E. Gerszten, A. S. Leon, J. S. Skinner, D. C. Rao, C. Bouchard, The HERITAGE Family Study: A Review of the Effects of Exercise Training on Cardiometabolic Health, with Insights into Molecular Transducers. *Med. Sci. Sports Exerc.* **54**, S1–S43 (2022).

10. C. Bouchard, P. An, T. Rice, J. S. Skinner, J. H. Wilmore, J. Gagnon, L. Pérusse, A. S. Leon, D. C. Rao, Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study. *J. Appl. Physiol. Bethesda Md 1985* **87**, 1003–1008 (1999).

11. S. N. Blair, Y. Cheng, J. S. Holder, Is physical activity or physical fitness more important in defining health benefits? *Med. Sci. Sports Exerc.* **33**, S379-399; discussion S419-420 (2001).

12. C. A. Celis-Morales, D. M. Lyall, J. Anderson, S. Iliodromiti, Y. Fan, U. E. Ntuk, D. F. Mackay, J. P. Pell, N. Sattar, J. M. R. Gill, The association between physical activity and risk of mortality is modulated by grip strength and cardiorespiratory fitness: evidence from 498 135 UK-Biobank participants. *Eur. Heart J.*, ehw249 (2016).

13. R. Ross, B. H. Goodpaster, L. G. Koch, M. A. Sarzynski, W. M. Kohrt, N. M. Johannsen, J. S. Skinner, A. Castro, B. A. Irving, R. C. Noland, L. M. Sparks, G. Spielmann, A. G. Day, W. Pitsch, W. G. Hopkins, C. Bouchard, Precision exercise medicine: understanding exercise response variability. *Br. J. Sports Med.* **53**, 1141–1153 (2019).

14. M. Klevjer, A. N. Nordeidet, A. F. Hansen, E. Madssen, U. Wisløff, B. M. Brumpton, A. Bye, Genome-Wide Association Study Identifies New Genetic Determinants of Cardiorespiratory Fitness: The Trøndelag Health Study. *Med. Sci. Sports Exerc.* **54**, 1534–1545 (2022).

15. A. Bye, M. Klevjer, E. Ryeng, G. J. J. D. Silva, J. B. N. Moreira, D. Stensvold, U. Wisløff, Identification of novel genetic variants associated with cardiorespiratory fitness. *Prog. Cardiovasc. Dis.* **63**, 341–349 (2020).

16. K. B. Hanscombe, E. Persyn, M. Traylor, K. P. Glanville, M. Hamer, J. R. I. Coleman, C. M. Lewis, The genetic case for cardiorespiratory fitness as a clinical vital sign and the routine prescription of physical activity in healthcare. *Genome Med.* **13**, 180 (2021).

17. L. G. Koch, S. L. Britton, Artificial selection for intrinsic aerobic endurance running capacity in rats. *Physiol. Genomics* **5**, 45–52 (2001).

18. V. A. Wagner, K. L. Holl, K. C. Clark, J. J. Reho, H.-J. Lehmler, K. Wang, J. L. Grobe, M. R. Dwinell, H. Raff, A. E. Kwitek, The Power of the Heterogeneous Stock Rat Founder Strains in Modeling Metabolic Disease. *Endocrinology* **164**, bqad157 (2023).

19. L. G. Koch, O. J. Kemi, N. Qi, S. X. Leng, P. Bijma, L. J. Gilligan, J. E. Wilkinson, H. Wisløff, M. A. Høydal, N. Rolim, P. M. Abadir, I. Van Grevenhof, G. L. Smith, C. F. Burant, Ø. Ellingsen, S. L. Britton, U. Wisløff, Intrinsic aerobic capacity sets a divide for aging and longevity. *Circ. Res.* **109**, 1162–1172 (2011).

20. U. Wisløff, S. M. Najjar, O. Ellingsen, P. M. Haram, S. Swoap, Q. Al-Share, M. Fernström, K. Rezaei, S. J. Lee, L. G. Koch, S. L. Britton, Cardiovascular risk factors emerge after artificial selection for low aerobic capacity. *Science* **307**, 418–420 (2005).

21. J. Y. Fleischman, F. Van Den Bergh, N. L. Collins, M. Bowers, D. A. Beard, C. F. Burant, Higher mitochondrial oxidative capacity is the primary molecular differentiator in muscle of rats with high and low intrinsic cardiorespiratory fitness. *Mol. Metab.* **76**, 101793 (2023).

22. Y. Ren, L. G. Koch, S. L. Britton, N. R. Qi, M. K. Treutelaar, C. F. Burant, J. Z. Li, Selection-, age-, and exercise-dependence of skeletal muscle gene expression patterns in a rat model of metabolic fitness. *Physiol. Genomics* **48**, 816–825 (2016).

23. J. Shintaku, J. M. Peterson, E. E. Talbert, J.-M. Gu, K. J. Ladner, D. R. Williams, K. Mousavi, R. Wang, V. Sartorelli, D. C. Guttridge, MyoD Regulates Skeletal Muscle Oxidative Metabolism Cooperatively with Alternative NF-κB. *Cell Rep.* **17**, 514–526 (2016).

24. M. T. Maurano, R. Humbert, E. Rynes, R. E. Thurman, E. Haugen, H. Wang, A. P. Reynolds, R. Sandstrom, H. Qu, J. Brody, A. Shafer, F. Neri, K. Lee, T. Kutyavin, S. Stehling-Sun, A. K. Johnson, T. K. Canfield, E. Giste, M. Diegel, D. Bates, R. S. Hansen, S. Neph, P. J. Sabo, S. Heimfeld, A. Raubitschek, S. Ziegler, C. Cotsapas, N. Sotoodehnia, I. Glass, S. R. Sunyaev, R. Kaul, J. A. Stamatoyannopoulos, Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**, 1190–1195 (2012).

25. M. Kasowski, S. Kyriazopoulou-Panagiotopoulou, F. Grubert, J. B. Zaugg, A. Kundaje, Y. Liu, A. P. Boyle, Q. C. Zhang, F. Zakharia, D. V. Spacek, J. Li, D. Xie, A. Olarerin-George, L. M. Steinmetz, J. B. Hogenesch, M. Kellis, S. Batzoglou, M. Snyder, Extensive Variation in Chromatin States Across Humans. *Science* **342**, 750–752 (2013).

26. J. D. Buenrostro, B. Wu, H. Y. Chang, W. J. Greenleaf, ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109** (2015).

27. H. S. Kaya-Okur, S. J. Wu, C. A. Codomo, E. S. Pledger, T. D. Bryson, J. G. Henikoff, K. Ahmad, S. Henikoff, CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930 (2019).

28. J. Ernst, M. Kellis, ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).

29. K. A. Overmyer, C. R. Evans, N. R. Qi, C. E. Minogue, J. J. Carson, C. J. Chermside-Scabbo, L. G. Koch, S. L. Britton, D. J. Pagliarini, J. J. Coon, C. F. Burant, Maximal Oxidative Capacity during Exercise Is Associated with Skeletal Muscle Fuel Selection and Dynamic Changes in Mitochondrial Protein Acetylation. *Cell Metab.* **21**, 468–478 (2015).

30. R. Kivelä, M. Silvennoinen, M. Lehti, R. Rinnankoski-Tuikka, T. Purhonen, T. Ketola, K. Pullinen, M. Vuento, N. Mutanen, M. A. Sartor, H. Reunanen, L. G. Koch, S. L. Britton, H. Kainulainen, Gene expression centroids that link with low intrinsic aerobic exercise capacity and complex disease risk. *FASEB J.* **24**, 4565–4574 (2010).

31. R. A. Howlett, S. D. Kirkton, N. C. Gonzalez, H. E. Wagner, S. L. Britton, L. G. Koch, P. D. Wagner, Peripheral oxygen transport and utilization in rats following continued selective breeding for endurance running capacity. *J. Appl. Physiol.* **106**, 1819–1825 (2009).

32. N. C. Gonzalez, R. A. Howlett, K. K. Henderson, L. G. Koch, S. L. Britton, H. E. Wagner, F. Favret, P. D. Wagner, Systemic oxygen transport in rats artificially selected for running endurance. *Respir. Physiol. Neurobiol.* **151**, 141–150 (2006).

33. R. A. Howlett, N. C. Gonzalez, H. E. Wagner, Z. Fu, S. L. Britton, L. G. Koch, P. D. Wagner, Selected Contribution: Skeletal muscle capillarity and enzyme activity in rats selectively bred for running endurance. *J. Appl. Physiol.* **94**, 1682–1688 (2003).

34. M. A. Aon, S. Cortassa, M. Juhaszova, J. A. González-Reyes, M. Calvo-Rubio, J. M. Villalba, A. D. Lachance, B. D. Ziman, S. J. Mitchell, K. N. Murt, J. E. C. Axsom, I. Alfaras, S. L. Britton, L. G. Koch, R. de Cabo, E. G. Lakatta, S. J. Sollott, Mitochondrial health is enhanced in rats with higher vs. lower intrinsic exercise capacity and extended lifespan. *NPJ Aging Mech. Dis.* **7**, 1 (2021).

35. E. Heyne, S. Zeeb, C. Junker, A. Petzinna, A. Schrepper, T. Doenst, L. G. Koch, S. L. Britton, M. Schwarzer, Exercise Training Differentially Affects Skeletal Muscle Mitochondria in Rats with Inherited High or Low Exercise Capacity. *Cells* **13**, 393 (2024).

36. A. J. Pinto, A. Bergouignan, P. C. Dempsey, H. Roschel, N. Owen, B. Gualano, D. W. Dunstan, Physiology of sedentary behavior. *Physiol. Rev.* **103**, 2561–2622 (2023).

37. S. Fishilevich, R. Nudel, N. Rappaport, R. Hadar, I. Plaschkes, T. Iny Stein, N. Rosen, A. Kohn, M. Twik, M. Safran, D. Lancet, D. Cohen, GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database J. Biol. Databases Curation* **2017**, bax028 (2017).

38. P. Ying, C. Chen, Z. Lu, S. Chen, M. Zhang, Y. Cai, F. Zhang, J. Huang, L. Fan, C. Ning, Y. Li, W. Wang, H. Geng, Y. Liu, W. Tian, Z. Yang, J. Liu, C. Huang, X. Yang, B. Xu, H. Li, X. Zhu, N. Li, B. Li, Y. Wei, Y. Zhu, J. Tian, X. Miao, Genome-wide enhancer-gene regulatory maps link causal variants to target genes underlying human cancer risk. *Nat. Commun.* **14**, 5958 (2023).

39. A. Sanyal, B. R. Lajoie, G. Jain, J. Dekker, The long-range interaction landscape of gene promoters. *Nature* **489**, 109–113 (2012).

40. C. P. Fulco, M. Munschauer, R. Anyoha, G. Munson, S. R. Grossman, E. M. Perez, M. Kane, B. Cleary, E. S. Lander, J. M. Engreitz, Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* **354**, 769–773 (2016).

41. J. Nasser, D. T. Bergman, C. P. Fulco, P. Guckelberger, B. R. Doughty, T. A. Patwardhan, T. R. Jones, T. H. Nguyen, J. C. Ulirsch, F. Lekschas, K. Mualim, H. M. Natri, E. M. Weeks, G. Munson, M. Kane, H. Y. Kang, A. Cui, J. P. Ray, T. M. Eisenhaure, R. L. Collins, K. Dey, H. Pfister, A. L. Price, C. B. Epstein, A. Kundaje, R. J. Xavier, M. J. Daly, H. Huang, H. K. Finucane, N. Hacohen, E. S. Lander, J. M. Engreitz, Genome-wide enhancer maps link risk variants to disease genes. *Nature* **593**, 238–243 (2021).

42. I. J. Miinalainen, W. Schmitz, A. Huotari, K. J. Autio, R. Soininen, E. Ver Loren van Themaat, M. Baes, K.-H. Herzig, E. Conzelmann, J. K. Hiltunen, Mitochondrial 2,4-dienoyl-CoA reductase deficiency in mice results in severe hypoglycemia with stress intolerance and unimpaired ketogenesis. *PLoS Genet.* **5**, e1000543 (2009).

43. V. Snetkova, J. A. Skok, Enhancer talk. *Epigenomics* **10**, 483–498 (2018).

44. R. Ren, Y. Fan, Z. Peng, S. Wang, Y. Jiang, L. Fu, J. Cao, S. Zhao, H. Wang, Characterization and perturbation of CTCF-mediated chromatin interactions for enhancing myogenic transdifferentiation. *Cell Rep.* **40**, 111206 (2022).

45. Y. Liu, X. Wan, H. Li, Y. Chen, X. Hu, H. Chen, D. Zhu, C. Li, Y. Zhang, CTCF coordinates cell fate specification via orchestrating regulatory hubs with pioneer transcription factors. *Cell Rep.* **42**, 113259 (2023).

46. W. Zhang, H. Tong, Z. Zhang, S. Shao, D. Liu, S. Li, Y. Yan, Transcription factor EGR1 promotes differentiation of bovine skeletal muscle satellite cells by regulating *MyoG* gene expression. *J. Cell. Physiol.* **233**, 350–362 (2018).

47. M. Xu, X. Chen, D. Chen, B. Yu, Z. Huang, FoxO1: a novel insight into its molecular mechanisms in the regulation of skeletal muscle differentiation and fiber type specification. *Oncotarget* **8**, 10662–10674 (2017).

48. A. M. J. Sanchez, R. B. Candau, H. Bernardi, FoxO transcription factors: their roles in the maintenance of skeletal muscle homeostasis. *Cell. Mol. Life Sci.* **71**, 1657–1671 (2014).

49. P. S. Zammit, Function of the myogenic regulatory factors Myf5, MyoD, Myogenin and MRF4 in skeletal muscle, satellite cells and regenerative myogenesis. *Semin. Cell Dev. Biol.* **72**, 19–32 (2017).

50. S. Ustanina, J. Carvajal, P. Rigby, T. Braun, The Myogenic Factor Myf5 Supports Efficient Skeletal Muscle Regeneration by Enabling Transient Myoblast Amplification. *Stem Cells* **25**, 2006–2016 (2007).

51. J. A. Villena, I. Martin, O. Viñas, B. Cormand, R. Iglesias, T. Mampel, M. Giralt, F. Villarroya, ETS transcription factors regulate the expression of the gene for the human mitochondrial ATP synthase beta-subunit. *J. Biol. Chem.* **269**, 32649–32654 (1994).

52. L. Fan, D. R. Sweet, D. A. Prosdocimo, V. Vinayachandran, E. R. Chan, R. Zhang, O. Ilkayeva, Y. Lu, K. S. Keerthy, C. E. Booth, C. B. Newgard, M. K. Jain, Muscle Krüppel-like factor 15 regulates lipid flux and systemic metabolic homeostasis. *J. Clin. Invest.* **131**, e139496 (2021).

53. X. Liao, R. Zhang, Y. Lu, D. A. Prosdocimo, P. Sangwung, L. Zhang, G. Zhou, P. Anand, L. Lai, T. C. Leone, H. Fujioka, F. Ye, M. G. Rosca, C. L. Hoppel, P. C. Schulze, E. D. Abel, J. S. Stamler, D. P. Kelly, M. K. Jain, Kruppel-like factor 4 is critical for transcriptional control of cardiac mitochondrial homeostasis. *J. Clin. Invest.* **125**, 3461–3476 (2015).

54. B. C. Wilson, L. Boehme, A. Annibali, A. Hodgkinson, T. S. Carroll, R. J. Oakey, V. C. Seitan, Intellectual disability-associated factor Zbtb11 cooperates with NRF-2/GABP to control mitochondrial function. *Nat. Commun.* **11**, 5469 (2020).

55. Z. Cheng, FoxO transcription factors in mitochondrial homeostasis. *Biochem. J.* **479**, 525–536 (2022).

56. S. Peng, W. Li, N. Hou, N. Huang, A Review of FoxO1-Regulated Metabolic Diseases and Related Drug Discoveries. *Cells* **9**, 184 (2020).

57. M. He, Z. Pei, A.-W. Mohsen, P. Watkins, G. Murdoch, P. P. Van Veldhoven, R. Ensenauer, J. Vockley, Identification and characterization of new long chain Acyl-CoA dehydrogenases. *Mol. Genet. Metab.* **102**, 418–429 (2011).

58. D. L. Nicolae, E. Gamazon, W. Zhang, S. Duan, M. E. Dolan, N. J. Cox, Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).

59. M. Oh, S. Jeong, K. Kang, S.-K. Kim, Enhancer Trajectories in Lineage Commitment: Regulatory Logic of States and Cooperation. *Biomolecules* **16**, 87 (2026).

60. J. P. Bothma, H. G. Garcia, S. Ng, M. W. Perry, T. Gregor, M. Levine, Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo. *eLife* **4**, e07956 (2015).

61. O. Corradin, A. Saiakhova, B. Akhtar-Zaidi, L. Myeroff, J. Willis, R. Cowper-Sal lari, M. Lupien, S. Markowitz, P. C. Scacheri, Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. *Genome Res.* **24**, 1–13 (2014).

62. E. Z. Kvon, R. Waymack, M. Gad, Z. Wunderlich, Enhancer redundancy in development and disease. *Nat. Rev. Genet.* **22**, 324–336 (2021).

63. Z. Wu, P. Puigserver, U. Andersson, C. Zhang, G. Adelmant, V. Mootha, A. Troy, S. Cinti, B. Lowell, R. C. Scarpulla, B. M. Spiegelman, Mechanisms controlling mitochondrial biogenesis and respiration through the thermogenic coactivator PGC-1. *Cell* **98**, 115–124 (1999).

64. R. C. Scarpulla, Transcriptional paradigms in mammalian mitochondrial biogenesis and function. *Physiol. Rev.* **88**, 611–638 (2008).

65. R. E. Gate, C. S. Cheng, A. P. Aiden, A. Siba, M. Tabaka, D. Lituiev, I. Machol, M. G. Gordon, M. Subramaniam, M. Shamim, K. L. Hougen, I. Wortman, S.-C. Huang, N. C. Durand, T. Feng, P. L. De Jager, H. Y. Chang, E. L. Aiden, C. Benoist, M. A. Beer, C. J. Ye, A. Regev, Genetic determinants of co-accessible chromatin regions in activated T cells across humans. *Nat. Genet.* **50**, 1140–1150 (2018).

66. B. Deplancke, D. Alpern, V. Gardeux, The Genetics of Transcription Factor DNA Binding Variation. *Cell* **166**, 538–554 (2016).

67. H. Kilpinen, S. M. Waszak, A. R. Gschwind, S. K. Raghav, R. M. Witwicki, A. Orioli, E. Migliavacca, M. Wiederkehr, M. Gutierrez-Arcelus, N. I. Panousis, A. Yurovsky, T. Lappalainen, L. Romano-Palumbo, A. Planchon, D. Bielser, J. Bryois, I. Padioleau, G. Udin, S. Thurnheer, D. Hacker, L. J. Core, J. T. Lis, N. Hernandez, A. Reymond, B. Deplancke, E. T. Dermitzakis, Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* **342**, 744–747 (2013).

68. S. M. Waszak, O. Delaneau, A. R. Gschwind, H. Kilpinen, S. K. Raghav, R. M. Witwicki, A. Orioli, M. Wiederkehr, N. I. Panousis, A. Yurovsky, L. Romano-Palumbo, A. Planchon, D. Bielser, I. Padioleau, G. Udin, S. Thurnheer, D. Hacker, N. Hernandez, A. Reymond, B. Deplancke, E. T. Dermitzakis, Population Variation and Genetic Control of Modular Chromatin Architecture in Humans. *Cell* **162**, 1039–1050 (2015).

69. F. Grubert, J. B. Zaugg, M. Kasowski, O. Ursu, D. V. Spacek, A. R. Martin, P. Greenside, R. Srivas, D. H. Phanstiel, A. Pekowska, N. Heidari, G. Euskirchen, W. Huber, J. K. Pritchard, C. D. Bustamante, L. M. Steinmetz, A. Kundaje, M. Snyder, Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. *Cell* **162**, 1051–1065 (2015).

70. R. Siersbæk, A. Rabiee, R. Nielsen, S. Sidoli, S. Traynor, A. Loft, L. L. C. Poulsen, A. Rogowska-Wrzesinska, O. N. Jensen, S. Mandrup, Transcription factor cooperativity in early adipogenic hotspots and super-enhancers. *Cell Rep.* **7**, 1443–1455 (2014).

71. A. Ferdous, S. Singh, Y. Luo, M. J. Abedin, N. Jiang, C. E. Perry, B. M. Evers, T. G. Gillette, M. Kyba, M. Trojanowska, J. A. Hill, Fli1 Promotes Vascular Morphogenesis by Regulating Endothelial Potential of Multipotent Myogenic Progenitors. *Circ. Res.* **129**, 949–964 (2021).

72. L. De Angelis, L. Berghella, M. Coletta, L. Lattanzi, M. Zanchi, M. G. Cusella-De Angelis, C. Ponzetto, G. Cossu, Skeletal myogenic progenitors originating from embryonic dorsal aorta coexpress endothelial and myogenic markers and contribute to postnatal muscle growth and regeneration. *J. Cell Biol.* **147**, 869–878 (1999).

73. D. Chen, X. Fan, N. Sun, K. Wang, L. Gong, J. M. Melero-Martin, W. T. Pu, Pioneer factor ETV2 safeguards endothelial cell specification by recruiting the repressor REST to restrict alternative lineage commitment. *Nat. Cardiovasc. Res.* **4**, 689–709 (2025).

74. C. Handschin, Y. M. Kobayashi, S. Chin, P. Seale, K. P. Campbell, B. M. Spiegelman, PGC-1alpha regulates the neuromuscular junction program and ameliorates Duchenne muscular dystrophy. *Genes Dev.* **21**, 770–783 (2007).

75. R. W. A. Souza, C. R. R. Alves, A. Medeiros, N. Rolim, G. J. J. Silva, J. B. N. Moreira, M. N. Alves, M. Wohlwend, M. Gebriel, L. Hagen, A. Sharma, L. G. Koch, S. L. Britton, G. Slupphaug, U. Wisløff, P. C. Brum, Differential regulation of cysteine oxidative post-translational modifications in high and low aerobic capacity. *Sci. Rep.* **8**, 17772 (2018).

76. E. L. Seifert, M. Bastianelli, C. Aguer, C. Moffat, C. Estey, L. G. Koch, S. L. Britton, M.-E. Harper, Intrinsic aerobic capacity correlates with greater inherent mitochondrial oxidative and H_2 O_2 emission capacities without major shifts in myosin heavy chain isoform. *J. Appl. Physiol.* **113**, 1624–1634 (2012).

77. D. A. Rivas, S. J. Lessard, M. Saito, A. M. Friedhuber, L. G. Koch, S. L. Britton, B. B. Yaspelkis, J. A. Hawley, Low intrinsic running capacity is associated with reduced skeletal muscle substrate oxidation and lower mitochondrial content in white skeletal muscle. *Am. J. Physiol.-Regul. Integr. Comp. Physiol.* **300**, R835–R843 (2011).

78. R. S. Rogers, E. M. Morris, J. L. Wheatley, A. E. Archer, C. S. McCoin, K. S. White, D. R. Wilson, G. M. E. Meers, L. G. Koch, S. L. Britton, J. P. Thyfault, P. C. Geiger, Deficiency in the Heat Stress Response Could Underlie Susceptibility to Metabolic Disease. *Diabetes* **65**, 3341–3351 (2016).

79. J. V. Virbasius, C. A. Virbasius, R. C. Scarpulla, Identity of GABP with NRF-2, a multisubunit activator of cytochrome oxidase expression, reveals a cellular role for an ETS domain activator of viral promoters. *Genes Dev.* **7**, 380–392 (1993).

80. J. Zhang, F. E. Frerman, J.-J. P. Kim, Structure of electron transfer flavoprotein-ubiquinone oxidoreductase and electron transfer to the mitochondrial ubiquinone pool. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 16212–16217 (2006).

81. S. Paquay, J. Duraffourd, M. Bury, I. P. Heremans, F. Caligiore, I. Gerin, V. Stroobant, J. Jacobs, A. Pinon, J. Graff, D. Vertommen, E. Van Schaftingen, J. P. Dewulf, G. T. Bommer, ACAD10 and ACAD11 allow entry of 4-hydroxy fatty acids into β-oxidation. *Cell. Mol. Life Sci. CMLS* **81**, 367 (2024).

82. E. H. Rashan, A. K. Bartlett, D. B. Khana, J. Zhang, R. Jain, G. Wade, L. A. Abriata, A. J. Smith, Z. N. Baker, T. Cook, A. Caldwell, A. R. Chevalier, P. Forny, B. F. Pfleger, M. D. Peraro, P. Yuan, D. Amador-

Noguez, J. A. Simcox, D. J. Pagliarini, ACAD10 and ACAD11 enable mammalian 4-hydroxy acid lipid catabolism. *Nat. Struct. Mol. Biol.*, doi: 10.1038/s41594-025-01596-4 (2025).

83. K. F. Petersen, S. Dufour, D. Befroy, R. Garcia, G. I. Shulman, Impaired mitochondrial activity in the insulin-resistant offspring of patients with type 2 diabetes. *N. Engl. J. Med.* **350**, 664–671 (2004).

