Integrative multi-omics uncovers skeletal muscle enhancer programming of cardiorespiratory fitness

Anne M. Weitzel[1], Peter Orchard[2], Charles R. Evans[3], Nandini Manickam[2], Mary K. Treutelaar[3], Steven L. Britton[1,4], Lauren Koch[5], Jun Z. Li[6], Stephen C.J. Parker[2,7,8,9], Charles F. Burant[1,2,3,9,10]

Abstract

Cardiorespiratory fitness (CRF) is a heritable trait associated with improved metabolic health and reduced mortality. To identify regulatory mechanisms associated with CRF variation, we generated and integrated 546 transcriptomic and epigenomic (chromatin accessibility and histone marks) profiles from 128 genetically heterogeneous rats selectively bred for high and low running capacity, a model that mirrors CRF-associated traits in humans. We found that selection for high CRF led to genetic convergence in skeletal muscle regulatory regions linked to genes involved in lipid metabolism and angiogenesis. We validated thousands of these genetic effects through generation and integration of 426 genotype, gene expression, and chromatin accessibility profiles in an independent HCR×LCR F2 population (n=147). Together, these 972 omics profiles show that genetic variation reshapes the chromatin landscape to support energy metabolism and oxygen delivery in skeletal muscle and overall CRF, offering a molecular framework to identify targets that reduce cardiometabolic disease risk.

Abbreviations & Glossary of Intentionally Used Words

- AME = analysis of motif enrichment
- BCAA = branched-chain amino acid
- caQTL = chromatin accessibility quantitative trait locus
- CRF = cardiorespiratory fitness
- 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
- SRF = serum response factor
- TF = transcription factor
- TSS = transcription start site
- VO_2max = maximal oxygen uptake

- 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)

[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
[9] Co-Senior Author
[10] Lead Contact

Introduction

Physical activity is recognized for its salutary effect on human health and longevity[1,2], while 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]. The genetic determinants of CRF and exercise training appear to be largely distinct[14], which may explain why individuals with low CRF who exercise still face higher risks of cardiometabolic disease and premature death than their non-exercising, high CRF counterparts[12]. Because CRF is a valuable clinical metric[6,14] and behavioral interventions have limited impact on modifying it, we seek to understand how CRF-associated genetic variation is related to diverse physiological phenotypes, including longevity.

To study intrinsic CRF, we use a genetically heterogeneous rat model selectively bred for oxidative capacity, generating two divergent populations of High Capacity Runner (HCR) and Low Capacity Runner (LCR) lines[15]. This well-characterized model was developed from the N/NIH founder stock based on a run-to-exhaustion test as the metric of CRF with genetic diversity maintained through a rotational breeding paradigm[15]. This selection strategy led to persistent and heritable line-specific phenotypes that recapitulate human traits associated with divergent CRF, including sex-independent differences in physiologic measures of cardiometabolic health and longevity[16,17]. The within-line positive correlation between running capacity and longevity emphasizes the genetic underpinning of oxidative capacity and its impact on longevity[16]. HCR, compared to LCR, exhibit greater skeletal[18] and cardiac[19] muscle mass, increased skeletal muscle mitochondria[20–22] and capillary density[18,23–25], and higher type I and IIa fiber content[23,26]. These features are accompanied by enhanced energy utilization through preferential fatty acid oxidation[27] and delayed depletion of muscle glycogen during intense exercise[27] – also hallmarks of people with high oxidative capacity[27–31]. Transcriptome profiling of skeletal muscle further demonstrated an upregulation of genes involved in oxidative phosphorylation, fatty acid oxidation, and branched-chain amino acid (BCAA) metabolism, along with reduced expression of inflammatory pathways in HCR rats compared to LCR rats[28,32]. While large-scale human population studies have identified genetic variants associated with CRF[33,34], there remains a gap in knowledge about the underlying skeletal muscle molecular changes that drive oxidative capacity. The genetic variants that were selected to produce the HCR and LCR rat model provide a unique and tractable system to study the pathways shaped by selection that underlie variation in intrinsic CRF.