84. S. J. Mihalik, B. H. Goodpaster, D. E. Kelley, D. H. Chace, J. Vockley, F. G. S. Toledo, J. P. DeLany, Increased levels of plasma acylcarnitines in obesity and type 2 diabetes and identification of a marker of glucolipotoxicity. *Obes. Silver Spring Md* **18**, 1695–1700 (2010).

85. T. R. Koves, J. R. Ussher, R. C. Noland, D. Slentz, M. Mosedale, O. Ilkayeva, J. Bain, R. Stevens, J. R. B. Dyck, C. B. Newgard, G. D. Lopaschuk, D. M. Muoio, Mitochondrial overload and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance. *Cell Metab.* **7**, 45–56 (2008).

86. S. A. Summers, B. H. Goodpaster, CrossTalk proposal: Intramyocellular ceramide accumulation does modulate insulin resistance. *J. Physiol.* **594**, 3167–3170 (2016).

87. M. N. Davies, L. Kjalarsdottir, J. W. Thompson, L. G. Dubois, R. D. Stevens, O. R. Ilkayeva, M. J. Brosnan, T. P. Rolph, P. A. Grimsrud, D. M. Muoio, The Acetyl Group Buffering Action of Carnitine Acetyltransferase Offsets Macronutrient-Induced Lysine Acetylation of Mitochondrial Proteins. *Cell Rep.* **14**, 243–254 (2016).

88. S. S. Korshunov, V. P. Skulachev, A. A. Starkov, High protonic potential actuates a mechanism of production of reactive oxygen species in mitochondria. *FEBS Lett.* **416**, 15–18 (1997).

89. J. N. Timmis, M. A. Ayliffe, C. Y. Huang, W. Martin, Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **5**, 123–135 (2004).

90. Z. Swigonová, A.-W. Mohsen, J. Vockley, Acyl-CoA dehydrogenases: Dynamic history of protein family evolution. *J. Mol. Evol.* **69**, 176–193 (2009).

91. A. Bye, M. Langaas, M. A. Høydal, O. J. Kemi, G. Heinrich, L. G. Koch, S. L. Britton, S. M. Najjar, Ø. Ellingsen, U. Wisløff, Aerobic capacity-dependent differences in cardiac gene expression. *Physiol. Genomics* **33**, 100–109 (2008).

92. J. Y. Fleischman, N. R. Qi, M. K. Treutelaar, S. L. Britton, L. G. Koch, J. Z. Li, C. F. Burant, Intrinsic cardiorespiratory fitness modulates clinical and molecular response to caloric restriction. *Mol. Metab.* **68**, 101668 (2023).

93. K. Bowden-Davies, J. Connolly, P. Burghardt, L. G. Koch, S. L. Britton, J. G. Burniston, Label-free profiling of white adipose tissue of rats exhibiting high or low levels of intrinsic exercise capacity. *Proteomics* **15**, 2342–2349 (2015).

94. N. T. Kamarulzaman, S. Makpol, The link between Mitochondria and Sarcopenia. *J. Physiol. Biochem.* **81**, 1–20 (2025).

95. G. P. Holloway, A. Bonen, L. L. Spriet, Regulation of skeletal muscle mitochondrial fatty acid metabolism in lean and obese individuals. *Am. J. Clin. Nutr.* **89**, 455S-462S (2009).

96. D. Serra, P. Mera, M. I. Malandrino, J. F. Mir, L. Herrero, Mitochondrial Fatty Acid Oxidation in Obesity. *Antioxid. Redox Signal.* **19**, 269–284 (2013).

97. P. Orchard, N. Manickam, C. Ventresca, S. Vadlamudi, A. Varshney, V. Rai, J. Kaplan, C. Lalancette, K. L. Mohlke, K. Gallagher, C. F. Burant, S. C. J. Parker, Human and rat skeletal muscle single-nuclei multi-omic integrative analyses nominate causal cell types, regulatory elements, and SNPs for complex traits. *Genome Res.* **31**, 2258–2275 (2021).

98. S. Henikoff, J. Henikoff, K. Ahmad, CUT&Tag-direct with CUTAC v3. [Preprint] (2020). https://doi.org/10.17504/protocols.io.bqwvmxe6.

99. H. Wickham, M. Averick, J. Bryan, W. Chang, L. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T. Pedersen, E. Miller, S. Bache, K. Müller, J. Ooms, D. Robinson, D. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani, Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).

100. M. Lawrence, W. Huber, H. Pagès, P. Aboyoun, M. Carlson, R. Gentleman, M. T. Morgan, V. J. Carey, Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).

101. S. Lee, D. Cook, M. Lawrence, plyranges: a grammar of genomic data transformation. *Genome Biol.* **20**, 4 (2019).

102. Z. Gu, R. Eils, M. Schlesner, Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

103. Z. Gu, Complex heatmap visualization. *iMeta* **1**, e43 (2022).

104. F. Mölder, K. P. Jablonski, B. Letcher, M. B. Hall, C. H. Tomkins-Tinch, V. Sochat, J. Forster, S. Lee, S. O. Twardziok, A. Kanitz, A. Wilm, M. Holtgrewe, S. Rahmann, S. Nahnsen, J. Köster, Sustainable data analysis with Snakemake. *F1000Research* **10**, 33 (2021).

105. F. J. Martin, M. R. Amode, A. Aneja, O. Austine-Orimoloye, A. G. Azov, I. Barnes, A. Becker, R. Bennett, A. Berry, J. Bhai, S. K. Bhurji, A. Bignell, S. Boddu, P. R. Branco Lins, L. Brooks, S. B. Ramaraju, M. Charkhchi, A. Cockburn, L. Da Rin Fiorretto, C. Davidson, K. Dodiya, S. Donaldson, B. El Houdaigui, T. El Naboulsi, R. Fatima, C. G. Giron, T. Genez, G. S. Ghattaoraya, J. G. Martinez, C. Guijarro, M. Hardy, Z. Hollis, T. Hourlier, T. Hunt, M. Kay, V. Kaykala, T. Le, D. Lemos, D. Marques-Coelho, J. C. Marugán, G. A. Merino, L. P. Mirabueno, A. Mushtaq, S. N. Hossain, D. N. Ogeh, M. P. Sakthivel, A. Parker, M. Perry, I. Piližota, I. Prosovetskaia, J. G. Pérez-Silva, A. I. A. Salam, N. Saraiva-Agostinho, H. Schuilenburg, D. Sheppard, S. Sinha, B. Sipos, W. Stark, E. Steed, R. Sukumaran, D. Sumathipala, M.-M. Suner, L. Surapaneni, K. Sutinen, M. Szpak, F. F. Tricomi, D. Urbina-Gómez, A. Veidenberg, T. A. Walsh, B. Walts, E. Wass, N. Willhoft, J. Allen, J. Alvarez-Jarreta, M. Chakiachvili, B. Flint, S. Giorgetti, L. Haggerty, G. R. Ilsley, J. E. Loveland, B. Moore, J. M. Mudge, J. Tate, D. Thybert, S. J. Trevanion, A. Winterbottom, A. Frankish, S. E. Hunt, M. Ruffier, F. Cunningham, S. Dyer, R. D. Finn, K. L. Howe, P. W. Harrison, A. D. Yates, P. Flicek, Ensembl 2023. *Nucleic Acids Res.* **51**, D933–D941 (2023).

106. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

107. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

108. H. Li, Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* (2013).

109. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

110. Picard Tools - By Broad Institute. http://broadinstitute.github.io/picard/.

111. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

112. Y. Zhang, T. Liu, C. A. Meyer, J. Eeckhoute, D. S. Johnson, B. E. Bernstein, C. Nusbaum, R. M. Myers, M. Brown, W. Li, X. S. Liu, Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).

113. P. Orchard, Y. Kyono, J. Hensley, J. O. Kitzman, S. C. J. Parker, Quantification, Dynamic Visualization, and Validation of Bias in ATAC-Seq Data with ataqv. *Cell Syst.* **10**, 298-306.e4 (2020).

114. A. N. Schep, B. Wu, J. D. Buenrostro, W. J. Greenleaf, chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).

115. H. M. Amemiya, A. Kundaje, A. P. Boyle, The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 9354 (2019).

116. K. Broman, broman: Karl Broman's R Code, version 0.80 (2022); https://CRAN.R-project.org/package=broman.

117. C. Davide Risso [Aut, RUVSeq, Bioconductor (2017); https://doi.org/10.18129/B9.BIOC.RUVSEQ.

118. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

119. I. Dolgalev, msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format, version 7.5.1 (2022); https://CRAN.R-project.org/package=msigdbr.

120. S. Durinck, Y. Moreau, A. Kasprzyk, S. Davis, B. De Moor, A. Brazma, W. Huber, BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* **21**, 3439–3440 (2005).

121. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **57**, 289–300 (1995).

122. C. Lee, S. Patil, M. A. Sartor, RNA-Enrich: a cut-off free functional enrichment testing method for RNA-seq with improved detection power. *Bioinformatics* **32**, 1100–1102 (2016).

123. G. Korotkevich, V. Sukhov, N. Budin, B. Shpak, M. N. Artyomov, A. Sergushichev, Fast gene set enrichment analysis. [Preprint] (2016). https://doi.org/10.1101/060012.

124. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* **2**, 100141 (2021).

125. Y. Zhou, B. Zhou, L. Pache, M. Chang, A. H. Khodabakhshi, O. Tanaseichuk, C. Benner, S. K. Chanda, Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat. Commun.* **10**, 1523 (2019).

126. K. W. Currin, M. R. Erdos, N. Narisu, V. Rai, S. Vadlamudi, H. J. Perrin, J. R. Idol, T. Yan, R. D. Albanus, K. A. Broadaway, A. S. Etheridge, L. L. Bonnycastle, P. Orchard, J. P. Didion, A. S. Chaudhry, F. Innocenti, E. G. Schuetz, L. J. Scott, S. C. J. Parker, F. S. Collins, K. L. Mohlke, B. B. Barnabas, S. Black, G. G. Bouffard, S. Y. Brooks, H. Coleman, L. Dekhtyar, J. Han, S. Ho, J. Kim, R. Legaspi, Q. L. Maduro, C. A. Masiello, J. C. McDowell, C. Montemayor, J. C. Mullikin, M. Park, N. L. Riebow, K. Schandler, B. Schmidt, C. Sison, S. Stantripop, J. W. Thomas, P. J. Thomas, M. Vemulapalli, A. C. Young, Genetic effects on liver chromatin accessibility identify disease regulatory variants. *Am. J. Hum. Genet.* **108**, 1169–1189 (2021).

127. M. A. DePristo, E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire, C. Hartl, A. A. Philippakis, G. Del Angel, M. A. Rivas, M. Hanna, A. McKenna, T. J. Fennell, A. M. Kernytsky, A. Y. Sivachenko, K. Cibulskis, S. B. Gabriel, D. Altshuler, M. J. Daly, A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).

128. S. Ramdas, A. B. Ozel, M. K. Treutelaar, K. Holl, M. Mandel, L. C. S. Woods, J. Z. Li, Extended regions of suspected mis-assembly in the rat reference genome. *Sci. Data* **6**, 39 (2019).

129. H. Li, A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).

130. P. Danecek, J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, H. Li, Twelve years of SAMtools and BCFtools. *GigaScience* **10**, giab008 (2021).

131. T. V. De Jong, Y. Pan, P. Rastas, D. Munro, M. Tutaj, H. Akil, C. Benner, D. Chen, A. S. Chitre, W. Chow, V. Colonna, C. L. Dalgard, W. M. Demos, P. A. Doris, E. Garrison, A. M. Geurts, H. M. Gunturkun, V. Guryev, T. Hourlier, K. Howe, J. Huang, T. Kalbfleisch, P. Kim, L. Li, S. Mahaffey, F. J. Martin, P. Mohammadi, A. B. Ozel, O. Polesskaya, M. Pravenec, P. Prins, J. Sebat, J. R. Smith, L. C. S. Woods, B.

Tabakoff, A. Tracey, M. Uliano-Silva, F. Villani, H. Wang, B. M. Sharp, F. Telese, Z. Jiang, L. Saba, X. Wang, T. D. Murphy, A. A. Palmer, A. E. Kwitek, M. R. Dwinell, R. W. Williams, J. Z. Li, H. Chen, A revamped rat reference genome improves the discovery of genetic diversity in laboratory rats. [Preprint] (2023). https://doi.org/10.1101/2023.04.13.536694.

132. S. Rubinacci, O. Delaneau, J. Marchini, Genotype imputation using the Positional Burrows Wheeler Transform. *PLOS Genet.* **16**, e1009049 (2020).

133. R. J. Hofmeister, D. M. Ribeiro, S. Rubinacci, O. Delaneau, Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. *Nat. Genet.* **55**, 1243–1249 (2023).

134. R. Hermsen, J. De Ligt, W. Spee, F. Blokzijl, S. Schäfer, E. Adami, S. Boymans, S. Flink, R. Van Boxtel, R. H. Van Der Weide, T. Aitman, N. Hübner, M. Simonis, B. Tabakoff, V. Guryev, E. Cuppen, Genomic landscape of rat strain and substrain variation. *BMC Genomics* **16**, 357 (2015).

135. S. Wright, The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. *Evolution* **19**, 395 (1965).

136. M. R. Paul Shannon, MotifDb, Bioconductor (2017); https://doi.org/10.18129/B9.BIOC.MOTIFDB.

137. G. Tan, B. Lenhard, TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* **32**, 1555–1556 (2016).

138. C. E. Grant, T. L. Bailey, W. S. Noble, FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).

139. R. C. McLeay, T. L. Bailey, Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* **11**, 165 (2010).

140. A. Fort, N. I. Panousis, M. Garieri, S. E. Antonarakis, T. Lappalainen, E. T. Dermitzakis, O. Delaneau, *MBV*: a method to solve sample mislabeling and detect technical bias in large combined genotype and sequencing assay datasets. *Bioinformatics* **33**, 1895–1897 (2017).

141. X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* **44**, 821–824 (2012).

142. J. R. Davis, L. Fresard, D. A. Knowles, M. Pala, C. D. Bustamante, A. Battle, S. B. Montgomery, An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *Am. J. Hum. Genet.* **98**, 216–224 (2016).

Acknowledgements

This research was supported in part through computational resources and services provided by Advanced Research Computing (ARC), a division of Information and Technology Services (ITS) at the University of Michigan, Ann Arbor. Next-generation sequencing was carried out in the Advanced Genomics Core at the University of Michigan. Dr. Nathan Xi and the University of Michigan Animal Metabolic, Physiological, & Behavioral Phenotyping Core assisted in the animal studies. Johanna Y Fleischman, Gayatri R Iyer, Martin J Walsh, and Alexander J Weitzel for reviewing the manuscript and providing advice as it was being prepared.

Funding:

Funding for the collection of HCR/LCR skeletal muscle samples was provided by a pilot award to CRE from the Michigan Nutrition Obesity Research Center, supported by National Institutes of Health grant P30 DK089503

A. Alfred Taubman Medical Research Institute and National Institutes of Health grant U24 DK112342 (CFB)

National Institutes of Health grant R01 DK117960 (SCJP)

University of Michigan Systems and Integrative Biology Training Grant T32 GM008322 (AMW)

The LCR-HCR rat model was funded by National Institutes of Health Office of Research Infrastructure Programs grant P40OD-021331 (LGK, SLB)

Author contributions:

Conceptualization: AMW, CRE, JZL, SCJP, CFB

Methodology: AMW, PO, CRE, SLB, LGK, SCJP, CFB

Validation: AMW, PO, SCJP, CFB

Formal analysis: AMW, PO

Investigation: AMW, CRE, NM, MKT

Resources: SLB, LGK, JZL, CFB

Data curation: AMW, PO

Writing – original draft: AMW, SCJP, CFB

Writing – review & editing: all authors

Visualization: AMW

Supervision: PO, JZL, SCJP, CFB

Project Administration: MKT, CRE, SCJP, CFB

Funding Acquisition: JZL, SCJP, CFB

Competing interests: The authors declare no competing interests.

Data, code, and materials availability:

This study did not generate new unique reagents.

Raw (FASTQ files) and aligned (BAM files) sequencing data will be deposited on NCBI GEO upon final submission. The HCR/LCR posterior probability matrices, HCR/LCR and F2 count matrices, HCR/LCR and F2 imputed genotypes, and the F2 Affymetrix Axiom panel will be publicly available by uploading to appropriate database by the date of publication. Processed data and results referenced in this manuscript are available in a Zenodo repository (DOI https://doi.org/10.5281/zenodo.19699175).

All original code will be deposited at a Zenodo repository to be publicly available by the date of publication.

Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Supplementary Materials

Materials and Methods

Supplementary Text (Supplemental Protocol S1: Nuclei extraction for ATAC-Seq library preparation)

Figs. S1 to S12

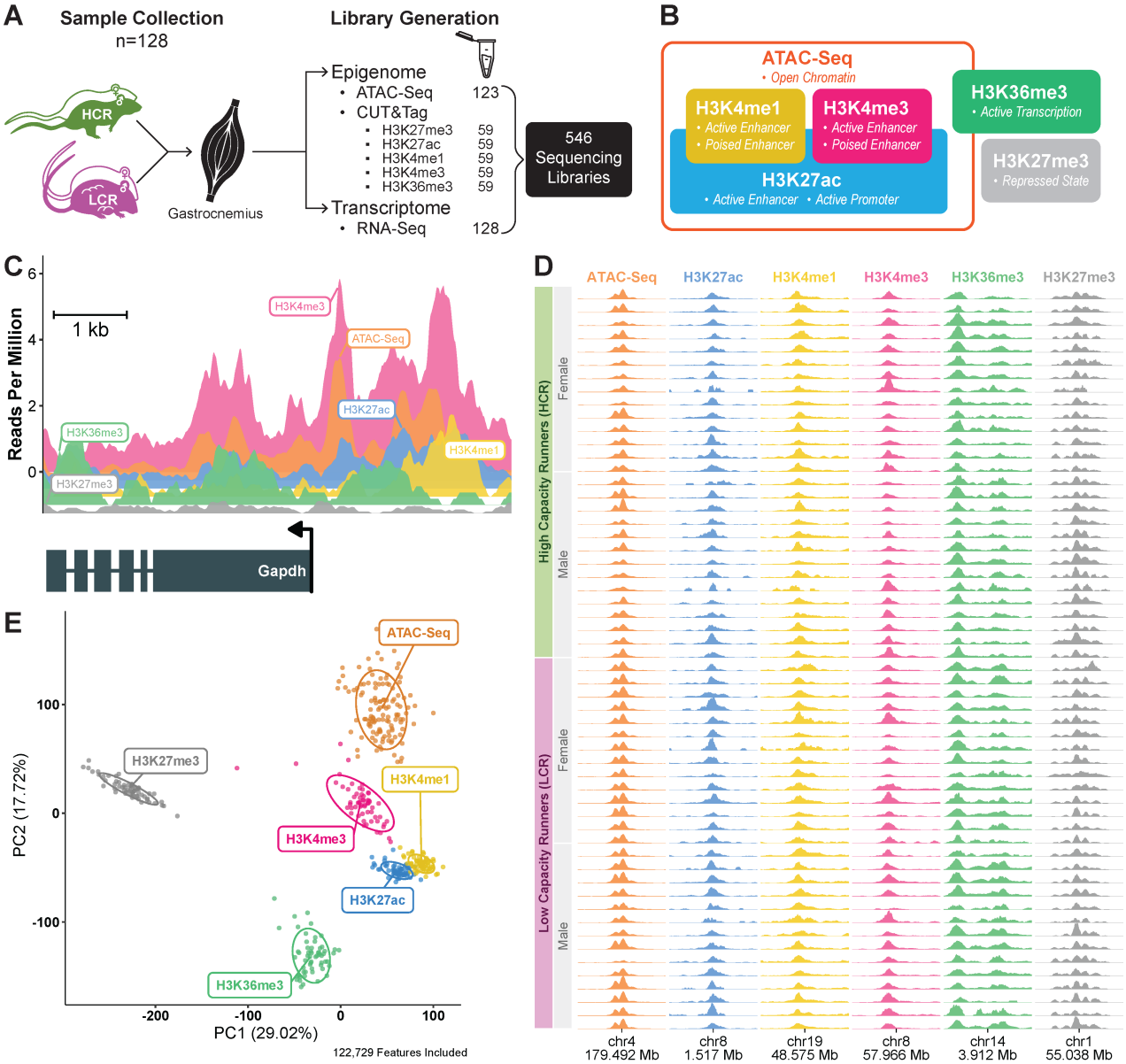


Fig. 1. Study Design and Data Overview

(**A**) Multi-omic profiling of gastrocnemius muscle from 128 rats (equal representation of HCR/LCR and sex), including RNA-Seq, ATAC-Seq, and CUT&Tag for five histone modifications, was carried out. See also Fig. S1A.

(**B**) Epigenomic mark key. H3K27ac, H3K4me1, and H3K4me3 are histone marks primarily located in open chromatin regions, most likely to coincide with ATAC-Seq signal. H3K27ac is associated with active enhancers and promoters, H3K4me1 has a strong presence in enhancers, H3K4me3 in promoters, H3K36me3 is present along the gene body of actively transcribed genes as it is laid down by polymerase II, and H3K27me3 is a repressive mark and associated with reduced transcription.

(**C**) Example signal tracks at the *Gapdh* locus in a single sample (rat R18804), showing from back to front, H3K4me3, ATAC-Seq, H3K27ac, H3K4me1, H3K36me3, and H3K27me3. *Gapdh* is transcribed on the negative strand.

(**D**) Signal reproducibility across the 56 samples with full epigenomic data (refer to Fig. S1A for details pertaining to library generation). Sequences within each modality are represented column-wise, where six unique genomic regions are displayed as the maximum RPM within 10bp bins, smoothed using a rolling mean (bin size = 50). Each column represents a 4 kb genomic window centered around the midpoint annotated on the x-axis. Y-axis ranges are fixed within each column with a minimum value of zero and modality-specific maxima: ATAC-Seq

(9.9 RPM), H3K27ac (5 RPM), H3K4me1 (10.3 RPM), H3K4me3 (15.4 RPM), H3K36me3 (6.6 RPM), and H3K27me3 (6.1 RPM). The presence and abundance of signal across genomic bins are consistent within modality.

(**E**) PCA of genome-wide signal within 10 kb bins (inverse rank normalized, blacklist filtered). Sequencing libraries cluster by modality, demonstrating reproducibility across samples.

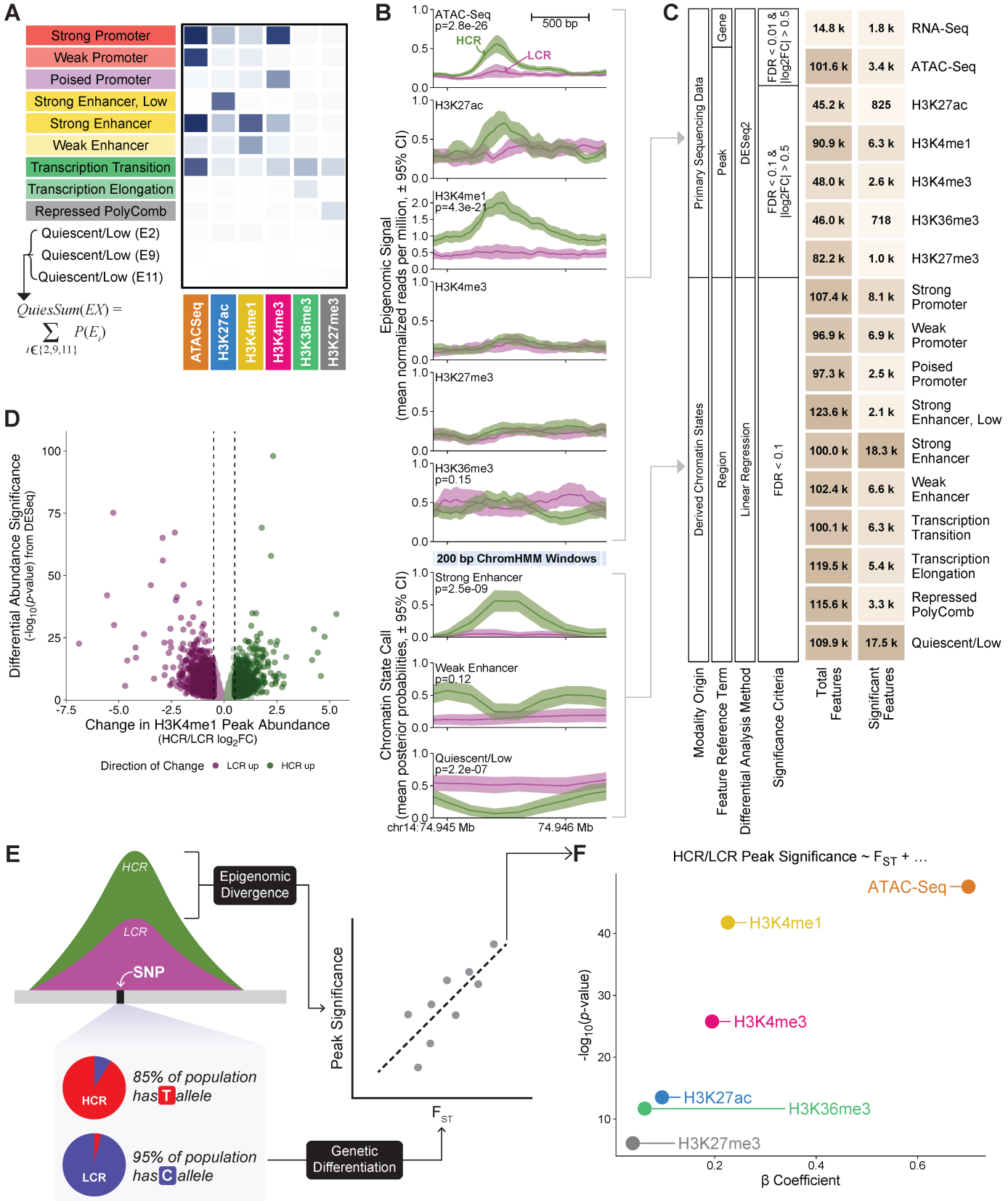


Fig. 2. Global Epigenomic Differences Between HCR and LCR

(**A**) ChromHMM emission plot showing chromatin state definitions where darker blue indicates a greater probability of mark presence in the emission state. The sum of three low signal states was used to merge into one representative, "Quiescent/Low", state for analysis.

(**B**) Example locus on chr14 showing normalized ATAC-Seq and CUT&Tag counts (reads per million) and ChromHMM posterior probabilities across 59 HCR (green) and LCR (purple) rats. For each sample, ATAC-Seq

and CUT&Tag signal was smoothed using a rolling mean applied to the maximum signal within 10 bp bins (bin size = 5), and values were summarized by plotting the mean ± 95% CI across samples. Posterior probabilities (mean ± 95% CI) are plotted at the midpoint of each 200 bp ChromHMM window, which represent the likelihood of the chromatin state occurring at these loci. Increased ATAC-Seq and H3K4me1 signal in HCR coincides with a rise in Strong Enhancer and dip in Quiescent/Low state probability, indicating a clear chromatin state difference between HCR and LCR. *P*-values from differential analyses are printed on each panel (not shown if not tested). Notably, ATAC-Seq libraries were prepared independently from CUT&Tag libraries and show highly concordant signals across ATAC-Seq and histone marks associated with open and active chromatin, further supporting the reproducibility and validity of these data.

(**C**) Summary of each data modality, including modality-specific terminology applied within this study, statistical thresholds, and the number of processed and significantly different HCR/LCR features. Significance thresholds were selected based on statistical power, with a more stringent cutoff applied to RNA-Seq (n = 128) and ATAC-Seq (n = 123), which had larger sample sizes than CUT&Tag (n = 59) and ChromHMM (n = 56). A $\log_2$FC threshold was included for CUT&Tag to compensate for the more lenient FDR, and for other DESeq-based analyses to maintain consistency. A magnitude of change threshold was not applied to ChromHMM-derived states due to the inverse normal transformation applied to posterior probabilities used in the linear model, as described in the Methods.

(**D**) Volcano plot of H3K4me1 peaks, highlighting significantly different features between HCR (green) and LCR (magenta). Significant features are highlighted as a darker color. See also Fig. S6.

(**E**) Schematic showing comparison of population differentiation (F_{ST}) and HCR/LCR epigenomic divergence carried out for each modality. F_{ST} is a standard metric for detecting loci under selection that ranges from 0 (no divergence) to 1 (complete fixation between HCR/LCR lines).

(**F**) Association between F_{ST} and HCR/LCR epigenomic peak significance. Linear models return a strong positive relationship between inverse rank normalized F_{ST} and HCR/LCR DESeq -$\log_{10}$(*p*-values) for each modality, while controlling for technical variables (GC content, FPKM) and genomic location (coding vs noncoding). The -$\log_{10}$(*p*-value) for the F_{ST} term in the model (y-axis) is plotted against the β coefficient (x-axis) and points are colored by modality.

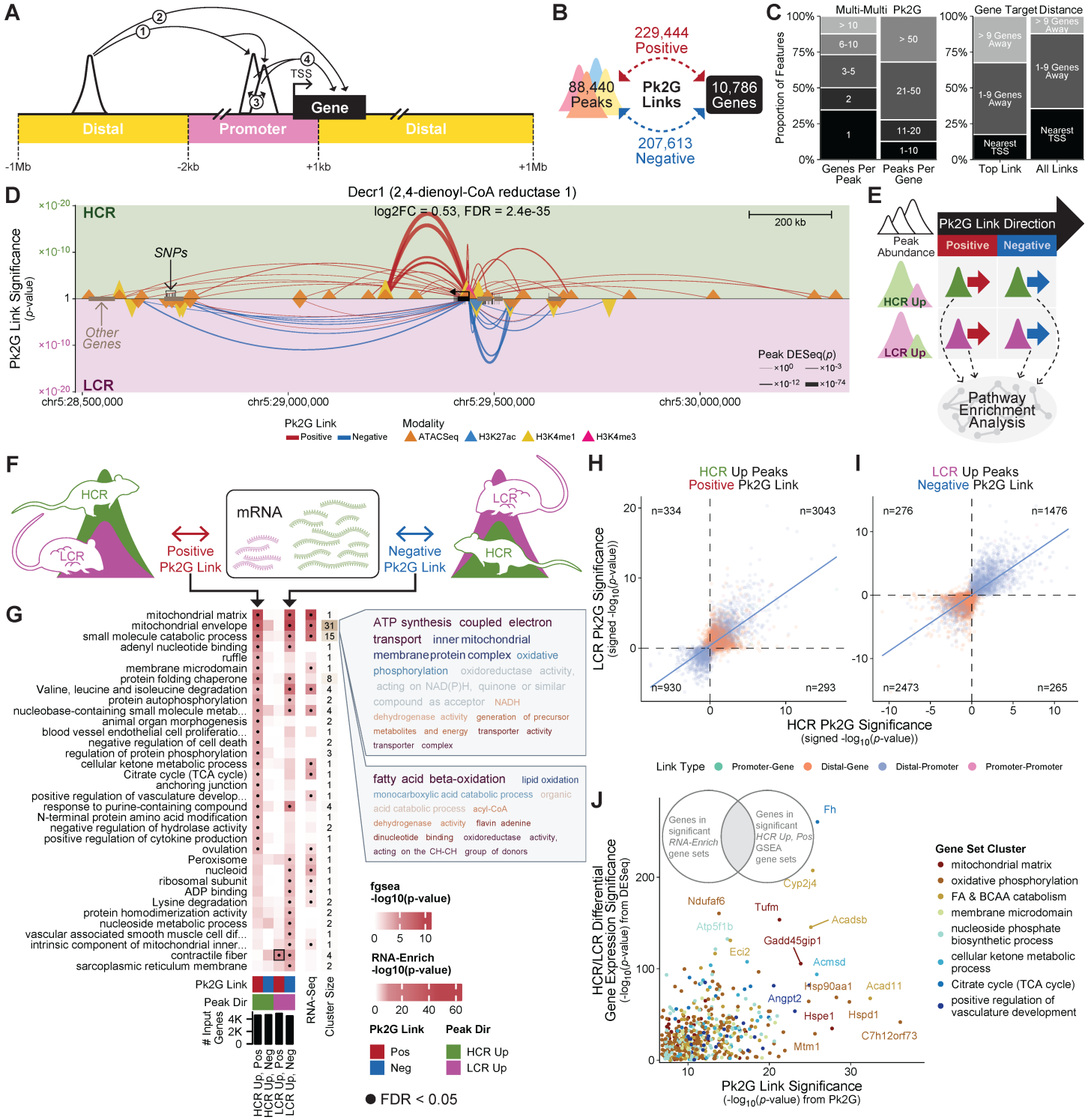


Fig. 3. Linking Epigenomic Features to Gene Expression

(**A**) Correlation framework mapping epigenomic peaks (ATAC-Seq, H3K27ac, H3K4me1, and H3K4me3) to gene expression. Distal regulatory elements (yellow, TSS ± 1 Mb) are correlated with (1) promoter peaks (pink, TSS -2 kb, TSS +1 kb) and (2) gene expression. Promoter peaks are correlated with (3) one another and (4) gene expression.

(**B**) Summary of the number of peaks and genes involved in the significant Pk2G links (FDR < 0.05) reflecting positive and negative peak-gene relationships.

(**C**) Pk2G link distribution and target location. Proportion of features within binned categories of genes linked per peak and peaks annotated per gene (Multi-Multi Pk2G), and genomic distance from peaks to their linked gene targets, relative to the nearest TSS (Gene Target Distance). Shading intensity reflects bin order.

(**D**) Pk2G links associated with a single gene, *Decr1* ($\log_2$FC = 0.53, FDR = 2.4×10^{-35}). Regulatory elements (triangles, colored by modality) are positioned by genomic distance from the TSS (arrow) of the gene (black rectangle). Peaks with DESeq $\log_2$FC > 0 (HCR up) are located above the midline, and LCR up peaks are shown below. Arc height denotes the significance of the Pk2G links, colored by direction of Pk2G link, with linewidth proportional to DESeq significance of the regulatory element. Vertical ticks at the midline represent SNPs with F_{ST} > 0.3 and additional rectangles represent other genes within view. See additional examples in Fig. S8.

(**E**) Schematic of four independent pathway enrichment analyses of activating and repressive regulatory relationships: 1) HCR up peaks positively linked to gene targets, 2) HCR up peaks negatively linked to gene targets, 3) LCR up peaks positively linked to gene targets, and 4) LCR up peaks negatively linked to gene targets. Genes were ranked by the Pk2G link significance.

(**F**) Illustration of the peak-gene relationships most strongly associated with the metabolic pathways (left = HCR up peaks positively associated with gene expression, right = LCR up peaks negatively associated with gene expression).

(**G**) Ranked gene set enrichment analysis of Pk2G targets linked to differential peaks for multiple trials (columns), with the number of input genes indicated in the bottom bar plot. Significant gene sets (FDR < 0.05) were kappa-reduced and the -$\log_{10}$(*p*-values) of the most significant category within the representative term is plotted, with '•' annotating significant clusters. The number of gene sets reduced into the cluster are annotated on the right side of the heatmap. The prominent gene sets collapsed into the two largest clusters are listed on the far-right side of the plot, with the size of the text associated with the number of times a gene set related to that term is present within the cluster. The additional comparison to the RNA-Enrich results shows concordance of links to RNA-Seq HCR/LCR differences.

(**H-I**) Comparison of line-specific Pk2G links enriched for metabolic pathways in the HCR up positive Pk2G (**H**) and LCR up negative Pk2G (**I**) pathway enrichment analysis. Signed -$\log_{10}$(*p*-values) are concordant between the HCR only (x-axis) and LCR only (y-axis) Pk2G analysis. Points are colored by the link type to detail negative and positive distal-promoter raw correlations contributing towards positive and negative Pk2G links, respectively.

(**J**) HCR-favored Pk2G mitochondrial metabolism genes compared to their differential expression. Comparison of the significance from the Pk2G correlation analysis included in the pathway enrichment analysis to the differential expression of the 537 genes that significantly enrich gene sets via GSEA of HCR Up, Pos gene links and RNA-Enrich differential gene expression. Y-axis shows -$\log_{10}$(*p*-value) from the HCR/LCR differential gene expression analysis, and the -$\log_{10}$(*p*-value) from the Pk2G link is plotted on the x-axis. Highly significant genes have strong relationships with epigenomic features, supporting coordinated regulation of the chromatin and gene expression.

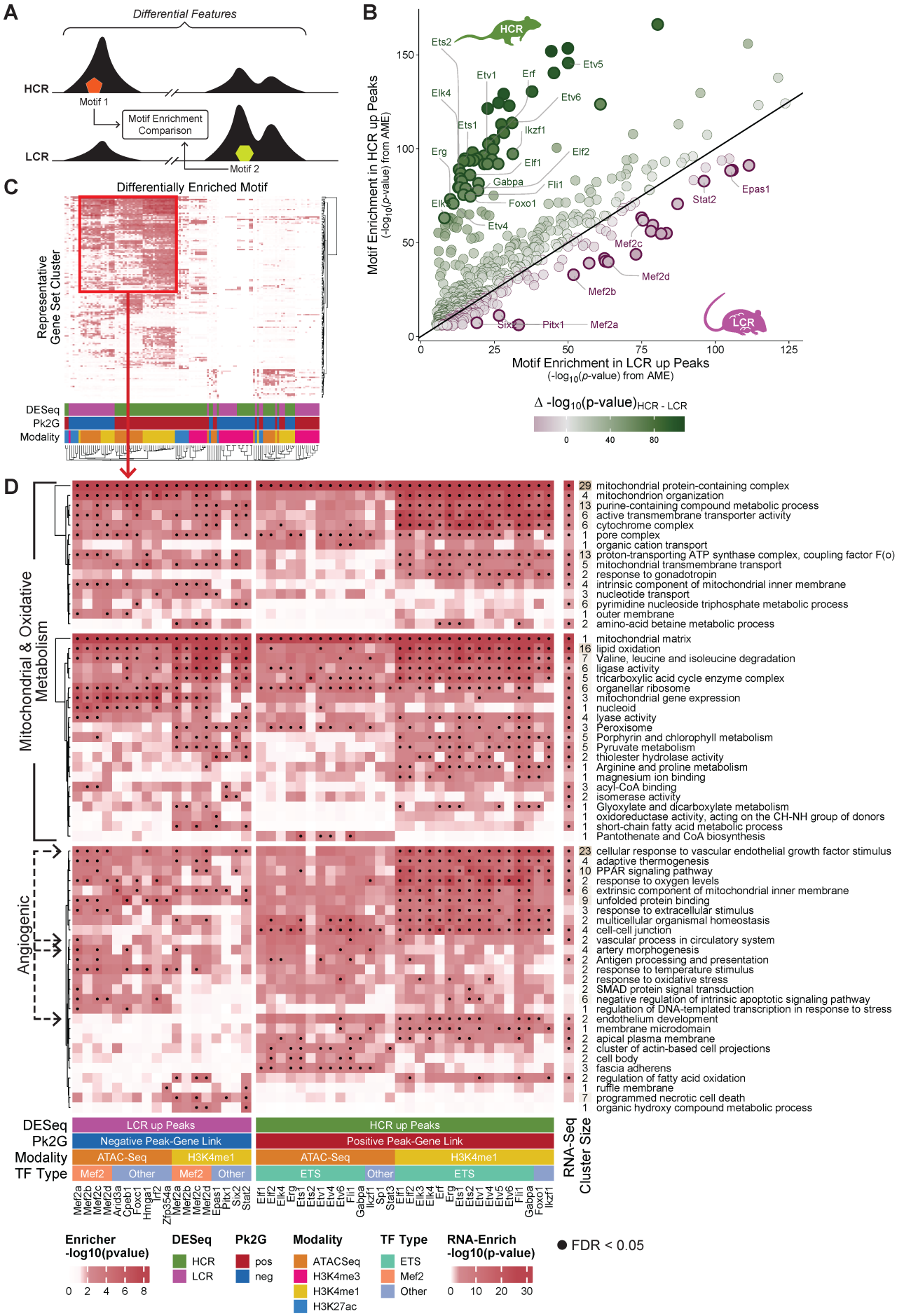

A
Differential Features
HCR
Motif 1
Motif Enrichment Comparison
LCR
Motif 2
B
HCR
LCR
Motif Enrichment in HCR up Peaks (-log10(p-value) from AME)
Motif Enrichment in LCR up Peaks (-log10(p-value) from AME)
Ets2
Etv5
Etv1
Erf
Elk4
Etv6
Ets1
Ikzf1
Erg
Elf2
Elf1
Gabpa
Fli1
Elk1
Foxo1
Etv4
Stat2
Epas1
Mef2c
Mef2d
Mef2b
Six2
Pitx1
Mef2a
Δ -log10(p-value)HCR - LCR
C
Differentially Enriched Motif
Representative Gene Set Cluster
DESeq
Pk2G
Modality
D
Mitochondrial & Oxidative Metabolism
Angiogenic
29 mitochondrial protein-containing complex
4 mitochondrion organization
13 purine-containing compound metabolic process
6 active transmembrane transporter activity
6 cytochrome complex
1 pore complex
1 organic cation transport
13 proton-transporting ATP synthase complex, coupling factor F(o)
5 mitochondrial transmembrane transport
2 response to gonadotropin
4 intrinsic component of mitochondrial inner membrane
3 nucleotide transport
6 pyrimidine nucleoside triphosphate metabolic process
1 outer membrane
2 amino-acid betaine metabolic process
1 mitochondrial matrix
16 lipid oxidation
7 Valine, leucine and isoleucine degradation
6 ligase activity
5 tricarboxylic acid cycle enzyme complex
6 organellar ribosome
3 mitochondrial gene expression
1 nucleoid
4 lyase activity
3 Peroxisome
5 Porphyrin and chlorophyll metabolism
5 Pyruvate metabolism
2 thiolester hydrolase activity
1 Arginine and proline metabolism
1 magnesium ion binding
3 acyl-CoA binding
2 isomerase activity
1 Glyoxylate and dicarboxylate metabolism
1 oxidoreductase activity, acting on the CH-NH group of donors
1 short-chain fatty acid metabolic process
1 Pantothenate and CoA biosynthesis
23 cellular response to vascular endothelial growth factor stimulus
4 adaptive thermogenesis
10 PPAR signaling pathway
2 response to oxygen levels
6 extrinsic component of mitochondrial inner membrane
9 unfolded protein binding
3 response to extracellular stimulus
2 multicellular organismal homeostasis
4 cell-cell junction
2 vascular process in circulatory system
4 artery morphogenesis
2 Antigen processing and presentation
2 response to temperature stimulus
2 response to oxidative stress
2 SMAD protein signal transduction
6 negative regulation of intrinsic apoptotic signaling pathway
1 regulation of DNA-templated transcription in response to stress
2 endothelium development
1 membrane microdomain
2 apical plasma membrane
2 cluster of actin-based cell projections
2 cell body
3 fascia adherens
2 regulation of fatty acid oxidation
1 ruffle membrane
7 programmed necrotic cell death
1 organic hydroxy compound metabolic process
DESeq
Pk2G
Modality
TF Type
LCR up Peaks
HCR up Peaks
Negative Peak-Gene Link
Positive Peak-Gene Link
ATAC-Seq
H3K4me1
Mef2
Other
ETS
RNA-Seq Cluster Size
Enricher -log10(pvalue)
DESeq
HCR
LCR
Pk2G
pos
neg
Modality
ATACSeq
H3K4me3
H3K4me1
H3K27ac
TF Type
ETS
Mef2
Other
RNA-Enrich -log10(p-value)
FDR < 0.05

Fig. 4. Global HCR/LCR Differential Motif Enrichment Analysis

(**A**) Schematic of motif enrichment comparisons between HCR up features with HCR alleles and LCR up features with LCR alleles, including significantly different features identified in Fig. 2C.

(**B**) Comparison of motif enrichment (analysis of motif enrichment (AME) -$\log_{10}$(*p*-values)) within H3K4me1 HCR up peaks (y-axis) and LCR up peaks (x-axis). Points further from the line of identity (slope = 1) and colored darker (green = HCR, magenta = LCR) represent motifs with strong differential enrichment, in which motifs within the top 10% Δ -$\log_{10}$(*p*-values) are outlined. See additional results in Fig. S9.

(**C**) Over-representation analysis of genes linked to differentially enriched motifs outlined in panel B and Fig. S9. The heatmap plots the -$\log_{10}$(*p*-values) across trials stratified by motif, modality, direction of peak change (HCR up or LCR up), and directionality of Pk2G links. Each row is a representative gene set cluster into which significant gene sets (FDR < 0.05) were kappa-reduced. Heatmap presents motifs collapsed into TFs with columns kappa-clustered by gene target similarity and rows hierarchically clustered by -$\log_{10}$(*p*-value). Distinct regulatory patterns result, with the most prominent signal driven by motifs enriched in HCR up ATAC-Seq and H3K4me1 peaks, outlined in red. See panel (D) for an expanded view of this cluster.

(**D**) Subset of the over-representation results from panel C, including the top 10 gene set clusters per column. The number of child gene sets within the representative cluster are annotated on the lefthand side of the heatmap. Results are compared to RNA-Enrich (left column) and the heatmap plots the -$\log_{10}$(*p*-value) of the most significant gene set within the cluster with '•' annotating significant clusters (FDR < 0.05). Rows are split into three clusters based on kappa similarity of genes, and then hierarchically by -$\log_{10}$(*p*-value) within cluster; columns are manually ordered by column annotation categories. To reduce sparsity in the heatmap, rows are filtered to include terms that are FDR significant in more than 5 "motif, modality, HCR/LCR direction, and Pk2G direction" trials. TF motif version is noted within parentheses.

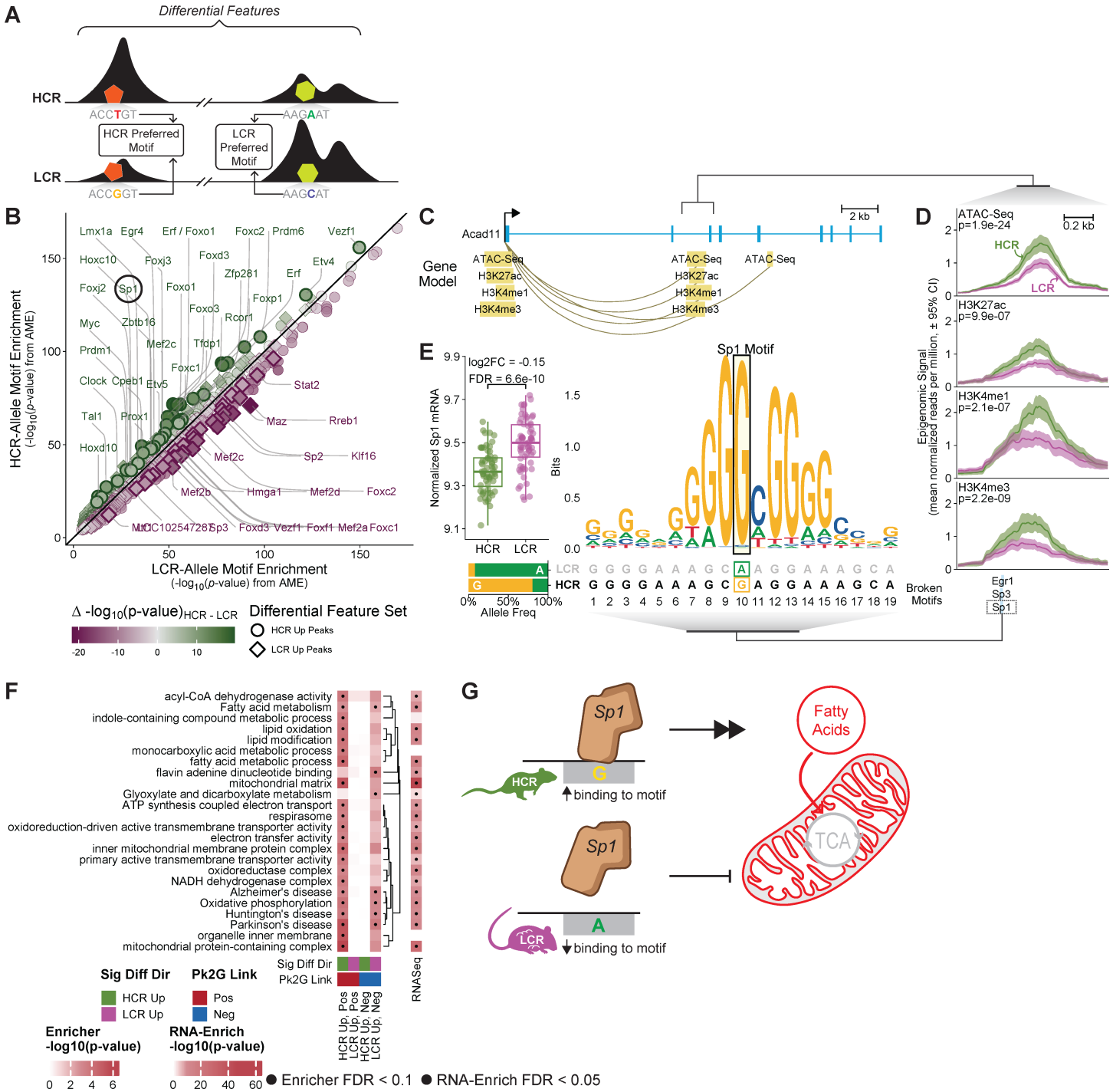


Fig. 5. Allele-Specific Influence on HCR/LCR Epigenomic Divergence

(**A**) Schematic of allele-specific differential motif enrichment analysis to detect TF binding differences due to fixed alleles within HCR up and LCR up regions. Motif enrichment is compared between HCR up regions with HCR alleles and HCR up regions with LCR alleles (left), as well as LCR up regions with LCR alleles and LCR up regions with HCR alleles (right).

(**B**) Comparison of allele-specific motif enrichment (AME -$\log_{10}$(*p*-values)) between HCR alleles (y-axis) and LCR alleles (x-axis) within HCR up (circles) and LCR up (diamonds) H3K4me1 peaks. Points further from the line of identity (slope = 1) indicate HCR allele (green) and LCR allele (magenta) driven motif enrichment. Motifs within the 10% Δ -$\log_{10}$(*p*-values) are outlined. See additional allele-specific results in Fig. S9.