Approximately 90% of trait-associated common genetic variants are concentrated in noncoding regions of the genome[35] where they can modulate transcription factor (TF) binding and epigenomic activity[36] in cell-specific and trait-relevant contexts. We propose that genetic variants underlying HCR/LCR divergence act through epigenomic regulatory mechanisms that influence genes involved in oxidative capacity. To address this, we generate the first comprehensive skeletal muscle epigenomic map of CRF using 128 HCR and LCR rats by profiling chromatin accessibility via assay for transposase-accessible chromatin (ATAC-Seq)[37] and five gene expression-relevant histone modifications (H3K27ac, H3K4me1, H3K4me3, H3K36me3, H3K27me3) using Cleavage Under Targets and Tagmentation (CUT&Tag)[38]. We integrate these data with transcriptome profiles and genetic variation to identify regulatory mechanisms and candidate variants that link CRF-associated genetic divergence to downstream molecular pathways, including skeletal muscle regulatory elements and target genes. Our findings demonstrate that CRF-associated genetic variation concentrates in skeletal muscle epigenomic regions that influence TF and chromatin activity associated with genes involved in energy utilization, vascular remodeling, and skeletal muscle development. To validate the effects of genetic variants on epigenomic regulation and gene expression, we performed integrative analyses of ATAC-Seq (n = 132), RNA-Seq (n = 147), and genotype data collected from an HCR×LCR F2 population consisting of 491 rats[39]. Together, our results show that the epigenome serves as a molecular convergence point through which genetic selection shapes

skeletal muscle physiology related to CRF. This work nominates regulatory elements, genes, and variants as candidate molecular drivers of CRF, establishing a mechanistic framework for therapeutic strategies aimed at reducing cardiometabolic risk and promoting healthy aging.

Results

Divergent skeletal muscle epigenomic signatures associate with intrinsic cardiorespiratory fitness

To understand epigenomic and transcriptomic patterns associated with CRF, we profiled lateral gastrocnemius muscle from 128 HCR/LCR rats balanced for sex and acute exercise conditions. Briefly, half of the cohort underwent a six-week endurance training regimen with subsequent tissue collection either prior to, during, or after a near-exhaustion acute exercise bout (Figure S1A). To isolate selection driven CRF divergence between lines, all data were adjusted for sex, training, and acute exercise (Figure S4E-F). We generated a total of 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) (Figure 1A). The histone marks (H3K27me3, H3K27ac, H3K4me1, H3K4me3, and H3K36me3) alongside chromatin accessibility enable discovery of active, poised, and repressed chromatin states (Figure 1B). For example, we show the epigenomic signal from an individual sample surrounding *Gapdh*, a housekeeping gene expressed in skeletal muscle (Figure 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 (Figure 1D) and principal component analysis (PCA) of genome-wide fragment counts showed clear within-modality clustering (Figure 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 in this study.

Chromatin accessibility and histone modification profiles were uniformly integrated across each muscle sample using a single ChromHMM[40] model to derive a common set of 12 chromatin states based on the presence, combination, and signal abundance of each mark across the genome (Figure 2A), producing posterior probabilities that represent the likelihood of each chromatin state occurring at genome-wide loci. Figure 2B demonstrates the strength of this integration at a representative locus where increased ATAC-Seq and H3K4me1 signal in HCR coincides with a ChromHMM-derived Strong Enhancer state, while the same genomic locus is classified as a Quiescent/Low state in LCR. The HCR/LCR differences observed across individual peaks collectively support a differential Strong Enhancer state. Processing data for gene expression, chromatin accessibility, histone modification peak abundance, and chromatin state posteriors resulted in 14,781 *genes*, 101,638 chromatin accessible *peaks*, 45k-91k histone modification *peaks*, and 97k-124k chromatin state *regions*, respectively (Figure 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 *feature*s.