(**C**) Example of a disrupted motif within differential peaks linked to *Acad11* expression. ATAC-Seq, H3K27ac, H3K4me1, and H3K4me3 peaks within the gene body of Acad11 have a positive Pk2G Link with Acad11.

(**D**) Peak signal at the motif-disrupted region linked to Acad11 (line = mean, shaded ± 95% CI). Signal was smoothed on a per-sample basis by applying a rolling mean (bin size = 20). The nominal *p*-value from the DESeq analysis is provided for each modality.

(**E**) Sequence comparison of HCR and LCR alleles within the Sp1 motif. The preferred Sp1 motif contains the G allele in HCR. 80% of the HCR population exhibit a T allele at this location, whereas 92% of the LCR population present the C allele.

(**F**) Over-representation analysis of genes linked to allele-disrupted motifs, when target genes specified by the column name are tested (e.g., 'HCR Up, Pos' = HCR alleles disrupt motifs within HCR up peaks that are positively linked to the gene target). Results are compared to RNA-Enrich (right column) and heatmap plots the $-\log_{10}$(*p*-value) from the respective tests, with '•' annotating significant gene sets (FDR < 0.1).

(**G**) Visual summary demonstrating how fixed alleles can disrupt TF binding and affect the expression of nuclear-encoded mitochondrial genes.

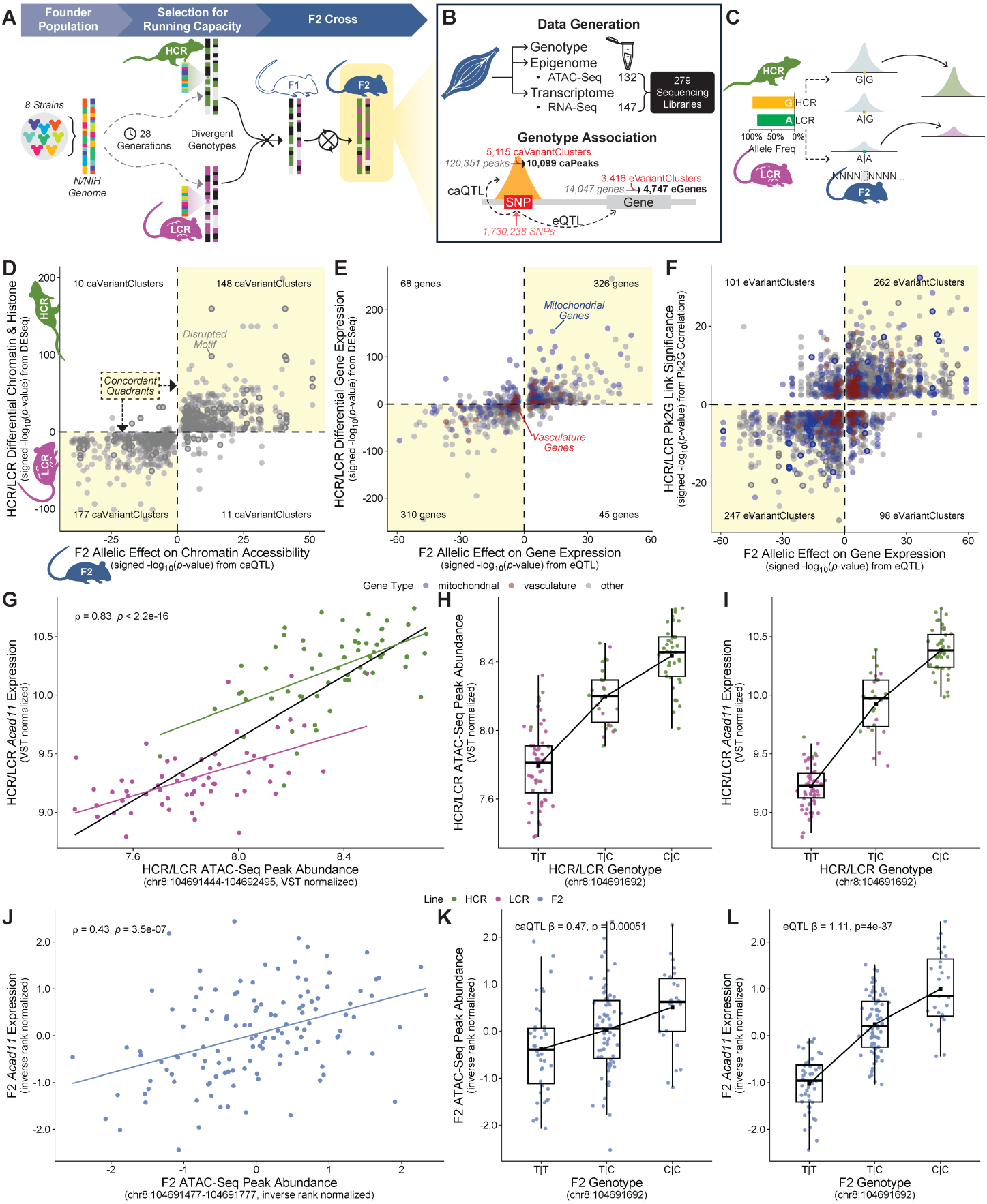


Fig. 6. Orthogonal Validation of Epigenomic Findings

(**A**) Simplified depiction of the previously generated F2 population. The far-left panel (Founder Population) illustrates the genetic mosaic of the N/NIH heterogeneous stock, which originated from eight inbred rat strains (represented by distinct colors). This heterogeneous stock served as the founder population for the HCR and LCR

lines. The center panel (Selection for Running Capacity) shows how 28 generations of artificial selection for intrinsic running capacity led to allele fixation and divergence between HCR (green) and LCR (magenta) populations. This process resulted in distinct genotypic and phenotypic profiles in each line. The far-right panel (F2 Cross) illustrates the crossing scheme used to generate the genetically diverse F2 population. First, HCR and LCR lines were crossed to produce an F1 population, illustrated by heterozygous loci inherited from both parental lines. The F1 generation was then randomly intercrossed, while avoiding sibling pairs, to produce the genetically recombinant F2 population containing a mixture of genetic variants derived from both HCR and LCR backgrounds.

(**B**) Overview of the validation study. F2 gastrocnemius gene expression (RNA-Seq) and chromatin accessibility (ATAC-Seq) were related to genotypes via eQTL and caQTL analyses, respectively. Linkage disequilibrium (LD) was accounted for by grouping (e/ca)Variants with $r^2 \geq 0.99$ to the most significant variant, labeled as (e/ca)VariantClusters.

(**C**) Schematic of directional concordance testing between F2 and HCR/LCR findings to validate proposed regulatory mechanisms driven by allelic convergence.

(**D-F**) Comparison of variant effects on chromatin and gene expression between HCR/LCR lines (y-axis) and F2 analysis (x-axis). Analyses were restricted to variants with $F_{ST} \geq 0.3$ between HCR and LCR. The effect allele is coded as the one with greater frequency in HCR than LCR. To assist with overplotting, only one occurrence of a variant with the highest r^2 to the top variant is included. Quadrants I and III (shaded) indicate directional concordance.

(**D**) HCR/LCR epigenomic peaks associated with open chromatin (ATAC-Seq, H3K27ac, H3K4me1, and H3K4me3) overlapped with F2 caPeaks. The $-\log_{10}$(*p*-values) from the caQTL analysis are signed by the caQTL slope. The $-\log_{10}$(*p*-values) from the HCR/LCR analysis are signed by DESeq $\log_2$FC (positive = greater signal in HCR). Variant-peak interactions were refined to caPeaks where the associated variant was within the peak or in strong LD ($r^2 > 0.8$) with the most significant caVariant (2,535 caPeaks linked to 1,582 caVariantClusters). Concordance = 94% (binomial *p*-value = 2.17×10^{-63}, expected = 50%, alternative = "greater"). Outlined points indicate peaks containing a disrupted motif based HCR/LCR allelic differences.

(**E**) Comparison of F2 eQTL effects versus HCR/LCR gene expression differences. Genes with HCR/LCR DESeq FDR < 0.01 and eGenes where a variant with $F_{ST} > 0.3$ was in strong LD with the top eVariant ($r^2 > 0.9$) were included for consideration. Concordance = 85% (binomial *p*-value = 1.63×10^{-89}, expected = 50%, alternative = "greater"). Points are colored by the biological relevance of the gene (blue = mitochondrial, red = vasculature, gray = other).

(**F**) Comparison of F2 eQTL versus HCR/LCR Pk2G links. The eQTL $-\log_{10}$(*p*-values) are signed by beta and HCR/LCR Pk2G $-\log_{10}$(*p*-values) are signed by Pk2G link direction. Concordance = 69% (binomial *p*-value = 3.15×10^{-17}, expected = 50%, alternative = "greater"). Outlined points indicate disrupted motif presence within the peak. Points are colored by the biological relevance of the gene (blue = mitochondrial, red = vasculature, gray = other).

(**G-I**) HCR/LCR VST normalized counts adjusted for additional experimental contrasts and technical variables and (**J-L**) F2 plots inverse normalized expression adjusted for sex, in which the relationship between the (**G,J**) regulatory peak and gene, (**H,K**) variant and regulatory peak, and (**I,L**) variant and gene expression are plotted where the line represents the mean across genotype combinations.

Methods

HCR/LCR Study Design

A total of 128 rats from generation 39 of the HCR/LCR line (38 weeks of age) were included in this study, including equal numbers of females and males (Fig. S1A). Animal procedures were approved by University of Michigan IACUC, Procol PRO00006206. Half of the rats, deemed the "training" group, underwent a six-week exercise regimen consisting of three days per week (M, W and F) of 30 minutes of supervised treadmill running up a 15-degree incline. The treadmill speed for each week of training was set at a constant speed at 75% of the average maximal running speed determined from a test performed the week prior. This weekly maximal running test was performed separately on a group of 3-4 rats from each line (LCR and HCR) and sex using the standard LCR/HCR test protocol, as described previously (*17*). When performed, this test took the place of the normal training procedure for that group of rats on Fridays. The other half of the rats (deemed the "sedentary" group) were subjected to once-weekly treadmill acclimatization on Tuesdays for six weeks. This consisted of 5 minutes of treadmill walking at a slow pace (5m / min) and 15 degree incline, followed by an additional 25 minutes of resting on a motionless treadmill, before being returned to their home cages. The training or acclimatization process was stagger-started over the course of three weeks to allow the final bout of exercise with dissection and sample collection to be performed after a consistent 72-96 hour washout period following the last bout of training.

Once training was complete and the washout period had passed, rats from both the "trained" and "sedentary" groups were subjected to an acute ascending-effort exercise bout using the standard LCR/HCR test protocol, with variation of the endpoint to allow sample collection at four different time points relative to initiation of exercise. The "rest" timepoint was collected immediately after animals were removed from their home cage. These rats did not perform any acute exercise on the day of sample collection. "Half-max" samples were collected from a group of rats that underwent the test protocol but were removed from the treadmill at half of their predicted maximum running time, which was determined by the group-specific averages measured on the last Friday of the training bout, as already described. The "half-max" timepoints were as follows: HCR females, 40 minutes; HCR males, 35 minutes; LCR females and males, 10 minutes. Samples collected at "maximum" exercise were collected from rats run to exhaustion according to the standard LCR/HCR test protocol. Samples collected during "recovery" were collected from rats that were run to exhaustion using the standard test protocol, then returned to their cage for 1 hour prior to dissection. The running time and distance achieved by animals in the "maximal" and "half-max" groups were recorded for each animal. At the time designated for dissection, which was immediately after removal from the treadmill for the "half-max" and "max" groups, and as described above for "rest" and "recovery" groups, rats were placed in an anesthesia induction chamber ready with isoflurane anesthetic vapor. Once anesthesia was induced (~15 seconds) the animal was transferred to a nose-cone connected to a metered anesthetic-delivery machine and were maintained under deep anesthesia during sample collection.

Accounting for Experimental Variability in the HCR/LCR Study

This study is part of a larger effort to investigate transcriptomic and epigenomic changes in skeletal muscle as a response to exercise training and an acute exercise bout in relation to CRF, but the presented results focus on the intrinsic differences between HCR and LCR. While the six-week training regimen resulted in increased running capacity in both HCR and LCR (Fig. S1D), there were minimal transcriptomic and epigenomic differences 48 hours after the last exercise bout (data not shown). However, an acute exercise bout induced considerable transcriptomic and epigenomic changes (data not shown). In this study, we included all samples when possible to maximize our power and carried out extensive QC on count data associated with RNA-Seq, ATAC-Seq, and CUT&Tag modalities. We normalized for technical and experimental variables (acute exercise, training group, and sex) to isolate HCR/LCR effect, as described in the applicable methods sections. A detailed analysis of exercise-induced experimental contrasts will be presented in future work.

Preparation of High-Throughput Sequencing Libraries

CUT&Tag Nuclei Isolation and Library Generation

Skeletal muscle nuclei were isolated from the gastrocnemius muscle of 60 HCR/LCR rats following the bulk nuclei isolation protocol published in Orchard P, 2021 (*97*). When isolating nuclei prior to CUT&Tag library preparation, modifications were made at the following steps: 3) a rotating plate was used to rock the samples for

3 minutes, 11 & 12) the nuclei pellet was spun at 500 x g, 13) a final concentration of 5 uM DAPI was used to count the nuclei. Throughout the protocol, CUT&Tag wash buffer was used to resuspend the sample and pre-wet the filters.

We profiled five histone modifications via CUT&Tag (*27*) following the "CUT&Tag-direct with CUTAC V.3" protocol (*98*). To reduce batch effect across experimental contrasts, samples were randomly assigned to five batches to carry out 360 CUT&Tag reactions at the bench (Fig. S1B). All antibodies were probed for a single sample in the same batch. In the case of library preparation failure, libraries were re-prepared in a later batch. A sixth preparation batch including four libraries was added to complete the library preparation. From the nuclei isolation step, 15K nuclei were prepared for each of the six CUT&Tag reactions per sample. Primary antibodies for histone modifications probed include H3K27me3 diluted 1:100 (Cell Signaling, 9733), H3K27ac diluted 1:100 (Abcam, ab4729), H3K4me1 diluted 1:100 (Cell Signaling, 5326), H3K4me3 diluted 1:100 (Diagenode, C15410003-50), and H3K36me3 diluted 1:100 (Active Motif, 61101). A negative control where no antibody was added was included for each sample. Bioanalyzer profiles were assessed to confirm success of experiment (no DNA in the sample), confirm nucleosomal patterns of histone modification samples, and measure DNA concentration for equimolar pooling of libraries to sequence. The 300 libraries where a primary antibody was included were sequenced. Libraries were re-randomized for sequencing batches, referred to as "flowcell" for the purposes of this study. Libraries were sequenced using paired-end, 150bp reads on the NovaSeq platform at the University of Michigan Advanced Genomics Core to an average depth of 7.7M paired-end reads per sample.

Five CUT&Tag libraries from a single sample were excluded due to it being a consistent outlier in PCA across all histone modifications. At the time of experiment, the muscle tissue also appeared discolored with an unusually powdery consistency. As a result, 59 samples (295 libraries) were included in our final CUT&Tag dataset.

ATAC-Seq Nuclei Isolation and Library Generation

Flash frozen gastrocnemius muscle samples from 128 HCR/LCR rats and 149 F2 rats were processed independently. In the HCR/LCR preparation, the samples were divided into approximately 21 batches comprising six samples each representing each of the six conditions, which were then re-randomized across sequencing batches (Fig. S1C). F2 samples were divided into 47 batches, with six samples in each batch. For each of the samples, tissue pieces weighing 50 mg were cut on dry ice using a frozen scalpel to prevent thawing. The samples were pulverized using liquid nitrogen and a cell crusher. Within each batch, each of the six samples was resuspended in 1 mL of ice-cold PBS to perform six different nuclei isolations. We developed a customized protocol (Supplemental Protocol S1) from the previously published ENCODE protocol, and used it to isolate nuclei, which is compatible with bulk ATAC seq. The nuclei were counted in Countess II FL Automated Cell Counter, and the appropriate volume of the suspension for 50K nuclei was spun down and used for the downstream transposition reaction (ENCODE Project Consortium 2012, 2012). The 50K nuclei were transposed, PCR amplified, and the libraries were generated in duplicate by adding appropriate sample indexes. Five HCR/LCR and eight F2 ATAC-Seq samples were excluded due to failure in pre-sequencing QC or unintentional duplication, resulting in 123 samples in the HCR/LCR and 141 samples in the F2 ATAC-Seq datasets. Libraries were sequenced using paired-end, 51 bp reads on the NovaSeq platform to an average depth of 63M paired-end reads per HCR/LCR sample (across 2-4 read groups) and 67M paired-end reads per F2 sample (across 1-4 read groups).

RNA-Seq Library Generation

Frozen gastrocnemius muscle from 128 HCR/LCR and 150 F2 samples were pulverized in liquid nitrogen. RNA was subsequently extracted using the Zymo Direct-zol kit. Stranded RNA-Seq libraries were generated using poly(A) selection and were sequenced using paired-end, 150bp reads on the HiSeq platform to an average depth of 39M paired-end reads per HCR/LCR sample and 37M paired-end reads per F2 sample.

Data Analysis

All analyses were performed using R Statistical Software (v4.3.2; 2023-10-31) and Python (v3.10.4). Commonly used R packages include tidyverse (*99*), GenomicRanges (*100*), plyranges (*101*), and ComplexHeatmap (*102*, *103*). Snakemake (*104*) was used to create pipelines used throughout the analysis. Sequencing data was aligned to the ensembl release (v109) of the rn7 genome (*105*).

Processing of RNA-Seq Data

Fastq files were trimmed using the TrimmomaticPE function (Trimmomatic (*106*) v0.36) with the following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads SLIDINGWINDOW:5:20. Reads were aligned using STAR (*107*) (v2.7.6a-aekjdpr) with the parameters "-c --outSAMtype BAM SortedByCoordinate". To create the count matrix, gene counts from QC.geneCounts.txt.gz files produced by QoRTs command QC with parameter –stranded and the same GTF file used in the alignment were used, returning 30,454 genes. The count matrix was filtered to remove non-autosomal genes and genes where all samples had 0 counts. For the HCR/LCR analysis, RNA-Seq counts were further filtered to include genes where more than 10 samples had a count value > 5, resulting in 14,781 genes in the dataset used in this analysis.

Processing of ATAC-Seq and CUT&Tag Data

Fastq file adapters were trimmed using cta version 0.1.2 (https://github.com/ ParkerLab/cta) and then aligned to the rn7 genome (ensembl file Rattus_norvegicus.mRatBN7.2.dna.toplevel.fa.gz) using the bwa (*108*) *mem* version 0.7.15 parameters "-I 200,200,5000" and then sorted using samtools (*109*) version 1.13. The bwa "-M" flag was used for aligning ATAC-Seq data only. ATAC-Seq libraries were prepared with multiple read groups, which were merged using samtools (*109*). Duplicates were marked using picard (*110*) *MarkDuplicates* version 2.25.5 with parameters "-ASSUME_SORTED true -VALIDATION_STRINGENCY LENIENT". Properly prepared and mapped autosomal reads with high quality alignments (> 30) were retained using samtools (*109*) *view* with parameters "-h -f 3 -F 4 -F 8 -F 256 -F 1024 -F 2048 -q 30". Bedtools (*111*) version 2.30.0 was used to remove the custom blacklist regions created specifically for this data. BAM files were converted to BED files and peaks were called using the *callpeak* command from MACS2 (*112*) version 2.2.4. Narrow peaks were called for H3K27ac, H3K4me1, and H3K4me3. Broad peaks were called for H3K27me3 and H3K36me3. Narrow and broad peaks were called for ATAC-Seq. Parameters "-g hs -B --keep-dup all -q 0.05" were included for all peak calls. H3K27me3 and H3K36me3 peak calling included the "--broad" flag. ATAC-Seq broad peaks had the added parameters, "--SPMR -f BED --nomodel --shift -100 --seed 762873 --extsize 200 --broad". ATAC-Seq narrow peak call had the added parameters, "--SPMR -f BED --nomodel --shift -37 --seed 762873 --extsize 73 --call-summits". Ataqv (*113*) was used to generate QC metrics used for further data normalization.

ATAC-Seq and CUT&Tag Count Matrix Construction for HCR/LCR Study

MACS2 called narrow peaks (H3K27ac, H3K4me1, H3K4me3), broad peaks (H3K27me3, H3K36me3), and narrow summits extended ± 150bp (ATAC-Seq) were used to create the consensus peak matrices. For each unique line + sex + acute exercise group, peaks were merged and the number of distinct samples that the peak occurred in was tallied. Peaks were retained if they were present in the majority of samples for each line + sex + acute exercise group (Fig. S2A-R). Peaks present in the majority of any of the 12-16 unique groups were merged to create the master peak matrix. In R, plyranges (*101*) *reduce_ranges* was used create the master peak set for each modality and chromVar (*114*) *getCounts* with arguments "paired = FALSE, by_rg = FALSE, format = 'bam'" was used to obtain counts within the consensus peak regions from the pruned BAM file.

Custom Blacklist Creation for rn7 Genome

A widely used blacklist for the rn7 genome was not available at the time of analysis, so a blacklist custom to this dataset was created following the process outlined in Amemiya, 2019 (*115*). A combination of 336 CUT&Tag and ATAC-Seq libraries prepared from 56 rats were used to create this blacklist region file. Briefly, 1 kb bins with 100bp overlapping regions were created using bedtools (*111*) version 2.30.0 *makewindows* and the number of reads present in each bin were counted. To detect erroneous signal, quantile normalization (broman::normalize (*116*)) was carried out on binned count matrices, for each data modality independently. The region was included in the final blacklist based on the presence of outlier signal in all six data modalities. To consider the fact that some datasets contain sequencing information for differing areas of the genome, the region was also included in the blacklist file if the median value of the quantile normalized matrix was 0. This resulted in 278 blacklist regions with cumulative width of 848 kb.

Latent Variable Discovery for ATAC-Seq, CUT&Tag, and RNA-Seq

RUVSeq (*117*) was used to generate and trial 1-10 latent variables. Control peaks consisted of features with a DESeq *p*-value > 0.1 across all known experimental and technical contrasts, with the number of control peaks

used ranging from 2,000-14,000+. Latent variable inclusion was determined by their ability to improve the number of significant DESeq features (Fig. S4A) and observed variability in PCA when added to the starting model (Fig. S4C-D), while ensuring that the latent variable(s) show a relationship with the ataqv (*113*) QC attributes and minimal relationship with the binarized experimental variables (Fig. S3A-G). When multiple models yielded similar results, the simpler model was preferred.

RNA-Seq, ATAC-Seq, and CUT&Tag Differential Analysis for HCR/LCR study

DESeq2 (*118*) was used to perform differential analysis on CUT&Tag histone modification peaks (H3K27ac, H3K4me1, H3K4me3, H3K27me3, H3K36me3), ATAC-Seq peaks, and RNA-Seq gene counts using the model specified in Fig. S4B. For the ATAC-Seq and RNA-Seq analyses, a stringent significance threshold of FDR < 0.01 and $|\log_2 FC| > 0.5$ was applied. Technical, experimental contrasts, and latent variables were included in the DESeq model. Given the smaller sample size in the CUT&Tag libraries, a more lenient cutoff of FDR < 0.1 and $|\log_2 FC| > 0.5$ was utilized for histone modifications.

Chromatin State Discovery

ChromHMM (*28*) was used to call chromatin states in 200bp bins across the genome. Custom rn7 files were created for chromosome sizes, coordinates, and anchor files using the R packages GenomicFeatures (*100*) and TxDb.Rnorvegicus.UCSC.rn7.refGene.

The 56 samples with complete ATAC-Seq, H3K27ac, H3K4me1, H3K4me3, H3K36me3, and H3K27me3 modalities (336 total libraries) were integrated to discover chromatin states. Custom blacklist regions were removed from BAM files, which were then down sampled to the lowest read depth (> 2M) per modality. If the lowest read depth was < 2M reads within a modality, then BAM files with read depths > 2M were down sampled to 2M and BAM files with read depths < 2M were not down sampled. There were three libraries with read depths ranging from 1.5-1.7M. ChromHMM state models ranging from 2-20 were assessed to choose the 12-state model, which presented minimal change in the log likelihood of the subsequent model and was most informative, given the combination of epigenomic marks. Emission states were manually assigned the following chromatin state names (with associated abbreviations): E1 = Strong Enhancer, Low (EnhS_E1), E3 = Weak Enhancer (EnhW_E3), E4 = Strong Enhancer (EnhS_E4), E5 = Weak Promoter (PromWk_E5), E6 = Transcription Transition (TxTrans_E6), E7 = Strong Promoter (PromA_E7), E8 = Inactive/Poised Promoter (PromPois_E8), E10 = Repressed PolyComb (ReprPC_E10), E12 = Transcription Elongation (TxnElon_E12). For example, we observe greater likelihood of a Strong Enhancer chromatin state call, in comparison to other states, at promoter regions of genes (Fig. S1E). Emission states E2, E9, and E11, returned states associated with low signal, were assigned the "Quiescent/Low" state and summed together into a single "Quiescent/Low", labeled as QuiesSum_EX in the results data tables.