We first asked whether epigenomic differences accompany transcriptional differences between HCR and LCR rats, interpreting HCR/LCR differences as CRF-associated changes (HCR up = high CRF; LCR up = low CRF). Using modality-specific methods (Figure 2C) that normalized for technical and experimental variables (Figure S4B), we identified thousands of differences in gene expression, chromatin accessibility, histone abundance, and chromatin state probabilities between HCR and LCR (Figure 2C). As an example, Figure 2D shows the magnitude of differences in H3K4me1 peak abundance with 3,707 HCR up and 2,560 LCR up peaks (FDR < 0.1 & $|\log_2 FC| > 0.5$), suggesting selection-driven alterations in enhancer elements. These results demonstrate widespread CRF-associated epigenomic divergence across active, poised, and repressed chromatin states (Figure 2C, Figure S6A-F).

Cardiorespiratory fitness-divergent epigenomic regions uncover angiogenesis and skeletal muscle remodeling pathways

To establish a transcriptional reference point for the CRF phenotype, we performed pathway enrichment analysis on the HCR/LCR differential gene expression. Genes upregulated in HCRs were predominantly enriched for mitochondrial-related processes including aerobic respiration, oxidative phosphorylation, fatty acid oxidation, and BCAA degradation (FDRs < 4.26×10^{-9}) (Figure S7A-B), consistent with previous studies[28,32]. A distinct signal related to angiogenesis was also identified (23 gene sets, FDRs = 6.9×10^{-5} to 3.6×10^{-2}, highlighted in Figure S7A). These transcriptional profiles align with known CRF phenotypes of enhanced mitochondrial function[27] and capillary density[18,23–25] in HCR, reinforcing the functional interpretability of these data to evaluate epigenomic features.

We next asked whether CRF-divergent epigenomic features recapitulate the biological pathways identified by RNA-Seq or reflect distinct biology. We assigned each differential HCR/LCR epigenomic feature (nominal *p*-value < 0.01) to genes with the nearest TSS and performed pathway enrichment analysis. Across epigenomic modalities, the 'nearest TSS' genes were enriched for between 1 and 1,004 gene sets (FDR < 0.05) that demonstrated strong internal concordance (mean kappa coefficient = 0.23), but had low overlap with the RNA-Seq enrichment (mean kappa coefficient = -0.11) (Figure 2E). For example, differential Strong Enhancer nearest TSS genes predominantly enriched for pathways involved in skeletal muscle development, structural remodeling, and muscle cell differentiation, which were largely absent in RNA-Seq enrichment (Figure 2F). Among the few shared enrichments, actin binding and ameboidal-type cell migration reflect known CRF phenotypes of increased muscle size and capillarization in HCR skeletal muscle compared to LCR[18,23–25]. This suggests that the CRF-divergent epigenome may regulate lineage-specific developmental programs, such as muscle remodeling and angiogenesis, that are complementary to CRF but only modestly captured by RNA-Seq.

On average, the epigenomic regions driving enrichment reside 51.8 kb from the nearest TSS, indicating distal enhancer activity. Enhancers frequently regulate genes that are not their nearest neighbors[41–44], making linear proximity annotation prone to misassignment. This may in part explain the limited enrichment for mitochondrial metabolism observed in the epigenomic data, despite its strong enrichment in RNA-Seq (Figure S7A). To address this limitation, we developed a correlation-based framework to directly link regulatory elements to gene expression, independent of genomic distance.