Chromatin State Differential Analysis

ChromHMM-derived posterior probabilities were prepared for group analysis by retaining 200bp regions where any line, sex, or acute exercise group called a non-quiescent state, based on the highest initial group averaged state probability. In the case of a tie, a non-quiescent state assignment was prioritized. This was followed by summing quiescent states into one "QuiesSum" state. This produced posterior probability matrices that included 1,160,951 200bp regions. For each chromatin state, 200bp regions where the sum of posterior probabilities across samples was greater than 0 were included for differential analysis. The count of merged regions included in this analysis is provided in Fig. 2C. Linear regression was applied to posterior probabilities that were inverse rank normalized (ties = "random") across samples, using the model $\sim Line + Sex + Acute\ Exercise\ Group$. *P*-values for the HCR/LCR line contrast were obtained and FDR-adjusted. Chromatin state regions with an FDR < 0.1 were considered significantly different. We excluded emission state E1 (labeled as "Strong Enhancer, Low") from further analyses because it showed consistently low ChromHMM posterior probabilities across samples. At most, 22/56 samples had a posterior > 0.1 for this state, and its 99th percentile posterior was 0.07. In contrast, the other states had many 200bp regions where all 56 samples returned a posterior > 0.1, with the 99th percentiles of posteriors being 0.65 for Repressed PolyComb, 0.83 for Transcription Transition, and > 0.96 for the remaining 7 chromatin states. The prevalence of near-zero posteriors within the E1 chromatin state is unlikely to return a "group call" of this state, which is determined from the averaged posteriors from samples within the group.

Pathway Enrichment Analysis

To ensure consistency across transcriptomic and epigenomic comparisons, a custom list of gene sets to probe was created by retrieving gene sets and their associated genes (ensembl IDs) from GOBP, GOMF, GOCC, and KEGG via msigdbr (*119*). Missing Ensembl or Entrez ID matches were filled in where possible using the *biomaRt* (*120*) package to retrieve gene information from the *rnorvegicus_gene_ensembl* v109 (*105*) database, ensuring consistency in gene set annotations. Categories with 10 to 500 genes were probed for pathway analysis. Resulting *p*-values were FDR adjusted for each individual database using the Benjamini-Hochberg method (*121*). Unless stated otherwise, categories returning an FDR < 0.05 were considered significant.

Pathway enrichment analysis was performed independently on all data modalities. RNA-Enrich (*122*) was used for pathway enrichment analysis of differential gene expression data which included all genes, ranked by *p*-value with direction of change provided.

To assess the processes that Pk2G Link target genes were involved in, HCR/LCR differential peaks passing the FDR threshold provided in Fig. 2C were queried. The list of input genes was separated by direction of peak and direction of relationship between the peak and target genes. The number of input genes for each trial is as follows: HCR up peaks positively linked to 4,600 genes, HCR up peaks negatively linked to 4,678 genes, LCR up peaks positively linked to 4,889 genes, and LCR up peaks negatively linked to 4,490 genes. The fgsea function from the fgsea (*123*) R package was used to carry out pathway enrichment on these sets of genes, ranked by significance of correlation.

The enricher function from the ClusterProfiler (*124*) R package was used to carry out over-representation analysis using specified target genes and all genes expressed in the skeletal muscle as the background list. When the target genes of regions contributing to allele-specific differential motif enrichment analysis were considered, the HCR up pos trial used 330 genes, HCR up neg trial used 279 genes, LCR up pos trial used 272 genes, and LCR up neg trial used 281 genes as the target list.

In efforts to reduce redundancy and clarify comparisons across independent trials, the clustering approach described by Metascape (*125*) was applied. Functions created to carry out gene set reduction by kappa clustering are published on GitHub (https://github.com/weitzela/reducekappa, commit 04eb74d). Briefly, pairwise kappa values were calculated to capture the similarity of genes associated with all categories across (one or multiple) trials. Kappa scores were min/max normalized to adjust for negative values. The kappa matrix was converted to a distance matrix by subtracting the kappa scores from 1, which was then used for hierarchical clustering (method = "average"). Clusters were identified using a threshold height of 0.7. The representative category for each cluster was selected as the term with the lowest *p*-value across all trials. For comparisons across multiple modalities, values associated with the most significant category within each cluster were retained for individual trials, even if that category wasn't named the "representative" one.

Correlation of Distal Regions with Promoter Regions and Gene Expression

To expand annotation links between epigenomic features and genes, we conducted a correlation analysis utilizing RNA-Seq, ATAC-Seq, H3K27ac, H3K4me1, and H3K4me3 data modalities. We extended the approach described in Currin et al. (*126*), detailed further here. This adaptable peak-gene annotation method is made publicly available as an R package (https://github.com/weitzela/PeakGeneNet).

We used the limma::removeBatchEffect function to adjust the DESeq VST counts (blind = FALSE) for technical variables such as prep date, flowcell, and latent variables discovered/used for differential analysis. To ensure that the line effect remained when evaluating the relationship between peaks and genes, non-*Line* experimental contrasts were further adjusted for via linear regression. Promoter regions were defined as regions 2 kb upstream and 1 kb downstream of the canonical TSS. Distal regions were defined as those that spanned 1 Mb ± TSS, excluding promoter regions. Promoter peaks consisted of ATAC-Seq, H3K4me3, and H3K27ac peaks that reside in the promoter region. Distal peaks consisted of ATAC-Seq, H3K4me1, and H3K27ac peaks residing in distal regions, along with H3K4me3 peaks within 10 kb ± TSS. The following four link types were assessed: 1) promoter peak to gene, 2) distal peak to gene, 3) distal peak to promoter peak, and 4) promoter peak to promoter peak. In this analysis, 12,289 genes had at least one *promoter peak* and were included for further analysis. We conducted pairwise Spearman correlations between the variables identified in the described link types, leading to 12,481,099

unique tests. The *p*-values were adjusted using FDR (BH) within each link type and modality pair. We considered tests with an adjusted *p*-value < 0.05 to be statistically significant. 566,772 significant peak-peak or peak-gene relationships were identified where 10,786 genes are annotated via these significant correlations.

Due to variable relationships between distal peaks→promoter peaks and promoter peaks→genes, all peak→gene relationships were consolidated into "Pk2G Links" with a consensus direction assigned using a prioritized, four-tier approach, where Tier 1 represents the highest confidence. Tier 1) Assign direction based on FDR significant links, bringing in correlation coefficients with nominal pvalue < 0.05 to inform about direction where necessary. All peak:gene directions are in agreement. Tier 2) For links with relationships with conflicting direction and all links are FDR significant, the direction assigned to the link with the most significant FDR is honored. For example, if the distal peak→Gene $r > 0$, the distal peak→promoter peak $r > 0$, and promoter peak→gene $r < 0$, the distal peak→gene relationship would return a positive relationship, while the distal peak→promoter peak →→ promoter peak→gene link would return a negative relationship. In this case, the distal peak→gene and distal peak→promoter peak FDRs would be compared. If the distal peak→gene FDR < distal peak→promoter peak FDR, then the consensus direction between the peak:gene would be positive. Tier 3) For links with conflicting direction, but not all links meet FDR significance, the relationship with the most significant nominal *p*-value < 0.05 was honored following a concept similar to assignment #2. Tier 4) Direction of ambiguous peaks where the distal peak→promoter peak is FDR significant, but neither promoter peak→gene or distal peak→gene correlations have a nominal *p*-value < 0.05, was assigned based on the correlation (*p*-value < 0.05) between the distal peak and other peaks significantly (FDR < 0.05) linked to the gene. The most FDR significant peaks were prioritized. Pk2G Links with a direction confidence tier < 4 were considered when identifying target genes in motif related analyses.

SNP Calling from HCR/LCR RNA-Seq Data

Using the processed RNA-Seq BAM files, reads with an edit distance > 2 and duplicated reads were removed. Subsequently, BAM file base quality scores were recalibrated using the SplitNCigarReads, BaseRecalibrator, and ApplyBQSR commands from GATK (*127*) version 4.3.0.0. Recalibration detection referred to the known variants found in the 8 HCR/LCR founder strains (*128*). SNPs were called using bcftools (*129*, *130*) version 1.12 *mpileup* (-Q 20 -I -a DP,AD -d 10000 -E) and *call* (-mv -a GQ) arguments. HCR and LCR line assignments were provided as the population parameter for SNP calling. Genotypes with GQ < 35 were marked as missing and variants were removed if QUAL < 20 or > 50% of the genotypes were missing. The variant sites were further filtered to retain those discovered in the founder population (*128*).

Genotyping of F2 Population

A custom Affymetrix Axiom panel of ~700,000 SNPs was designed to genotype the 613 F2 animals using DNA extracted from frozen liver tissue; SNPs in this panel were selected based on criteria previously described in Ren et al. (*22*). These SNPs were based on the rn4 version of the rat genome assembly, then lifted over to the rn6 genome using LiftoverVcf (picard v2.8.1) (*110*) and then filtered against an rn6 blacklist (https://github.com/shwetaramdas/maskfiles/blob/master/rataccessibleregionsmaskfiles/strains_intersect.bed), resulting in 380,990 autosomal SNPs. The genotypes were bidirectionally lifted over from the rn6 to the rn7 genome using GATK (*127*) v4.3.0.0 *LiftoverVcf* (--RECOVER_SWAPPED_REF_ALT false). Variants were first assigned location-based IDs in the rn6 VCF using bcftools v1.7 (*130*) *annotate* (-x ID --set-id +'%CHROM_%POS'), lifted to rn7, and then back to rn6. Variants that returned the same genomic position in the reverse (rn7→rn6) lift were retained using bcftools v1.7 (*130*) *view*. This resulted in 375,220 autosomal SNPs in the rn7 VCF. Genotypes were imputed from this panel for samples that have RNA-Seq data as well (described below).

Genotype Imputation

Genetic Map Preprocessing

The rn7 genetic map published by Jong et al. (*131*) was used. Discrepancies throughout the file were addressed for autosomal chromosomes by the following preprocessing steps: 1) removal of sites where contig ≠ chr, 2) linear transformation of cM values for chromosomes with negative relationships between position and average cM, 3) removal of outlier cM:Position pairs, 4) sorted by position and smoothed the cM values by running mean weighted

by standard deviation (bin size = 500). These steps removed 1,626 entries and resulted in a genetic map with 149,209 positions.

Genotype Imputation

IMPUTE5 (*132*) version 1.2.0 and SHAPEIT5 (*133*) version 5.1.1 were used to carry out genotype imputation. The reference panel published by Hermsen et al. (*134*), was prephased using IMPUTE5 *xcftools_static view* (-m 0.03125). Target files were prephased using SHAPEIT5 (*133*) *phase_common_static*, providing the phased reference panel and referring to the preprocessed genetic map. Genomic regions were chunked for imputation using IMPUTE5 *imp5Chunker_v1.2.0_static*. Genotype imputation was carried out using IMPUTE5 *impute5_v1.2.0_static*, setting the effective population size to 52 and providing the phased reference panel, phased target file, preprocessed genetic map, and computed chunk and buffer regions. Post imputation, genotypes were filtered to retain sites with a minor allele count (MAC) > 10 and posterior probability > 0.97 for all samples. As a quality control measure, imputed genotypes were compared to reference alleles derived from ATAC-Seq BAM files to verify consistency between the proportions of reference and alternate alleles at both homozygous and heterozygous sites (Fig. S5).

Calculation of Fixation Index for HCR/LCR

The population differentiation between HCR and LCR lines was assessed by computing the Fixation index (F_{ST}) (*135*) to compare reference allele frequencies. The following formula was used $F_{ST} = \frac{(RAF_{HCR} - RAF_{LCR})^2}{4 \times \overline{RAF} \times (1 - \overline{RAF})}$. F_{ST} scores range from 0 to 1, where a score of 1 is interpreted as complete population differentiation. $F_{ST} > 0.25$ is described as very high differentiation (*135*). A conservative threshold of $F_{ST} > 0.3$ is applied to identify 163,767 that exhibited high population differentiation between HCR and LCR.

Comparison of F_{ST} to Peak Significance

We performed a linear regression using the model: $-log_{10}(p\ value)_{Feature\ DESeq:\frac{HCR}{LCR}} \sim GC\ content + FPKM + variant\ location\ (noncoding\ vs\ coding) + F_{ST}\ (inverse\ rank\ normalized)$. In instances where multiple variants reside within one epigenomic feature, the variant with the maximum F_{ST} value was retained to ensure unique epigenomic peak observations. GC content, FPKM, and variant location are included in the model to control for feature-specific artifacts.

Motif Analysis

Motif Inclusion

The MotifDb (*136*) R package was used to query motifs cataloged for humans and rats from cisbp, JASPAR, HOCOMOCO, and other databases. The initial 6,131 motifs were filtered to include TFs present in the RNA-Seq data. In instances of duplicate JASPAR motifs, the most recent version was kept. Redundancy was reduced by removing duplicated motifs (identical universal motif PPMs or TFBSTools::PWMSimilarity (*137*) = 0), keeping only one instance of the particular motif per TF. This filtering resulted in 1,962 motifs from 516 transcription factors.

Preparation of Canonical Sequences

Sequences within significantly different peaks (defined in Fig. 2C) and ChromHMM regions (FDR < 0.1) were used as the input sequences for motif enrichment and PWM scans. Canonical HCR and LCR sequences were prepared for each feature set by replacing the reference allele with HCR and LCR alleles meeting the following criteria: 1) the variant site has an $F_{ST} > 0.3$ and 2) the alternate allele frequency within the given line is > 0.5. For regions where multiple SNPs fall within a close proximity (e.g., within the same feature), the one combination of ref/alt alleles that was most represented in the line was used for replacement. The rationale for this is to include relevant combinations of sites that indicate population differentiation while preventing over-representation of the input sequences in the analysis.

Haplotype-Aware PWM Scans

FIMO (v5.5.5) (*138*) was used to carry out PWM scans for HCR and LCR sequences that extended ± 35bp of a variant with $F_{ST} > 0.3$. Disrupted motifs are defined as those with Δ FIMO $\log_{10}$(*p*-value) > 1.

Differential Motif Enrichment Analysis
Motif enrichment was performed, using AME (v5.5.5) (*139*) with the arguments --control --shuffle-- --scoring max, to specify the shuffled background model and scoring based on the maximum PWM match. The input sequences (pre-trimmed) were retrieved for each modality & direction of change (e.g., HCR up, LCR up) based on the following significance thresholds: A) ATAC-Seq peaks with FDR < 0.01 & abs($\log_2$FC) > 0.5; B) H3K4me3, H3K4me1, and H3K27ac peaks with FDR < 0.1 & abs($\log_2$FC) > 0.5; and C) ChromHMM states associated with open chromatin (EnhS_E1, EnhW_E3, EnhS_E4, PromWk_E5, PromA_E7, and PromPois_E8) with FDR < 0.1. For ChromHMM states, directionality was assigned by comparing mean posterior probabilities within significant regions between groups.

To control for differences in genomic territory associated with HCR up and LCR up regions within a given modality, we trimmed the larger set of features by removing the least significant regions until the length of sequences spanned within ~1 kb of the smaller set of features. Final sequence lengths used for enrichment analysis per modality were: ATAC-Seq (872.3 kb), H3K4me3 (870.8 kb), H3K4me1 (2.5 Mb), H3K27ac (233.2 kb), EnhS_E1 (135.5 kb), EnhW_E3 (1.4 Mb), EnhS_E4 (1.3 Mb), PromWk_E5 (590.7 kb), PromA_E7 (339.5 kb), and PromPois_E8 (351.5 kb).

Motif enrichment was carried out independently for each modality and direction of change (e.g., HCR up and LCR up). For each modality and direction of change, two AME trials were carried out using 1) HCR canonical sequences and 2) LCR canonical sequences.

Differentially enriched motifs were identified per modality using two approaches: a global comparison and an allele-specific comparison. In the global analysis, HCR sequences in HCR up regions were compared to LCR sequences in LCR up regions. For the allele-specific analysis, two comparisons took place: 1) HCR vs. LCR sequences in HCR up regions, and 2) HCR vs. LCR sequences in LCR up regions. For both global and allele-specific trials, we compared the resulting AME -log10(*p*-values). Motifs were considered differentially enriched if the absolute difference in -log10(*p*-values) ranked in the top 10% of all Δ -log10(*p*-values) in the comparison.

In the allele-specific analysis, we identified regions that contributed toward the differential motif enrichment and contained motifs with disrupted PWMs. Starting with the AME results, regions were initially filtered to retain those that exclusively returned a true positive and either 1) was absent from the motif match in the counter group (e.g., when assessing HCR sequences, there was not a motif match returned in the LCR sequence), or 2) the difference in the PWM scores for the assessed region was greater than 1 (target group region – counter group region). The regions were further filtered to retain those that contained a variant with an $F_{ST} > 0.3$, following the assumption that regions without SNPs are likely to produce the same or similar PWM scores across different tests. Regions with disrupted motifs were further refined by directly comparing PWM scans of HCR and LCR sequences, retaining those with Δ FIMO -$\log_{10}$(*p*-value) > 1.

Sample Swap Detection for F2 Analysis
QTLtools (*140*) (v1.2) command mbv was used to detect sample swaps, using the F2 genotypes that were bidirectionally lifted over from rn6 to rn7. The flag --filter-mapping-quality 255 was applied when processing RNA-Seq BAM files, while this flag was not used for the pruned ATAC-Seq BAM files. When assessing ATAC-Seq sample swaps, matches with *n_het_covered > 5* were considered. Samples displaying a high degree of heterozygous matches were retained and sample ID swaps were corrected where necessary. One RNA-Seq and two ATAC-Seq samples were excluded due to low heterozygous consistency (<63%) with any genotype profile.

ATAC-Seq Count Matrix Construction for F2 Analysis
In efforts to improve performance in the caQTL scan, a consensus peak matrix was uniquely prepared for the F2 samples. The peak count matrix was created by including peaks (narrow summits extended ±150bp) exhibited by at least 20% of samples in either male or female groups. High scoring regions were selected as the "candidate" region for all overlapping regions to create a non-redundant peak set.

Sample Drop Criteria for F2 Analysis
We dropped outliers in PC space prior to performing the scans. We performed DESeq2 (*118*)-style size factor normalization of the count matrices (as implemented in pyqtl, https://github.com/broadinstitute/pyqtl), dropped features where samples all had 0 counts, filtered to features with at least 10 normalized counts in at least 10

percent of samples, filled any remaining 0s with a value equal to (½ of the minimum non-zero value for that feature), $\log_{10}$ transformed the matrix, and then performed PCA on this normalized matrix. Outliers were determined based on visual inspection of the top 10 PCs. The final sample sizes for the (e/ca)QTL scans were 147 RNA-Seq samples and 132 ATAC-Seq samples.

F2 eQTL & caQTL Analysis for Detection of eGenes and caPeaks

We performed cis-expression and chromatin accessibility quantitative trait locus scans (eQTL, caQTL) using the F2 RNA-Seq and ATAC-Seq data. We used a linear mixed model, implemented in GEMMA v. 0.98.5 (gemma -bfile $plink_prefix -n $phenotype_index -snps variants-to-test.txt -lmm -k $kinship_matrix -o $output_prefix -c covariates.txt) (*141*). We computed the kinship matrix using GEMMA (gemma -bfile $plink_prefix -gk 1). For eQTL scans we tested all gene - variant pairs where the variant was within 1 Mb of the gene TSS. For caQTL scans we tested all peak - variant pairs where the variant was within 100 kb of the peak center. We normalized the count matrix as follows prior to the scan: 1) Filter to autosomal features, 2) Perform TMM normalization, 3) Filter to features with at least 5 (unnormalized) counts in at least 30% of samples, 4) Exclude features w/o any variants within the above-specified cis window, and 5) Inverse normal transform.

As covariates we included the top N gene expression / chromatin accessibility PCs (PCs calculated on the inverse normal transformed matrix), where N was selected in order to maximize the number of eGenes / caPeaks (this was determined by running (e/ca)QTL scans with 0, 5, 10, 15, … PCs, and then determining the number of eGenes / caPeaks yielded by each scan; the final N was 20 for RNA-seq and 5 for ATAC-seq).

After testing all feature - variant pairs, we performed two levels of multiple testing adjustment. First, we adjusted for the multiple tests performed against each feature using a slightly modified version of eigenMT (*142*) (eigenMT was modified in order to collapse variants in perfect LD before calculating eigenvalues; this modified version is available as git commit c89bf4e at https://github.com/porchard/eigenMT, and was run with arguments eigenMT.py --cis_dist $CIS_WINDOW --var_thresh 0.99 --window 10000 --collapse-redundant-variants --CHROM $chrom --QTL qtl.txt --GEN $genotype_matrix --GENPOS $genpos --PHEPOS $phepos --OUT $output). EigenMT estimates the number of independent tests per feature; we then multiplied the most significant observed *p*-value for each feature by the estimated number of tests (capping the product at 1, if greater than 1) to get a feature-level *p*-value. We then adjusted for the multiple testing across features, performing a Benjamini-Hochberg correction (*121*) on the feature-level *p*-values to get an FDR for each feature. We then applied an FDR threshold of 5% to determine whether each gene or peak had a significant (e/ca)QTL. For each eGene / caPeak, the variant with the most significant *p*-value was set as the eVariant / caVariant (due to linkage disequilibrium, in many cases multiple variants had the same *p*-value; for downstream analyses sensitive to this, we randomly selected one of these variants to be the representative eVariant / caVariant).

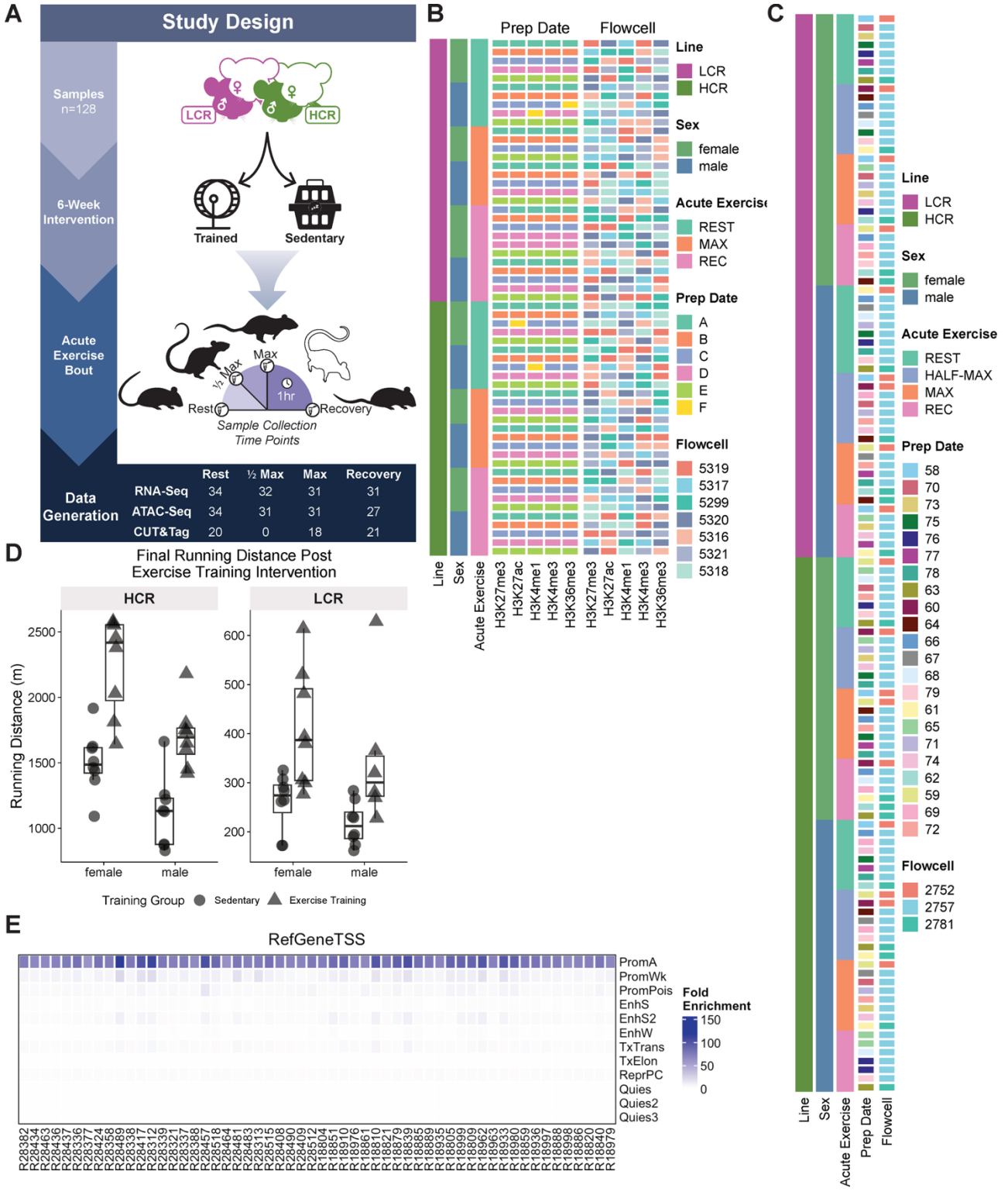


Fig. S1. Detailed Breakdown of Study and Randomization, related to Fig. 1

(**A**) Study design that included 128 HCR and LCR rats, balanced for sex. Half of the cohort underwent a six-week endurance training regimen while the remaining rats were sedentary for that same time period. Rats were split into four sample collection time groups and subject to an acute exercise bout with subsequent tissue collection either prior to, during, or after a near-exhaustion acute exercise bout. High throughput sequencing libraries were generated from gastrocnemius muscle, with the number of libraries produced per-acute exercise bout group provided. Note that CUT&Tag libraries were not produced for samples in the half-max group, and fewer CUT&Tag libraries were produced for the remaining acute exercise bout groups.