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

Functional connections between epigenomic regulatory elements and gene transcription were directly inferred by applying a peak-to-gene (Pk2G) correlation framework. We mapped positive and negative regulatory associations between gene expression and both distal (TSS ±1 Mb) and promoter (TSS -2 kb, TSS +1 kb) from ATAC-Seq, H3K27ac, H3K4me1, and H3K4me3 modalities (Figure 3A). We identified 229,444 positive and 207,613 negative significant (FDR < 0.05) regulatory Pk2G links involving 88,440 peaks from open and active chromatin modalities and 10,786 genes. Consistent with our earlier findings, only 17.5% of peaks were most significantly linked to their nearest gene, whereas 50.1% were linked to genes 2-10 genes away and 32.5% to genes located more than 10 genes away, consistent with the biology that distal regulatory elements bypass nearby genes to modulate distal targets[41–43]. This map of gene regulation is illustrated at the *Decr1* locus, a mitochondrial gene involved in fatty acid beta-oxidation[45], where multiple regulatory elements show both positive and negative correlation with expression (Figure 3B). Additional examples in Figure S8 highlight the CRF-related regulatory architecture that connects chromatin activity to transcription. This Pk2G approach connects epigenomic peaks to biological function through gene expression, providing the framework to interpret how selection shaped the regulatory programs underlying the variation in CRF.

To identify biological processes potentially under primary control of the CRF-divergent epigenomic features, we performed pathway enrichment analysis on the Pk2G target genes linked to differential HCR/LCR peaks (nominal *p*-value < 0.01), ranked by the significance of correlation of the Pk2G link. Given that chromatin accessibility and histone modifications can both activate and repress transcription, we conducted four independent analyses: 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. Notably, genes positively linked to LCR up peaks enriched only for contractile fiber (boxed in Figure 3C), driven by the predominance of glycolytic type II fiber genes (*Actn3*, *Myl1*, *Casq1*), possibly reflecting the greater reliance on glucose as a fuel source during acute exercise in the LCR, resulting in lower endurance capacity[27]. In contrast, genes positively linked to HCR up peaks and negatively linked to LCR up peaks enriched for gene sets related to mitochondrial processes (aerobic respiration, fatty acid and BCAA metabolism) as well as endothelial angiogenesis (Figure 3C), mirroring RNA-Seq enrichment patterns (Figure S7A). This concordance between Pk2G and RNA-Seq derived enrichment supports the idea that selection has shaped distal epigenomic elements regulating energy metabolism and vascularization central to CRF-related phenotypes.

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, 124 HCR up peaks were positively linked to 130 target genes. Of these, 124 genes overlapped with the 463 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 (Figure 3D). Regulation of this fatty acid and BCAA metabolism gene subset may establish a metabolic framework that dictates mitochondrial efficiency through preferential fatty acid and BCAA utilization, core processes associated with CRF.

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

Having established that CRF-associated epigenomic features regulate nuclear-encoded mitochondrial genes and genes important for angiogenesis, we next asked whether distinct patterns of TF binding contribute to the divergent transcriptional and epigenomic profiles between HCR and LCR. We compared motif enrichment globally (Figure 4A) between significantly different HCR up regions and LCR up regions passing the conservative significance thresholds provided in Figure 2C, probing motifs for TFs that are expressed in the skeletal muscle based on presence in the RNA-Seq data. Applying this approach on epigenomic modalities associated with open chromatin, we identified 146 differential motifs, represented by 78 TFs (Figure 4B, Figure S9B). These results demonstrate widespread divergence in TF binding opportunity between HCR and LCR populations, independent of their expression profiles (Figure S10B).

We next characterized the epigenomic regions containing differential motifs through over-representation analysis of Pk2G target genes, stratified by modality, motif, and directionality between epigenomic marks and target gene expression. These motifs predominantly occur in differential ATAC-Seq and H3K4me1 regions, highlighted by the heatmap cluster with the most abundant signal in Figure 4C. HCR up ATAC-Seq and H3K4me1 peaks with greater E26 transformation-specific (ETS) family TF binding potential positively regulate genes involved in oxidative phosphorylation and mitochondrial structure (Figure 4D). Notably, HCR up H3K4me1 peaks’ positive gene targets are further enriched for fatty acid oxidation, BCAA oxidation, and vascular development gene sets (Figure 4D), refining the regulatory processes identified in Figure 3C. This suggests that the ETS TF function, in coordination with the H3K4me1 mark, plays an important role in orchestrating gene expression patterns that biologically optimize oxidative phosphorylation and oxygen delivery through enhanced capillary structure. These adaptations are directly linked to increased VO_2 max and related phenotypes, including lower body weight and enhanced cardiovascular function[6,46–51]. In contrast, LCR up peaks with differential binding potential for serum response factor (SRF) family TF motifs negatively regulate genes involved in these same mitochondrial

processes. These putative repressive and activating mechanisms suggest bidirectional and orthogonal selection for epigenomic regulation occurring in CRF-associated gene enhancers.

Selection of regulatory variants coincide with epigenomic divergence in HCR/LCR

Given the heritability of CRF, we asked whether epigenomic differences between HCR and LCR align with population-level genetic differentiation between the two lines, as expected under selection for running capacity. We imputed genotypes from single nucleotide polymorphisms (SNPs) called from RNA-Seq data, which yielded 637,713 variant sites across the genome. Differences in allele frequency between HCR and LCR were quantified by calculating per-variant fixation indices (F_{ST}), a standard metric for detecting loci under selection that ranges from 0 (no divergence) to 1 (complete fixation between HCR/LCR lines). To assess whether allele frequency divergence is more pronounced in differential epigenomic peaks, we performed linear regression between HCR/LCR epigenomic peak significance ($-\log_{10}(p\text{-value})$) and F_{ST} values of variants residing within each peak (Figure 5A). Allele frequency divergence increased with the strength of epigenomic differences, particularly for ATAC-Seq and H3K4me1 peaks (Figure 5B). These findings imply that selection of regulatory variants has contributed to chromatin remodeling at enhancers, influencing mitochondrial and angiogenic gene networks underlying CRF. Line-specific alleles within these enhancer regions may disrupt TF binding motifs, acting as regulatory pivot points that distinguish epigenomic and transcriptomic profiles based on high and low aerobic capacity.

Cardiorespiratory fitness associated allelic fixation modulates epigenomic regulation of energy metabolism via transcription factor binding

Building on our observation that fixed variants are enriched in enhancer regions associated with CRF, we next examined whether the presence of population-specific alleles directly alter TF binding affinity. We performed allele-specific differential motif enrichment analysis, testing the same epigenomic regions included in the global approach. To test the impact of fixed alleles ($F_{ST} > 0.3$) on motif recognition, we compared TF motif enrichment under two conditions for each modality: 1) HCR up regions with HCR alleles vs. HCR up regions with LCR alleles and 2) LCR up regions with LCR alleles vs. LCR up regions with HCR alleles (Figure 5C). Comparing enrichment significance across these trials, we identified 246 differentially enriched motifs that 121 TFs may bind (Figure 5D, Figure S9A). Of these, 92 motifs were unique to HCR up features, 76 to LCR up features, and 78 were shared. Importantly, 91.9% of the differentially enriched motifs do not exhibit significantly different expression of the encoded TF gene (Figure S10A), suggesting that selection acts primarily through altered TF binding potential rather than TF abundance.

Differentially enriched motifs predominantly belong to the TF families of Cys2-His2 zinc finger (zf-C2H2), ETS, Forkhead, and basic helix-loop-helix (bHLH), all with established roles in skeletal muscle identity[52–58] and mitochondrial function[59–65]. The functions of these TF families are consistent with our earlier pathway enrichment analyses in which skeletal muscle development was captured by peaks annotated to the nearest TSS (Figure 2F) and mitochondrial processes emerging from Pk2G links (Figure 3C). These TF families highlight how molecular adaptations that support enhanced aerobic capacity may be rooted in the presence of population alleles that enhance TF-mediated regulation of gene expression.