(**B**) Randomization schema of CUT&Tag libraries.

(**C**) Randomization schema of ATAC-Seq libraries.

(**D**) To assess the effect of the exercise training, we performed a linear regression using the formula $final\ running\ distance \sim line + sex + exercise\ training\ group$. The results showed that running distance significantly increased following training (β = 404m, *p*-value = 2.97×10^{-07}).

(**E**) QC plot from ChromHMM analysis, showing the enrichment for the Strong Promoter chromatin state at TSSs across the genome. The likeliness of ATAC-Seq and H3K4me3 presence associated with the Strong Promoter state is expected at TSSs.

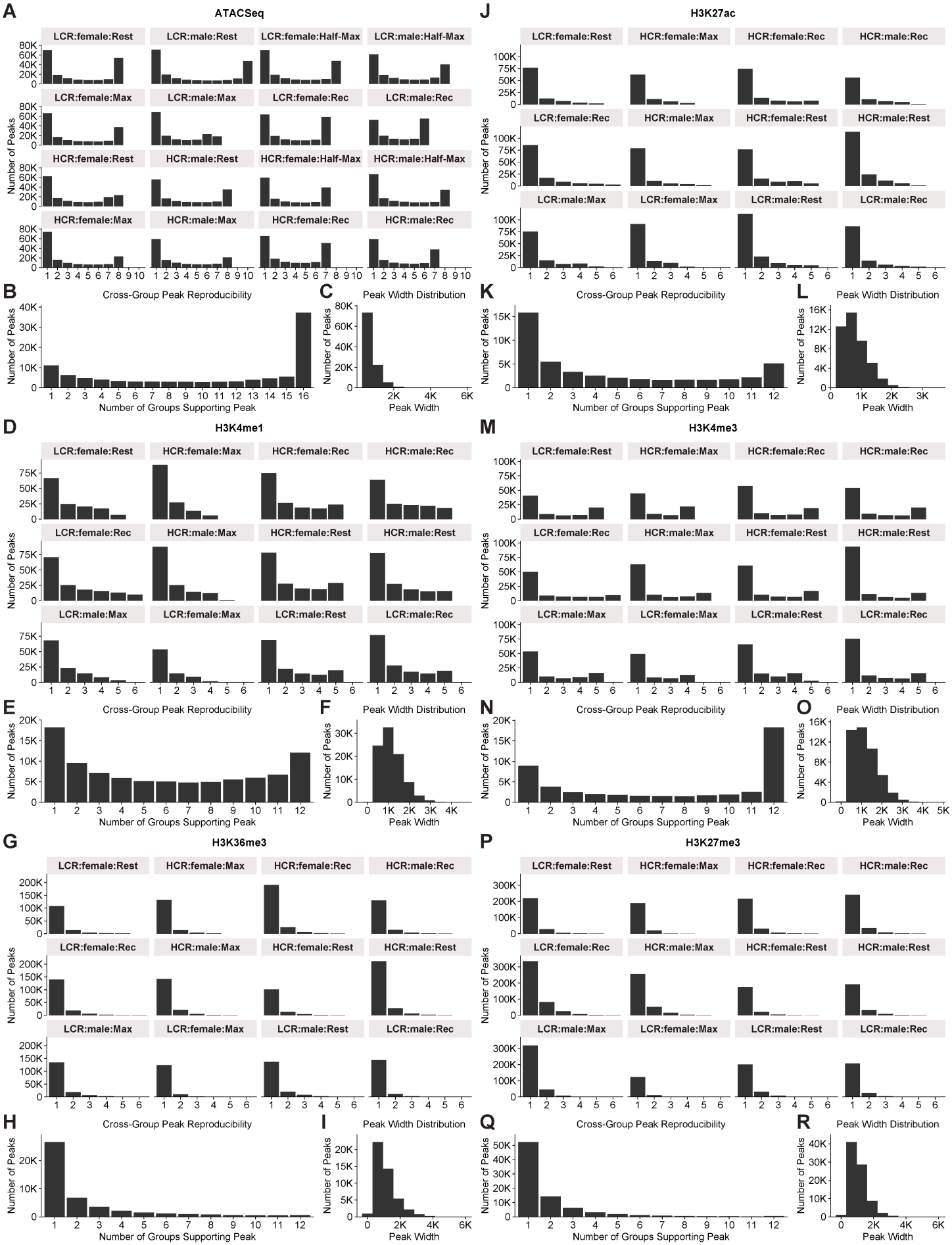
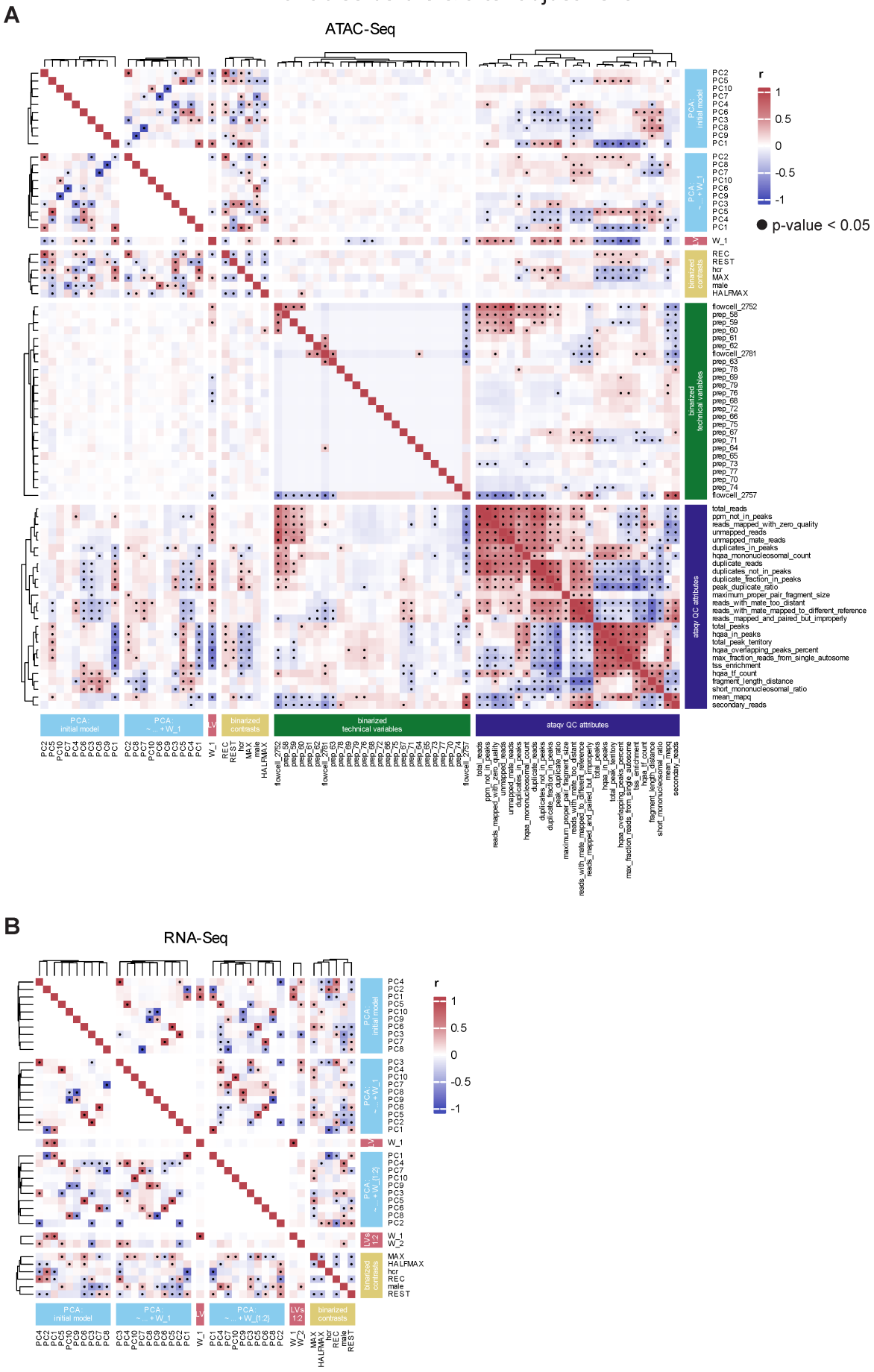

A
ATACSeq
LCR:female:Rest
LCR:male:Rest
LCR:female:Half-Max
LCR:male:Half-Max
LCR:female:Max
LCR:male:Max
LCR:female:Rec
LCR:male:Rec
HCR:female:Rest
HCR:male:Rest
HCR:female:Half-Max
HCR:male:Half-Max
HCR:female:Max
HCR:male:Max
HCR:female:Rec
HCR:male:Rec
Number of Peaks
B
Cross-Group Peak Reproducibility
Number of Groups Supporting Peak
C
Peak Width Distribution
Peak Width
J
H3K27ac
K
L
D
H3K4me1
E
F
M
H3K4me3
N
O
G
H3K36me3
H
I
P
H3K27me3
Q
R

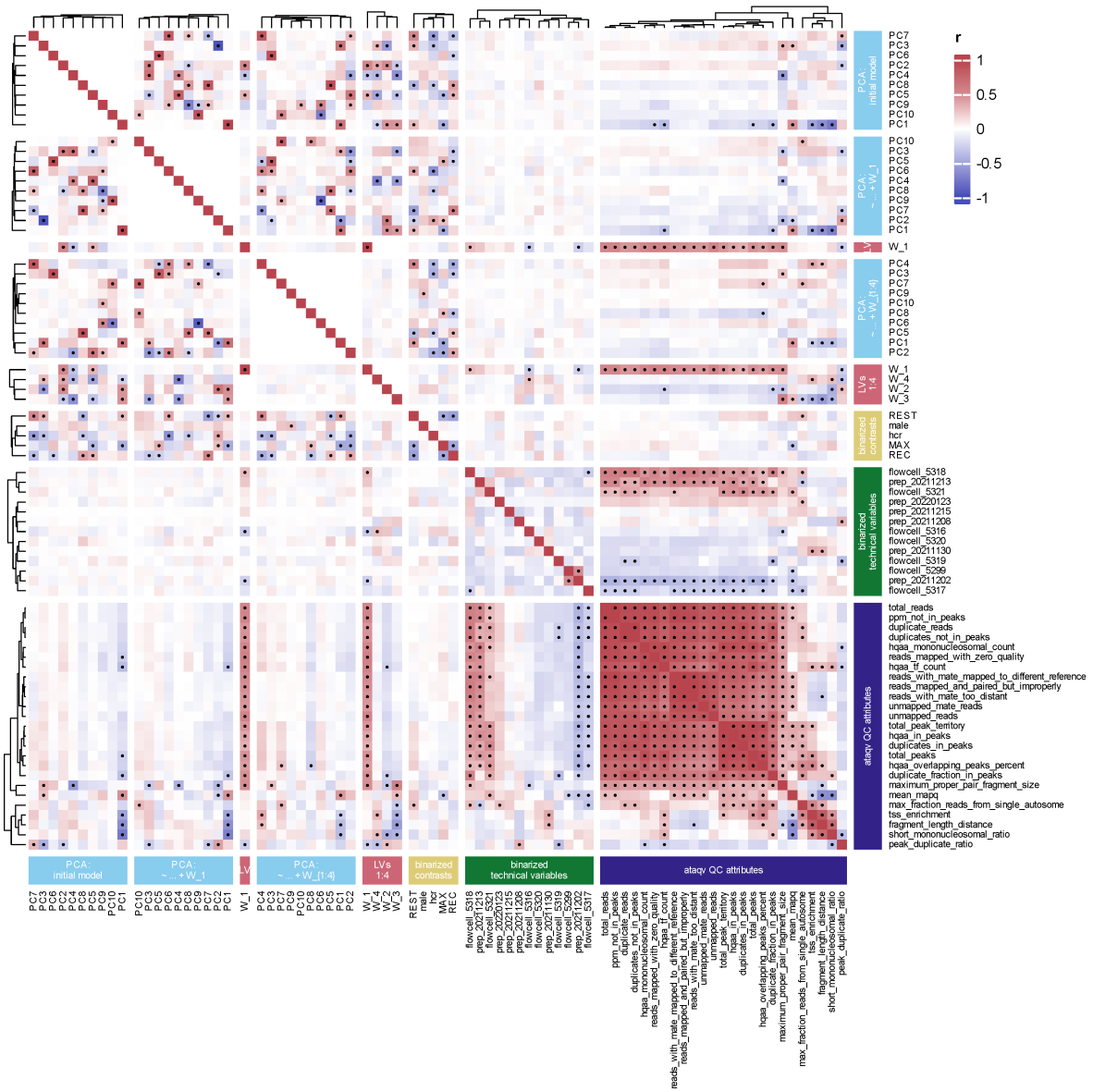
Fig. S2. QC Statistics for the Creation of Master Peak Matrices, related to Fig. 2
(**A-R**) Within-group and cross-group reproducibility of peaks called by MACS2, considered when constructing the master peak matrix for each epigenomic sequencing modality.

Comparison of PCs, latent variables, and experimental variables before & after adjustment

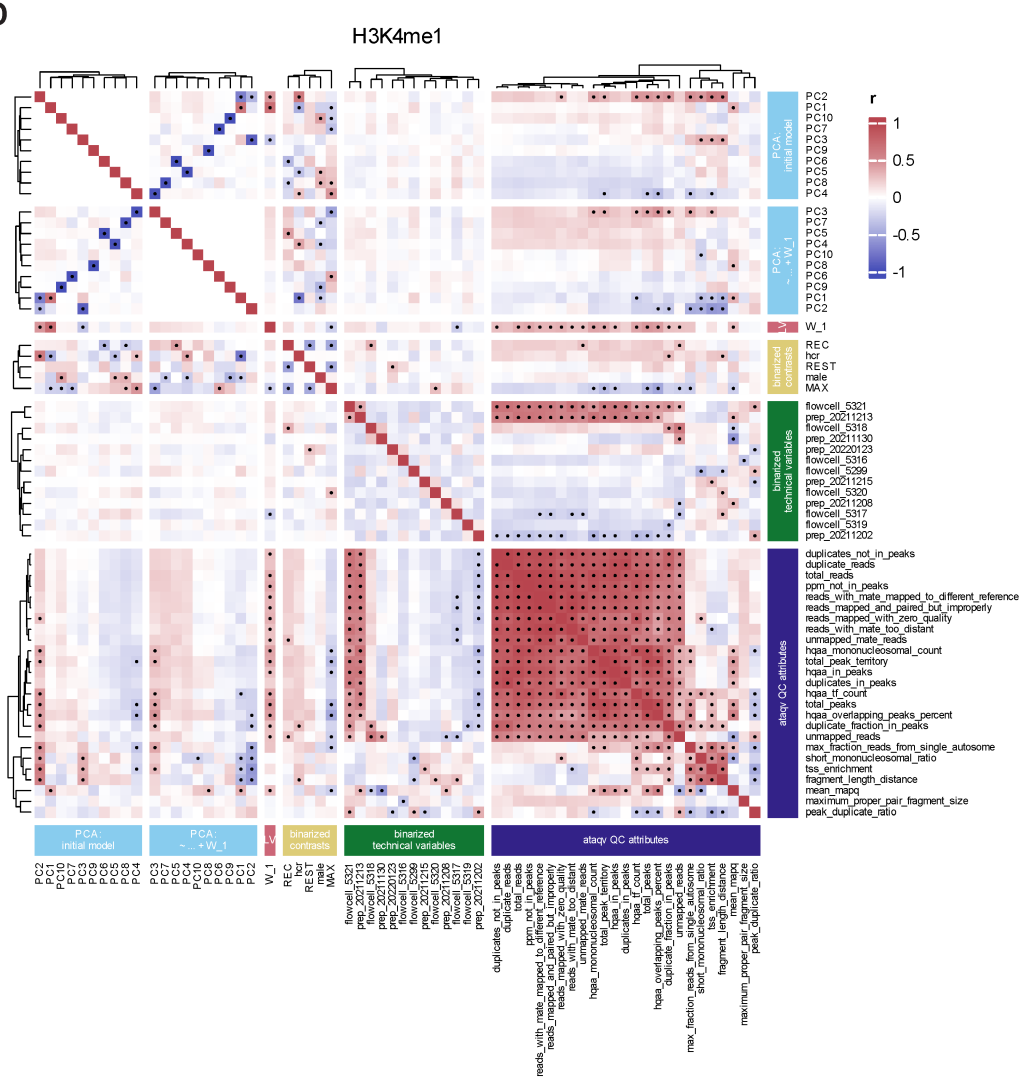


C

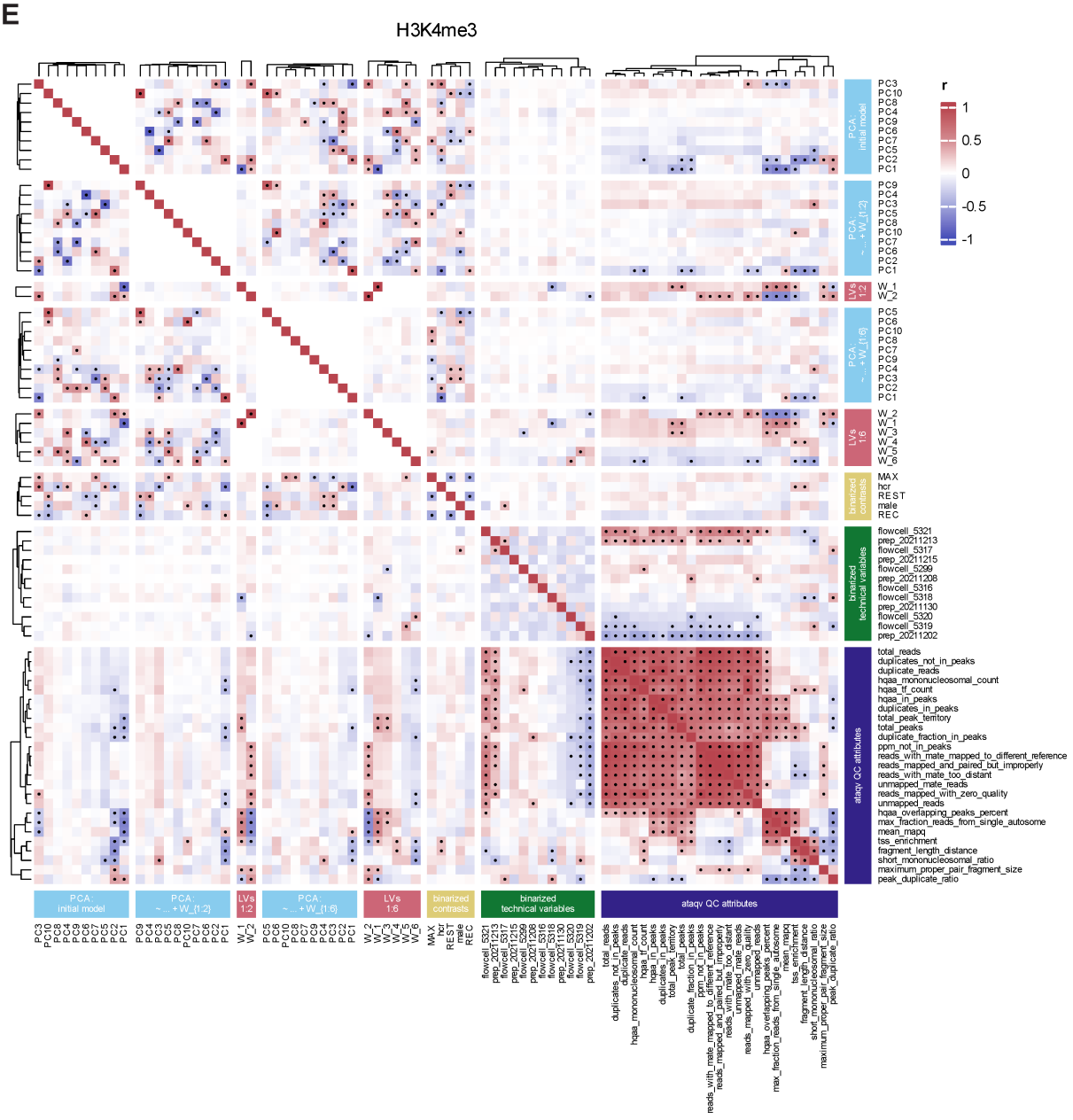
H3K27ac


r
1
0.5
0
-0.5
-1
PCA: initial model
PC7
PC3
PC6
PC2
PC4
PC8
PC5
PC9
PC10
PC1
PCA: ~ ... + W_1
PC10
PC3
PC5
PC6
PC4
PC8
PC9
PC7
PC2
PC1
W_1
PCA: ~ ... + W_{1:4}
PC4
PC3
PC7
PC9
PC10
PC8
PC6
PC5
PC1
PC2
LVs 1:4
W_1
W_4
W_2
W_3
binarized contrasts
REST
male
hcr
MAX
REC
binarized technical variables
flowcell_5318
prep_20211213
flowcell_5321
prep_20220123
prep_20211215
prep_20211208
flowcell_5316
flowcell_5320
prep_20211130
flowcell_5319
flowcell_5299
prep_20211202
flowcell_5317
ataqv QC attributes
total_reads
ppm_not_in_peaks
duplicate_reads
duplicates_not_in_peaks
hqaa_mononucleosomal_count
reads_mapped_with_zero_quality
hqaa_tf_count
reads_with_mate_mapped_to_different_reference
reads_mapped_and_paired_but_improperly
reads_with_mate_too_distant
unmapped_mate_reads
unmapped_reads
total_peak_territory
hqaa_in_peaks
duplicates_in_peaks
total_peaks
hqaa_overlapping_peaks_percent
duplicate_fraction_in_peaks
maximum_proper_pair_fragment_size
mean_mapq
max_fraction_reads_from_single_autosome
tss_enrichment
fragment_length_distance
short_mononucleosomal_ratio
peak_duplicate_ratio

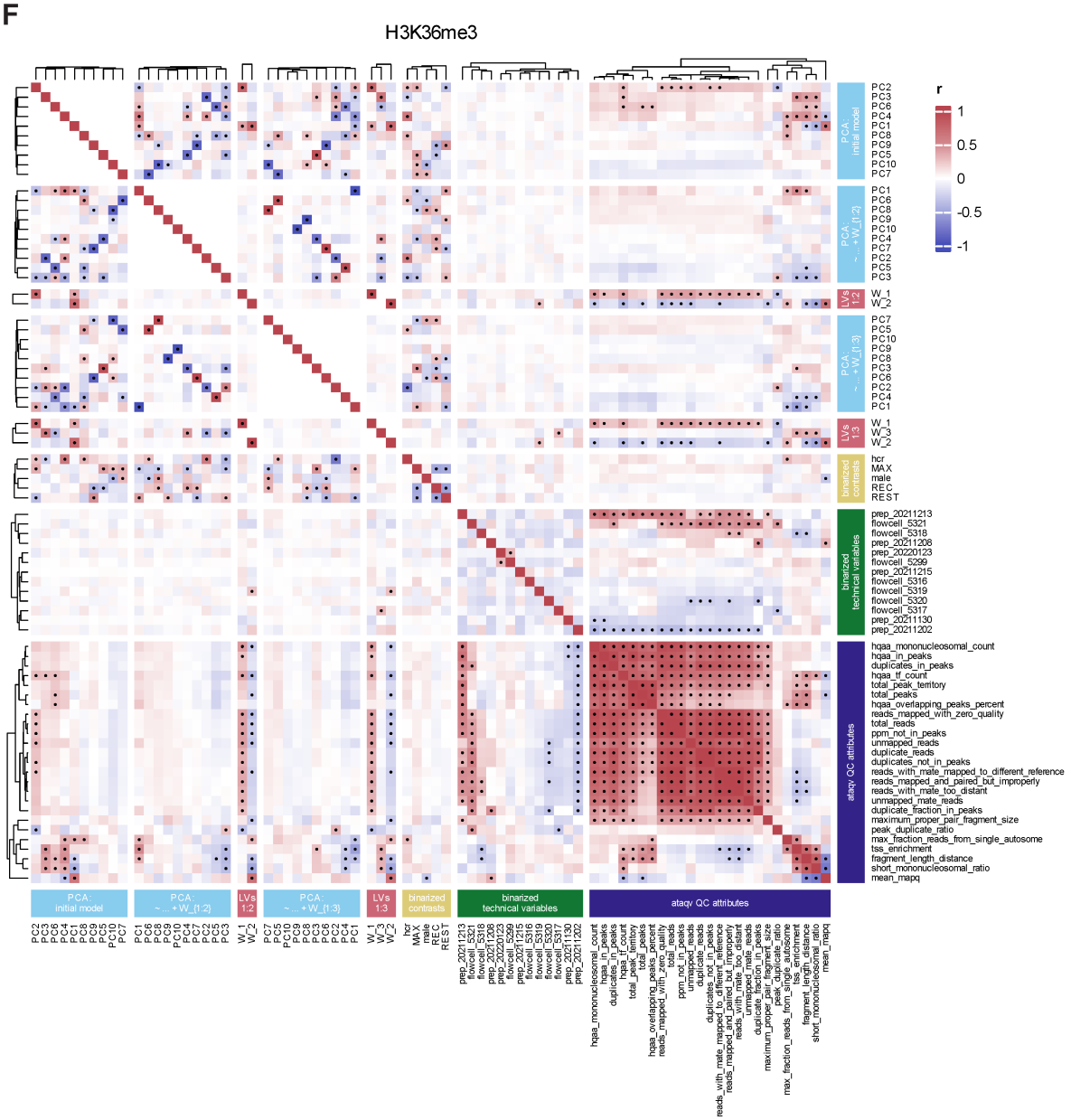
D


H3K4me1
r
1
0.5
0
-0.5
-1
PCA: initial model
PC2
PC1
PC10
PC7
PC3
PC9
PC6
PC5
PC8
PC4
PCA: ~ ... + W_1
PC3
PC7
PC5
PC4
PC10
PC8
PC6
PC9
PC1
PC2
W
W_1
binarized contrasts
REC
hcr
REST
male
MAX
binarized technical variables
flowcell_5321
prep_20211213
flowcell_5318
prep_20211130
prep_20220123
flowcell_5316
flowcell_5299
prep_20211215
flowcell_5320
prep_20211208
flowcell_5317
flowcell_5319
prep_20211202
ataqv QC attributes
duplicates_not_in_peaks
duplicate_reads
total_reads
ppm_not_in_peaks
reads_with_mate_mapped_to_different_reference
reads_mapped_and_paired_but_improperly
reads_mapped_with_zero_quality
reads_with_mate_too_distant
unmapped_mate_reads
hqaa_mononucleosomal_count
total_peak_territory
hqaa_in_peaks
duplicates_in_peaks
hqaa_tf_count
total_peaks
hqaa_overlapping_peaks_percent
duplicate_fraction_in_peaks
unmapped_reads
max_fraction_reads_from_single_autosome
short_mononucleosomal_ratio
tss_enrichment
fragment_length_distance
mean_mapq
maximum_proper_pair_fragment_size
peak_duplicate_ratio