We identified 799 disrupted motifs linking allelic fixation to gene expression through epigenomic regulation. The impact of allele-specific binding on predicted TF binding affinity is illustrated by the Sp1 motif containing variant chr8:104691692 ($F_{ST} = 0.54$), where the HCR-enriched G allele (HCR allele frequency = 0.80) strengthens the motif match at position 10 compared to the LCR-enriched C allele (LCR allele frequency = 0.92, Figure 5G). Although *Sp1* is not differentially expressed between lines ($\log_2FC = -0.15$, FDR = 6.64×10^{-10}; Figure 5G), the chr8:104691692 G allele is observed alongside elevated ATAC-Seq, H3K27ac, H3K4me1, and H3K4me3 profiles (Figure 5F), all of which have positive Pk2G links with *Acad11* expression (Figure 5E). *Acad11* is a fatty acid dehydrogenase[66] and is upregulated in HCR ($\log_2FC = 1.04$, FDR = 3.59×10^{-66}). The chr8:104691692 G

allele likely drives this epigenomic and transcriptomic profile but would require validation in an independent and more genetically diverse population.

To generalize our biological understanding beyond the individual Sp1 motif, we assessed the functional relevance of all disrupted motifs through over-representation analysis of Pk2G target genes. Active and repressive mechanisms were explored by combining all modalities and conducting four independent analyses: 1) HCR up peaks with HCR preferred motifs positively linked to gene targets, 2) HCR up peaks with HCR preferred motifs negatively linked to gene targets, 3) LCR up peaks with LCR preferred motifs positively linked to gene targets, and 4) LCR up peaks with LCR preferred motifs negatively linked to gene targets. Target genes positively linked to HCR preferred motifs significantly enriched for gene sets involved in fatty acid metabolism and aerobic respiration (FDR < 0.1) (Figure 5H). These results demonstrate how allelic fixation can directly influence mitochondrial and lipid gene expression and metabolism (Figure 5I), establishing a molecular mechanism by which genetic divergence shapes the physiological traits underlying CRF.

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

Thus far, our results demonstrate many epigenomic differences between HCR and LCR skeletal muscle that correspond with gene expression patterns reflective of intrinsic CRF. In addition, we show a higher concentration of fixed alleles in CRF-divergent epigenomic regions and exemplify how selection may shape TF binding affinity. Assessing how genetic variants are associated with chromatin accessibility and gene expression in a separate orthogonal population would support that these variants are important in driving these differences. We performed a large-scale validation study using a HCR×LCR cross 'F2' population we created 11 generations prior to the current study, with ~13% less inbreeding than the generation 39 HCR/LCR, to conduct phenotype association studies[39] (Figure 6A).

To enable genetic association with gene expression and chromatin accessibility in the F2 population, we genotyped the F2 population and performed ATAC-Seq (n = 132) and RNA-Seq (n = 147) on gastrocnemius muscle samples. These additional data allowed us to evaluate genotypes at 1,730,238 imputed SNPs for association with the 1) expression of 14,047 genes through expression quantitative trait locus (eQTL) analysis and 2) accessibility of 120,351 ATAC-Seq peaks through chromatin accessibility quantitative trait locus (caQTL) analysis (Figure 6B). To account for linkage disequilibrium (LD), we grouped variants with $r^2 \geq 0.99$ to the most significant variant and refer to them as (e/ca)VariantClusters where appropriate. We identified 4,747 genes with a significant eQTL (eGenes, FDR < 0.05) regulated by 3,416 eVariantClusters and 10,099 peaks with a significant caQTL (caPeaks, FDR < 0.05) regulated by 5,115 caVariantClusters. Because the HCR×LCR recombination events produced an F2 population with a broad range in running capacity[39], these associations provide a framework to explore how genetic variation may contribute to transcriptional and epigenomic regulation relevant to intrinsic CRF. To further refine variant-peak interactions, we focused on caPeaks where the associated variant was within the peak or in strong LD ($r^2 > 0.8$) with the most significant caVariant. This resulted in a set of 2,535 caPeaks linked to 1,582 caVariantClusters. This map of allelic effects on the skeletal muscle epigenome and transcriptome in the F2 population serves as a robust validation framework for mechanisms identified in the HCR/LCR model.