E


H3K4me3
r
1
0.5
0
-0.5
-1
PCA: initial model
PCA: ~ ... + W_{1:2}
LVs 1:2
PCA: ~ ... + W_{1:6}
LVs 1:6
binarized contrasts
binarized technical variables
ataqv QC attributes
PC3
PC10
PC8
PC4
PC9
PC6
PC7
PC5
PC2
PC1
PC9
PC4
PC3
PC5
PC8
PC10
PC7
PC6
PC2
PC1
W_1
W_2
PC5
PC6
PC10
PC8
PC7
PC9
PC4
PC3
PC2
PC1
W_2
W_1
W_3
W_4
W_5
W_6
MAX
hcr
REST
male
REC
flowcell_5321
prep_20211213
flowcell_5317
prep_20211215
flowcell_5299
prep_20211208
flowcell_5316
flowcell_5318
prep_20211130
flowcell_5320
flowcell_5319
prep_20211202
total_reads
duplicates_not_in_peaks
duplicate_reads
hqaa_mononucleosomal_count
hqaa_tf_count
hqaa_in_peaks
duplicates_in_peaks
total_peak_territory
total_peaks
duplicate_fraction_in_peaks
ppm_not_in_peaks
reads_with_mate_mapped_to_different_reference
reads_mapped_and_paired_but_improperly
reads_with_mate_too_distant
unmapped_mate_reads
reads_mapped_with_zero_quality
unmapped_reads
hqaa_overlapping_peaks_percent
max_fraction_reads_from_single_autosome
mean_mapq
tss_enrichment
fragment_length_distance
short_mononucleosomal_ratio
maximum_proper_pair_fragment_size
peak_duplicate_ratio

F


H3K36me3
r
1
0.5
0
-0.5
-1
PCA: initial model
PCA: ~ ... + W_{1:2}
LVs 1:2
PCA: ~ ... + W_{1:3}
LVs 1:3
binarized contrasts
binarized technical variables
ataqv QC attributes
PC2
PC3
PC6
PC4
PC1
PC8
PC9
PC5
PC10
PC7
W_1
W_2
W_3
hcr
MAX
male
REC
REST
prep_20211213
flowcell_5321
flowcell_5318
prep_20211208
prep_20220123
flowcell_5299
prep_20211215
flowcell_5316
flowcell_5319
flowcell_5320
flowcell_5317
prep_20211130
prep_20211202
hqaa_mononucleosomal_count
hqaa_in_peaks
duplicates_in_peaks
hqaa_tf_count
total_peak_territory
total_peaks
hqaa_overlapping_peaks_percent
reads_mapped_with_zero_quality
total_reads
ppm_not_in_peaks
unmapped_reads
duplicate_reads
duplicates_not_in_peaks
reads_with_mate_mapped_to_different_reference
reads_mapped_and_paired_but_improperly
reads_with_mate_too_distant
unmapped_mate_reads
duplicate_fraction_in_peaks
maximum_proper_pair_fragment_size
peak_duplicate_ratio
max_fraction_reads_from_single_autosome
tss_enrichment
fragment_length_distance
short_mononucleosomal_ratio
mean_mapq

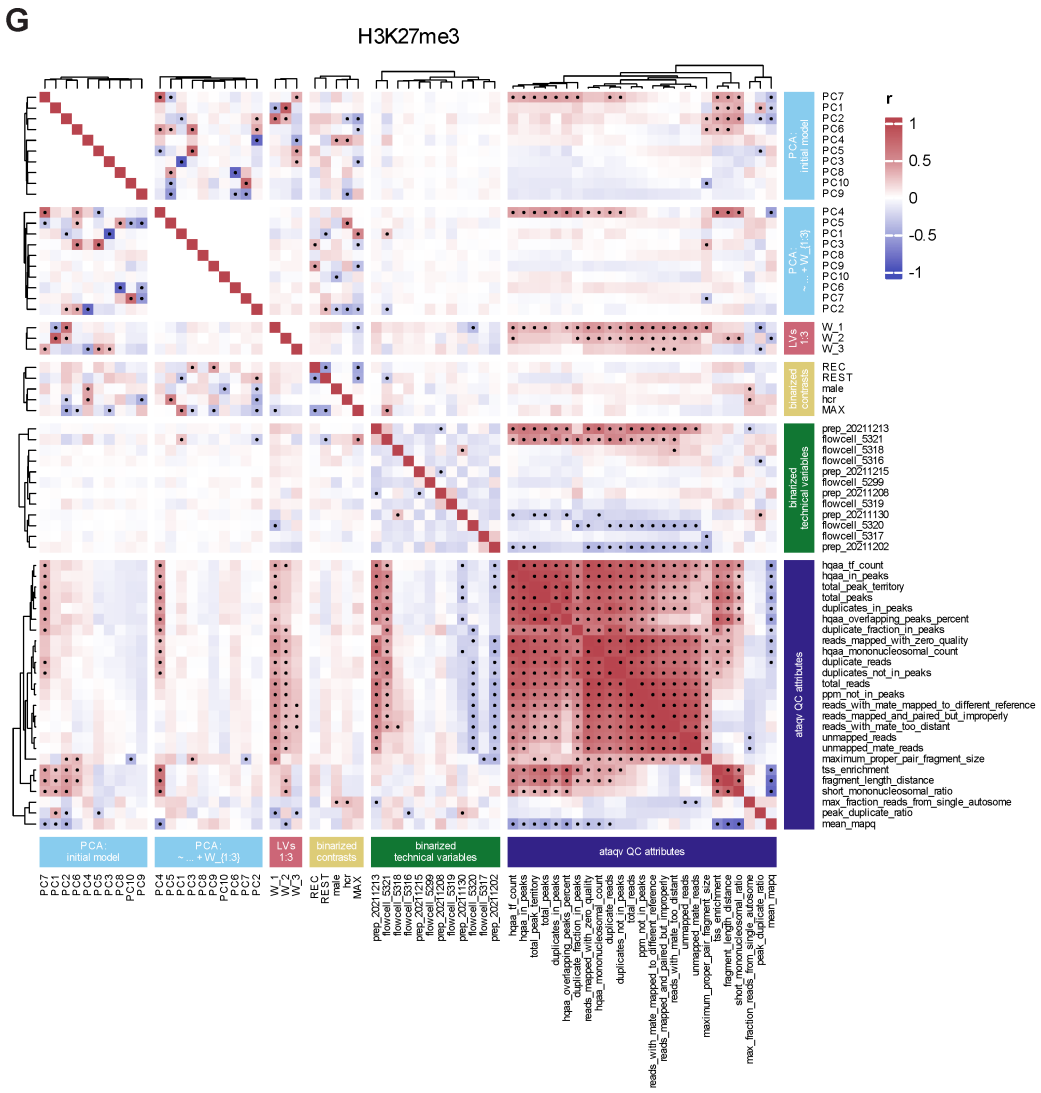


Fig. S3. Comparison of PCs, Latent Variables, and Experimental Variables Before and After Adjustment of ATAC-Seq, CUT&Tag, and RNA-Seq Data, related to Fig. 2

(**A-G**) Heatmap of correlation between binarized experimental contrasts (yellow), considered latent variables (red), and PCs before and after inclusion of latent variables (light blue), with each other and QC metrics associated with epigenomic sequencing data (dark blue).

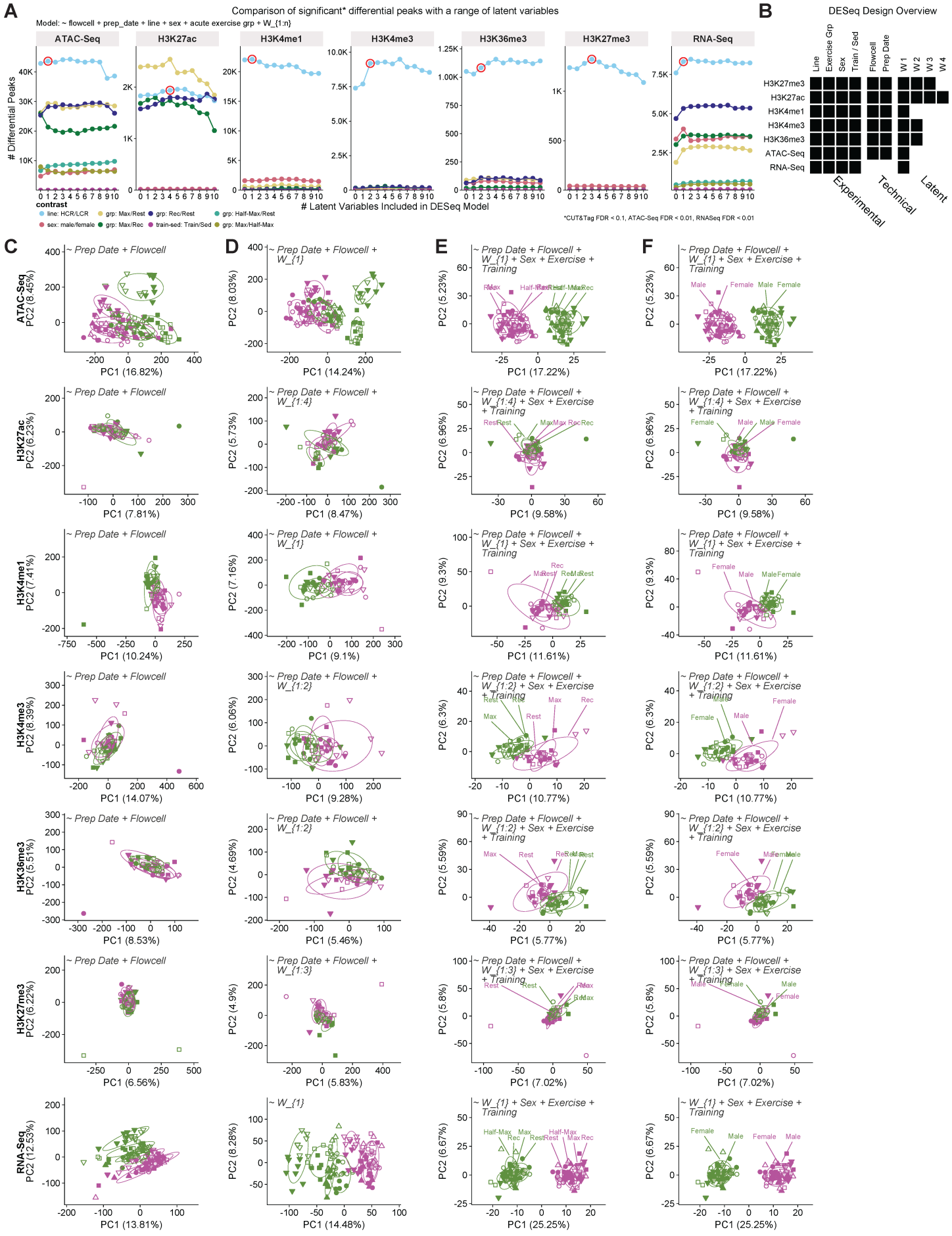

A
Comparison of significant* differential peaks with a range of latent variables
Model: ~ flowcell + prep_date + line + sex + acute exercise grp + W_{1:n}
ATAC-Seq
H3K27ac
H3K4me1
H3K4me3
H3K36me3
H3K27me3
RNA-Seq
Differential Peaks
Latent Variables Included in DESeq Model
contrast
line: HCR/LCR
grp: Max/Rest
grp: Rec/Rest
grp: Half-Max/Rest
sex: male/female
grp: Max/Rec
train-sed: Train/Sed
grp: Max/Half-Max
*CUT&Tag FDR < 0.1, ATAC-Seq FDR < 0.01, RNASeq FDR < 0.01
B
DESeq Design Overview
Line
Exercise Grp
Sex
Train / Sed
Flowcell
Prep Date
W 1
W 2
W 3
W 4
Experimental
Technical
Latent
C
D
E
F
~ Prep Date + Flowcell
~ Prep Date + Flowcell + W_{1}
~ Prep Date + Flowcell + W_{1} + Sex + Exercise + Training
~ Prep Date + Flowcell + W_{1:4}
~ Prep Date + Flowcell + W_{1:4} + Sex + Exercise + Training
~ Prep Date + Flowcell + W_{1:2}
~ Prep Date + Flowcell + W_{1:2} + Sex + Exercise + Training
~ Prep Date + Flowcell + W_{1:3}
~ Prep Date + Flowcell + W_{1:3} + Sex + Exercise + Training
~ W_{1}
~ W_{1} + Sex + Exercise + Training
ATAC-Seq PC2 (8.45%)
PC1 (16.82%)
PC2 (8.03%)
PC1 (14.24%)
PC2 (5.23%)
PC1 (17.22%)
H3K27ac PC2 (6.23%)
PC1 (7.81%)
PC2 (5.73%)
PC1 (8.47%)
PC2 (6.96%)
PC1 (9.58%)
H3K4me1 PC2 (7.41%)
PC1 (10.24%)
PC2 (7.16%)
PC1 (9.1%)
PC2 (9.3%)
PC1 (11.61%)
H3K4me3 PC2 (8.39%)
PC1 (14.07%)
PC2 (6.06%)
PC1 (9.28%)
PC2 (6.3%)
PC1 (10.77%)
H3K36me3 PC2 (5.51%)
PC1 (8.53%)
PC2 (4.69%)
PC1 (5.46%)
PC2 (5.59%)
PC1 (5.77%)
H3K27me3 PC2 (6.22%)
PC1 (6.56%)
PC2 (4.9%)
PC1 (5.83%)
PC2 (5.8%)
PC1 (7.02%)
RNA-Seq PC2 (12.53%)
PC1 (13.81%)
PC2 (8.28%)
PC1 (14.48%)
PC2 (6.67%)
PC1 (25.25%)
Line Sex
LCR F
LCR M
HCR F
HCR M
Exercise
REST
HALF-MAX
MAX
REC

Fig. S4. Identification of Latent Variables, DESeq2 Models, and Variation of Data Before and After Adjustment, related to Fig. 2

(**A**) Comparison of the count of significant differential peaks (y-axis) resulting between all experimental contrasts (colored) when a range of latent variables are included in the DESeq model (x-axis). The number of latent variables that we chose to include in all subsequent models (e.g., DESeq, adjusting of VST transformed counts for correlation and data visualization) for each modality is circled in red.

(**B**) DESeq design overview for each modality subject to differential analysis by this method.

(**C**) PCA plots of VST transformed counts for each modality normalized for technical variables by adjusting based on the formula printed on each plot.

(**D**) PCA plots of VST transformed counts for each modality normalized for technical and latent variables (formula printed on the plot). This shows how outliers are rescued, and group clustering is improved by normalizing for unknown variation. Additionally, these plots show how strong the line effect is relative to acute exercise and sex prior to the removal of this effect.

(**E-F**) VST transformed counts where the line effect was isolated by adjusting for additional experimental (sex, acute exercise, and training) and technical variables (prep batch, sequencing batch, latent variables) via linear regression. The linear model included the line variable to protect this effect. These plots confirm that acute exercise (**E**) and sex groups (**F**) are centered post-removal of the additional experimental and technical contrasts, exemplifying successful isolation of the HCR/LCR effect in this data. Centering of training samples are not displayed because that experimental contrast resulted in minimal molecular changes.

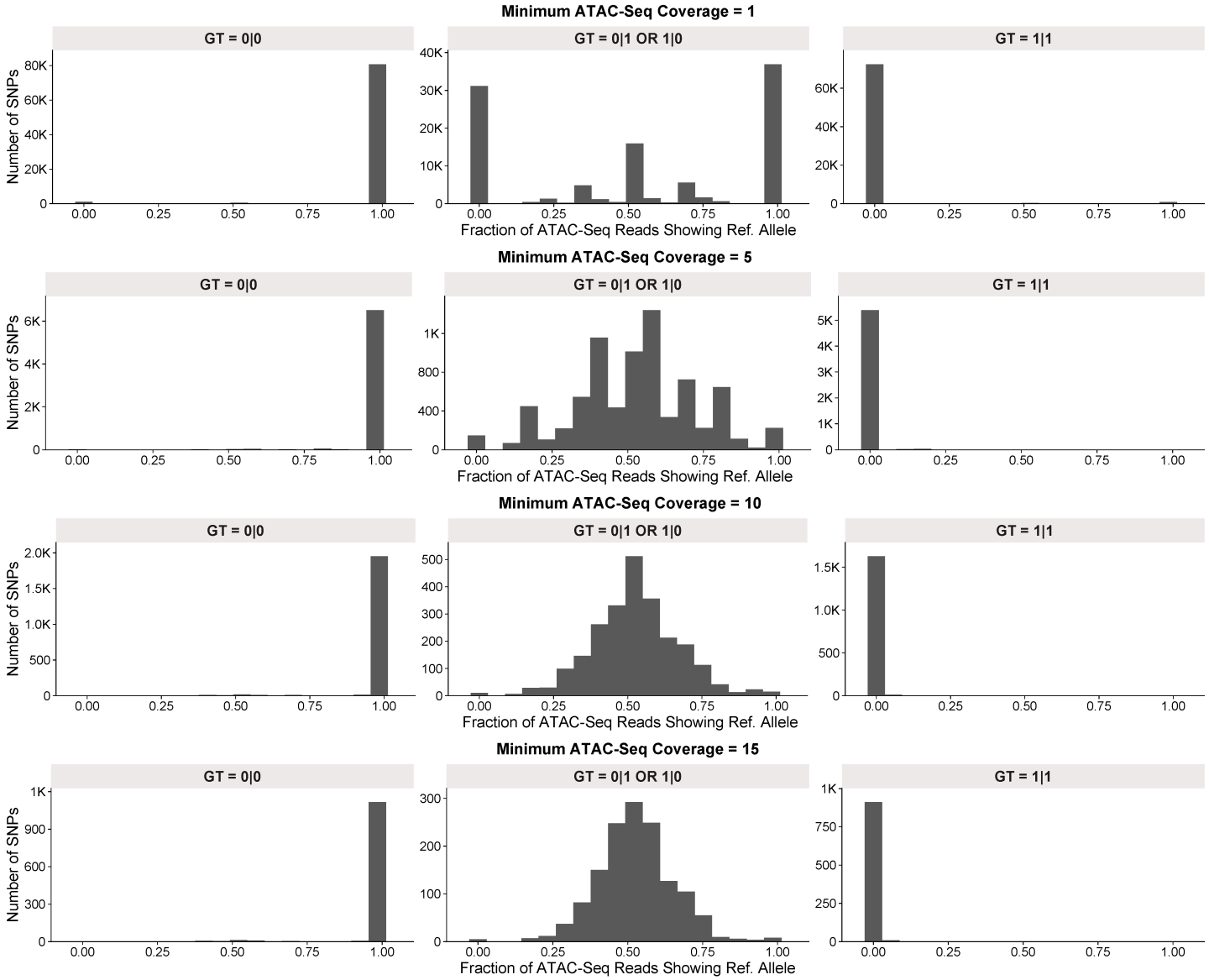


Fig. S5. Genotype Imputation QC, related to Fig. 5

Distribution plots showing the proportion of reference alleles present in ATAC-Seq data at variant sites imputed as homozygous reference (0|0), heterozygous (1|0 OR 0|1), and homozygous alternate (1|1) at varying ATAC-Seq depth thresholds. Reference alleles are predominantly observed at sites imputed as homozygous reference, and vice versa at sites imputed as homozygous alternate. For sites imputed as heterozygous, the distribution of observed genotypes is centered around 0.5.

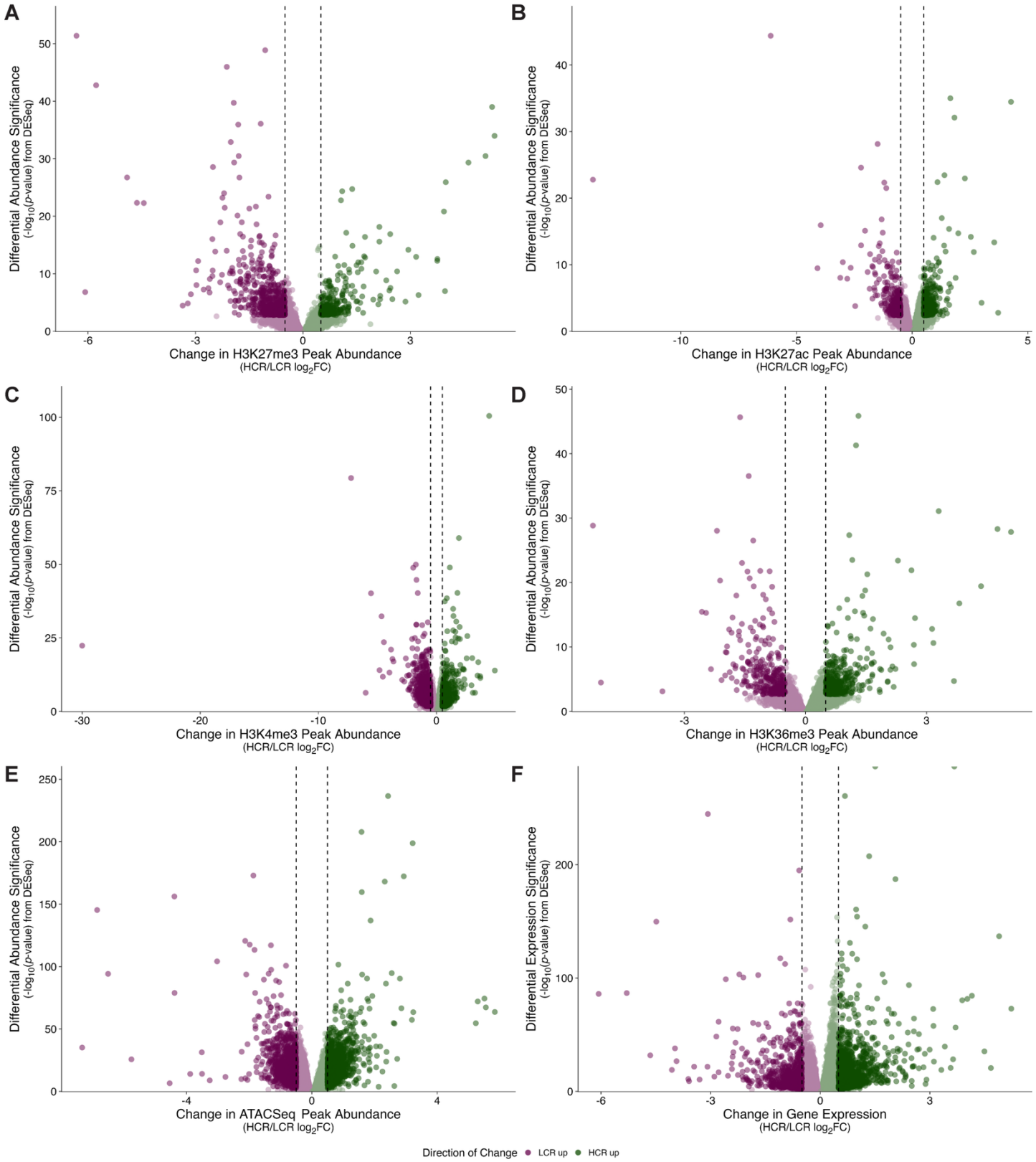


Fig. S6. Differential Chromatin Accessibility, Histone Modification Abundance, and Gene Expression, related to Fig. 2

Volcano plots created from the HCR/LCR DESeq output, with emphasis on the regions that pass the significance threshold provided in Fig. 2C for each additional sequencing-based modality: (**A**) H3K27me3, (**B**) H3K27ac, (**C**) H3K4me3, (**D**) H3K36me3, (**E**) ATAC-Seq, (**F**) RNA-Seq (gene expression).

Fig. S7. Differential RNA-Seq Pathway Enrichment Results, related to Fig. 2

RNA-Seq pathway enrichment analysis. HCR/LCR differential gene expression significantly enriched for 342 gene sets via RNA-Enrich analysis (FDR < 0.05), which were collapsed into 149 representative clusters and

plotted separately by upregulation in HCR (**A**) and LCR (**B**). The rows are hierarchically clustered based on the $\log_2$ odds ratio of the RNA-Enrich results plotted. The number of gene sets within the hierarchical cluster is annotated on the lefthand side of the heatmap. The largest cluster in panel (**A**), mitochondrial protein-containing complex, is expanded upon to the right to show the individual terms a part of that cluster.

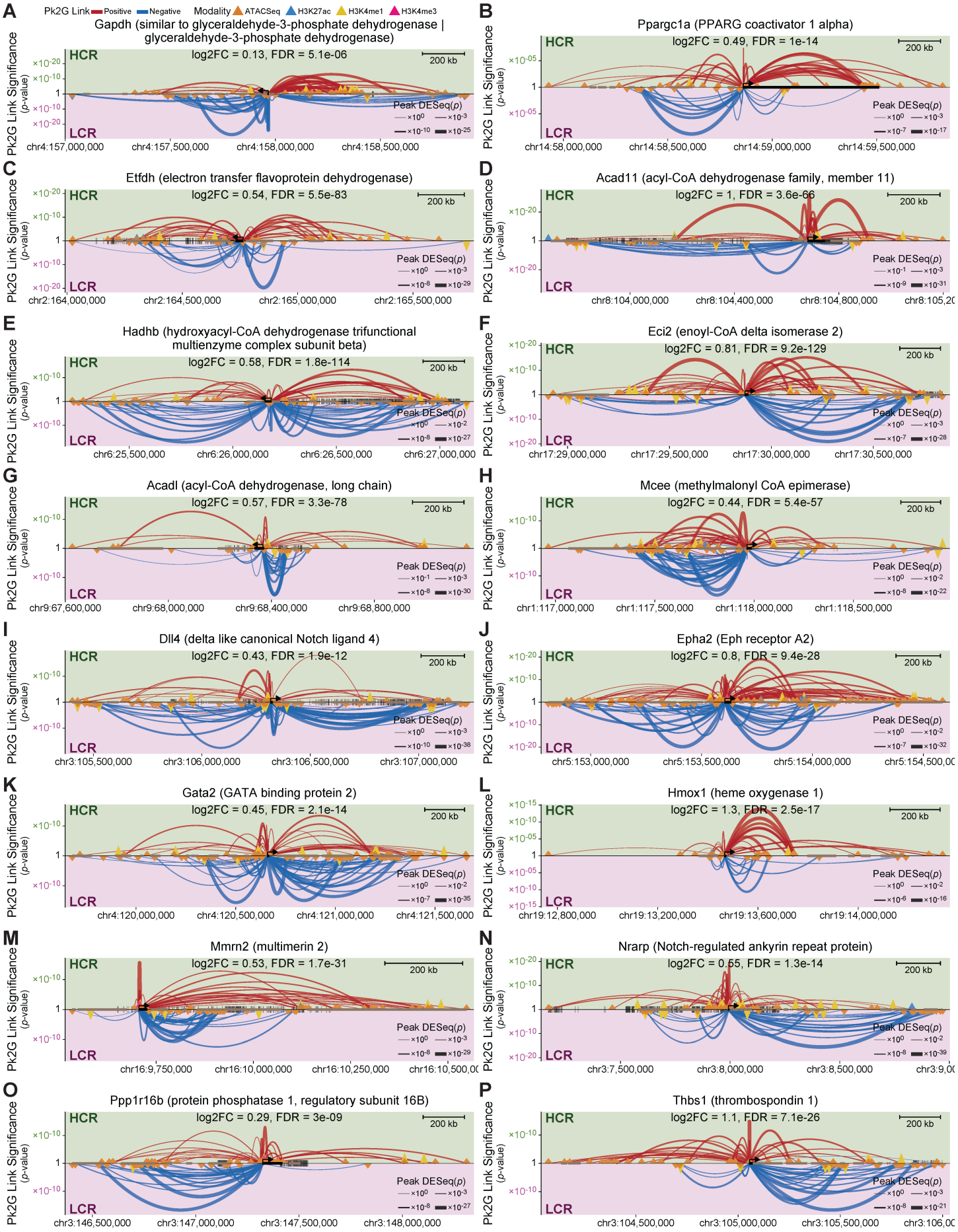

A
Pk2G Link Positive Negative
Modality ATACSeq H3K27ac H3K4me1 H3K4me3
Gapdh (similar to glyceraldehyde-3-phosphate dehydrogenase | glyceraldehyde-3-phosphate dehydrogenase)
log2FC = 0.13, FDR = 5.1e-06
HCR
LCR
200 kb
Peak DESeq(p)
Pk2G Link Significance (p-value)
chr4:157,000,000 chr4:157,500,000 chr4:158,000,000 chr4:158,500,000
B
Ppargc1a (PPARG coactivator 1 alpha)
log2FC = 0.49, FDR = 1e-14
chr14:58,000,000 chr14:58,500,000 chr14:59,000,000 chr14:59,500,000
C
Etfdh (electron transfer flavoprotein dehydrogenase)
log2FC = 0.54, FDR = 5.5e-83
chr2:164,000,000 chr2:164,500,000 chr2:165,000,000 chr2:165,500,000
D
Acad11 (acyl-CoA dehydrogenase family, member 11)
log2FC = 1, FDR = 3.6e-66
chr8:104,000,000 chr8:104,400,000 chr8:104,800,000
E
Hadhb (hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit beta)
log2FC = 0.58, FDR = 1.8e-114
chr6:25,500,000 chr6:26,000,000 chr6:26,500,000 chr6:27,000,000
F
Eci2 (enoyl-CoA delta isomerase 2)
log2FC = 0.81, FDR = 9.2e-129
chr17:29,000,000 chr17:29,500,000 chr17:30,000,000 chr17:30,500,000
G
Acadl (acyl-CoA dehydrogenase, long chain)
log2FC = 0.57, FDR = 3.3e-78
chr9:67,600,000 chr9:68,000,000 chr9:68,400,000 chr9:68,800,000
H
Mcee (methylmalonyl CoA epimerase)
log2FC = 0.44, FDR = 5.4e-57
chr1:117,000,000 chr1:117,500,000 chr1:118,000,000 chr1:118,500,000
I
Dll4 (delta like canonical Notch ligand 4)
log2FC = 0.43, FDR = 1.9e-12
chr3:105,500,000 chr3:106,000,000 chr3:106,500,000 chr3:107,000,000
J
Epha2 (Eph receptor A2)
log2FC = 0.8, FDR = 9.4e-28
chr5:153,000,000 chr5:153,500,000 chr5:154,000,000
K
Gata2 (GATA binding protein 2)
log2FC = 0.45, FDR = 2.1e-14
chr4:120,000,000 chr4:120,500,000 chr4:121,000,000 chr4:121,500,000
L
Hmox1 (heme oxygenase 1)
log2FC = 1.3, FDR = 2.5e-17
chr19:12,800,000 chr19:13,200,000 chr19:13,600,000 chr19:14,000,000
M
Mmrn2 (multimerin 2)
log2FC = 0.53, FDR = 1.7e-31
chr16:9,750,000 chr16:10,000,000 chr16:10,250,000
N
Nrarp (Notch-regulated ankyrin repeat protein)
log2FC = 0.55, FDR = 1.3e-14
chr3:7,500,000 chr3:8,000,000 chr3:8,500,000
O
Ppp1r16b (protein phosphatase 1, regulatory subunit 16B)
log2FC = 0.29, FDR = 3e-09
chr3:146,500,000 chr3:147,000,000 chr3:147,500,000 chr3:148,000,000
P
Thbs1 (thrombospondin 1)
log2FC = 1.1, FDR = 7.1e-26
chr3:104,500,000 chr3:105,000,000 chr3:105,500,000

Fig. S8. Pk2G Examples, related to Fig. 3

Peak-gene regulatory architecture surrounding (**A**) housekeeping gene *Gapdh*, (**B**) mitochondrial biogenesis regulator *Ppargc1a*, (**C-H**) fatty acid oxidation genes, and (**I-P**) angiogenesis-related genes. Regulatory elements (triangles, colored by modality) are positioned by genomic distance from the TSS (arrow) of the gene (black rectangle). Peaks with DESeq $\log_2$FC > 0 (HCR up) are located above the midline, and LCR up peaks are shown below. Arc height denotes the significance of the Pk2G links, colored by direction of Pk2G link, with linewidth proportional to DESeq significance of the regulatory element. Vertical ticks at the midline represent SNPs with $F_{ST} > 0.3$ and additional rectangles represent other genes within view.

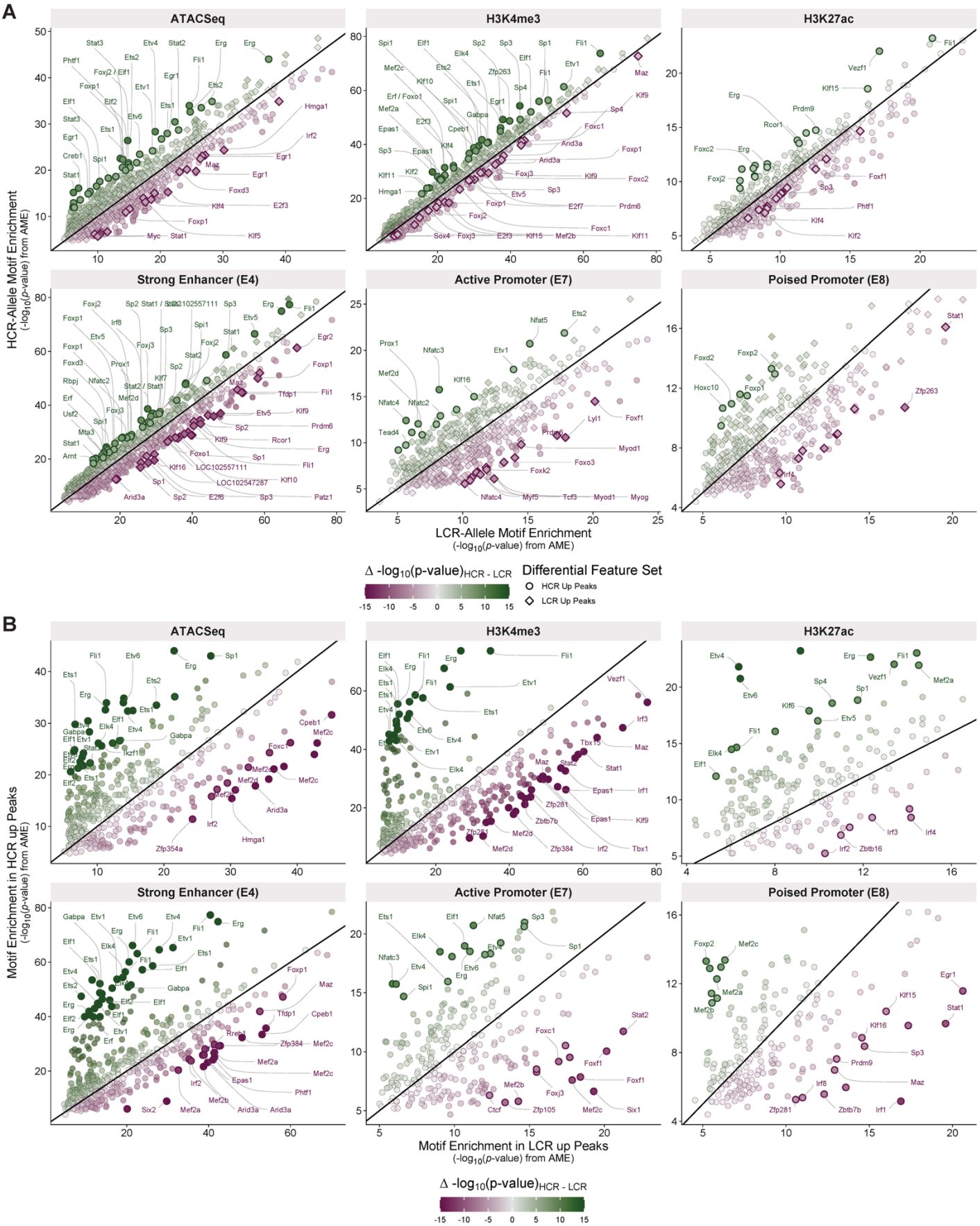


Fig. S9. Differential Motif Enrichment, related to Fig. 4 and Fig. 5

Comparison of motif enrichment from the (**A**) allele-specific and (**B**) global differential motif enrichment analysis across HCR up and LCR up peaks, colored in the same manner as the main-text figures.

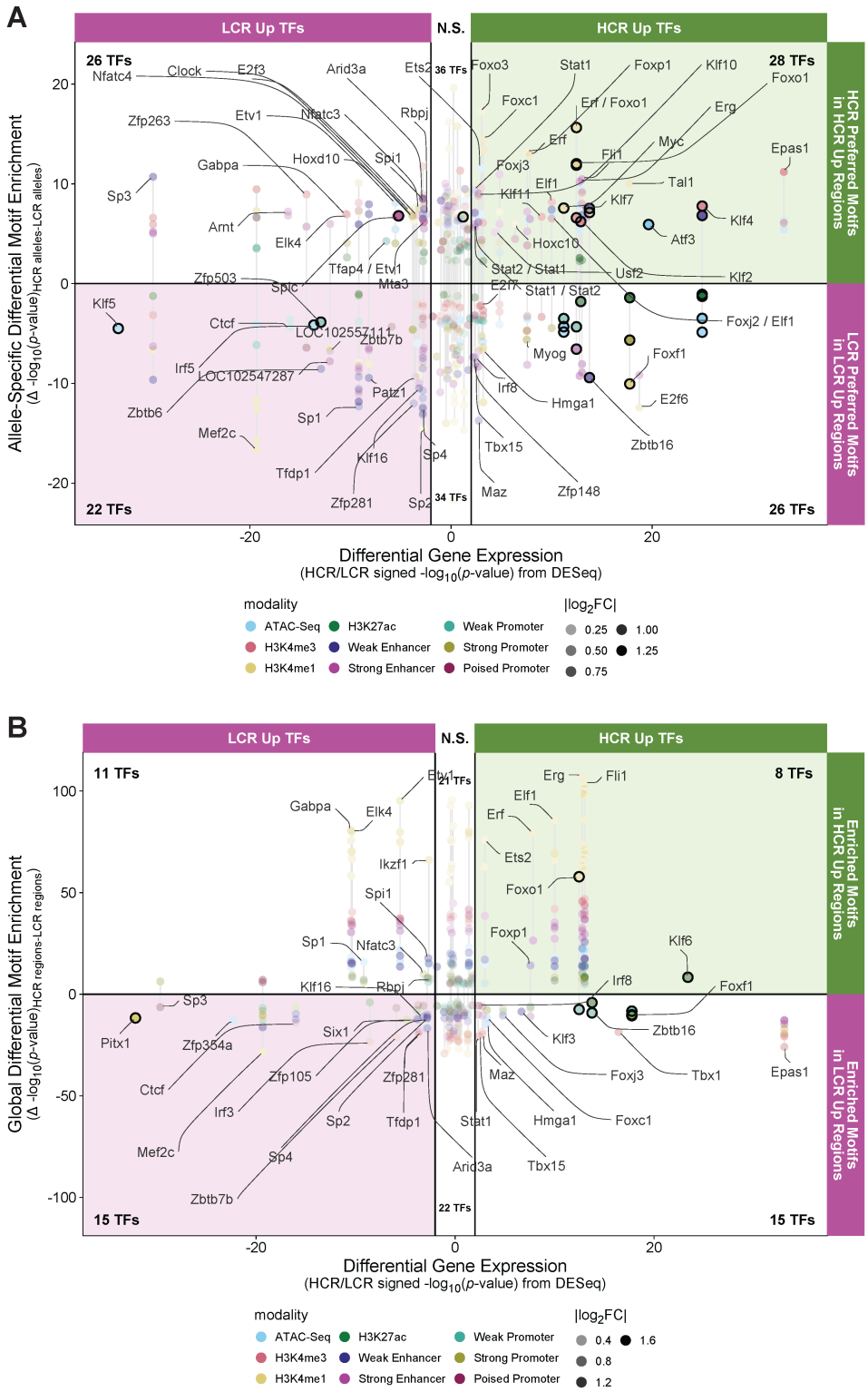


Fig. S10. Relationship Between Differential Motifs and Transcription Factor Gene Expression, related to Fig. 4 and Fig. 5

Comparison of the differential motif enrichment significance (signed Δ -$\log_{10}$(AME *p*-value)) across all modalities to the significance of differential TF gene expression (HCR/LCR). The significance of differential gene expression is represented as signed -$\log_{10}$(FDR) from DESeq. TFs that are differentially enriched in multiple modalities are connected by a vertical gray line, indicating multiple AME *p*-values with consistent DESeq significance. Black vertical lines at -2 and 2 correspond to a DESeq FDR threshold of < 0.01. The fill intensity represents the magnitude of change ($|\log_2 FC|$). Points with $|\log_2 FC| > 0.5$ are outlined in black. (**A**) Comparison specific to the allele-specific analysis. (**B**) Comparison specific to the global motif enrichment analysis.

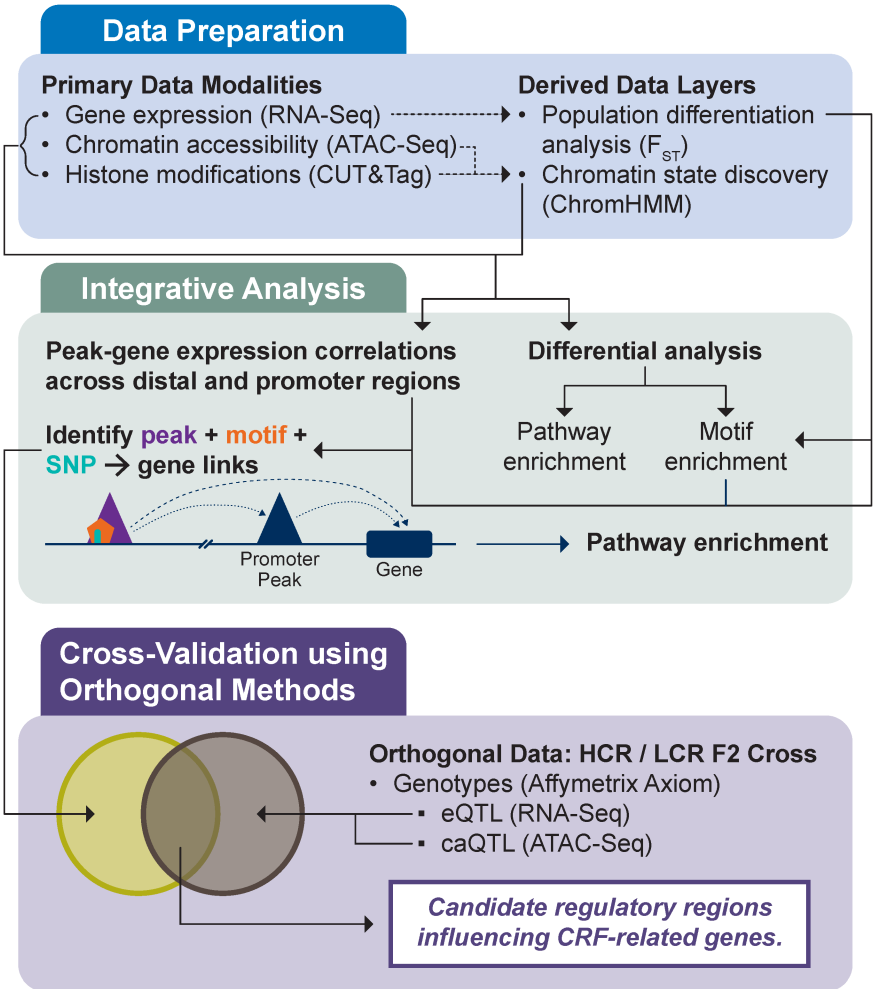


Fig. S11. Multi-Modal Data Framework
Overview of the data modalities included in the study, categorized by whether the data used is primary or derived. Primary data modalities are sequencing libraries prepared from skeletal muscle tissue. Derived data layers were generated by utilizing the primary data modalities. The green box shows the workflow of analyzing various datasets independently and then integrated to identify variant:peak:gene links. The purple box presents the concept of the validation study.

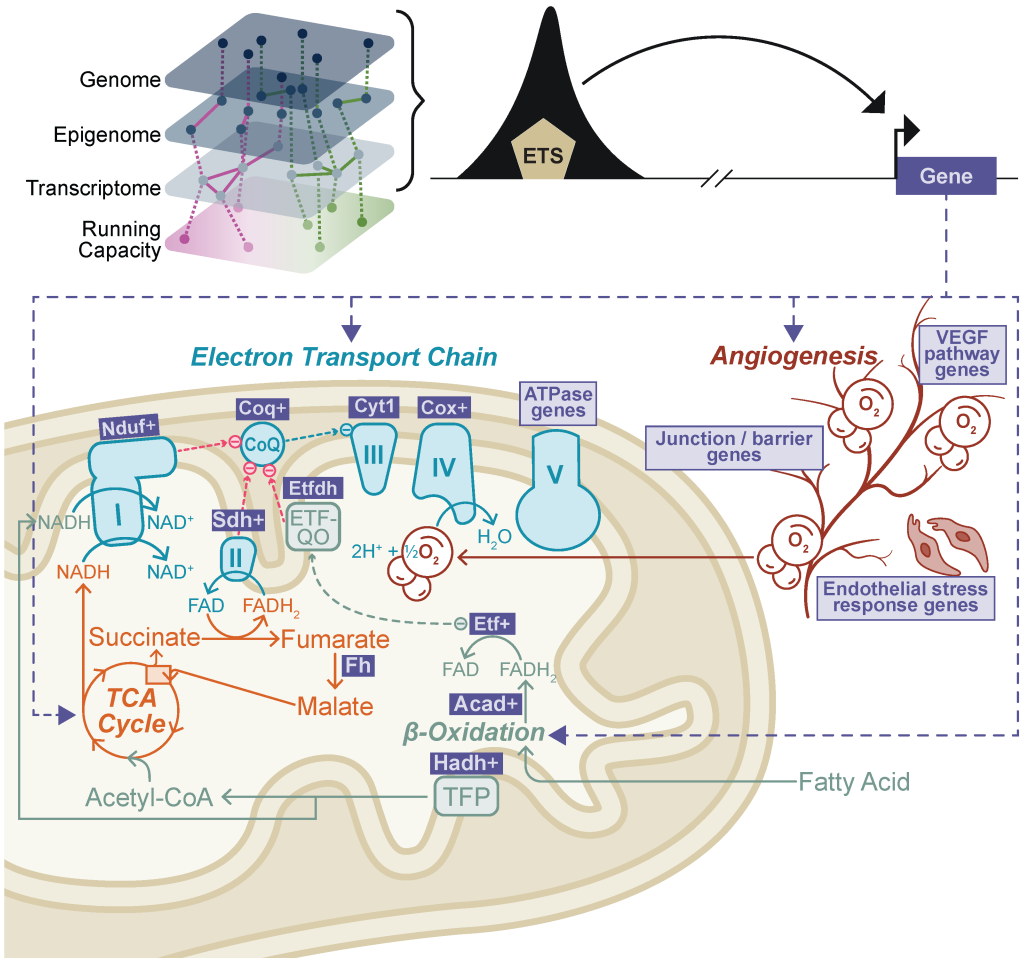


Fig. S12. ETS-mediated regulation of mitochondrial energy metabolism genes supports oxidative efficiency in high CRF muscle

Cartoon representation of ETS TF motifs in enhancer regions upregulating genes involved in fatty acid oxidation (ACADs, *Hadh*, *Etfa*/*Etfb*, *Etfdh*), TCA flux (*Fh*: $\log_2$FC = 0.68, FDR = 1.47×10^{-257}, rank = 3), electron transport (subunits of complexes I-V), and endothelial pathways, which all support efficient oxidative metabolism in HCR skeletal muscle. Gene symbols followed with the symbol `+` indicate that multiple genes within that family are targeted. TFP = trifunctional protein.