Next, we assessed whether the variants associated with gene expression or chromatin accessibility in the F2 population also correlate, with the same direction of effect, with the gene expression and chromatin activity differences observed between the HCR/LCR lines. For example, if one allele of a genetic variant predominates in the HCR line and the other allele predominates in the LCR line, and the HCR-favored allele positively associates with ATAC-Seq peak signal in the F2 population (relative to the LCR-favored allele), then one would expect the ATAC-Seq peak to have higher signal in the HCR line relative to the LCR line (Figure 6C). When comparing the allelic effects on chromatin accessibility observed in the F2 population, for caVariants with F_{ST} >

0.3 between the HCR/LCR lines, with the chromatin accessibility and histone modification differences (H3K27ac, H3K4me1, H3K4me3) observed between the HCR/LCR lines, we found high concordance (94%) of direction of effect, strongly supporting the hypothesis that these variants exert similar regulatory effects across populations (binomial *p*-value = 2.17×10^{-63}, expected = 50%, alternative = "greater"; Figure 6D).

Similarly, we assessed if F2 allelic effects on gene expression parallel HCR/LCR changes by comparing the signed $-\log_{10}$(*p*-value) of eGenes where a variant with $F_{ST} > 0.3$ was in strong LD with the top eVariant ($r^2 > 0.9$) to the signed $-\log_{10}$(*p*-value) of HCR/LCR gene expression changes. This resulted in 85% concordance (binomial *p*-value = 1.63×10^{-89}, expected = 50%, alternative = "greater"; Figure 6E), indicating that variant-driven gene expression differences in F2 largely translate to the HCR/LCR population. Finally, we tested whether eQTL effects align with Pk2G links, which capture peak-gene expression relationships in the HCR/LCR data (Figure 6F). If a regulatory peak containing an eVariant functionally links to gene expression, the effect allele should show consistent impact across datasets. We observed 69% concordance (binomial *p*-value = 3.15×10^{-17}, expected = 50%, alternative = "greater"), showing a significant overlap in allele-driven epigenomic modifications influencing transcription between the F2 and HCR/LCR model.

Considering the co-occurrence of variant-gene and variant-chromatin links allows us to pinpoint variants that may influence TF binding affinity, which can provide insights into the regulatory mechanisms behind physiological traits associated with innate running capacity. As an example, we show the validation of variant chr8:104691692 within the Sp1 motif (Figure 5E-G). The ATAC-Seq peak is positively correlated with *Acad11* expression in HCR and LCR (Figure 6G) and F2 (Figure 6J). In the HCR/LCR model, the presence of the C allele is positively associated with the intronic-residing ATAC-Seq peak (Figure 6H) and *Acad11* expression (Figure 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}; Figure 6K) and expression of *Acad11* (eQTL $\beta = 1.11$, *p*-value = 4×10^{-37}; Figure 6L). This suggests that an increase in *Sp1* binding due to the A-allele present at variant chr8:104691692 in animals with increased running capacity leads to greater chromatin accessibility, facilitating the binding of additional transcriptional machinery, which leads to an increased expression of *Acad11*. This is one of the 293 examples validated in our study that likely contributes toward shaping the regulatory landscape in relation to the CRF phenotype.

Ultimately, the concordance between the genetic effects on chromatin accessibility/gene expression in the F2 population and the epigenomic/gene expression differences observed between the HCRs and LCRs suggests that genetic differences between the HCRs and LCRs drive many of the observed epigenomic and gene expression differences between the lines. It also lends independent support to the reliability of the differential HCR/LCR gene expression and chromatin activity analyses. Because a large fraction of genetic variants associated with complex traits likely exert their effects via gene regulation[35,67], the (e/ca)Variants we have identified in the F2 population should be of particular interest when searching for functional, CRF-associated variants in the HCR/LCR lines and their effects on chromatin accessibility and gene expression.

Discussion

Intrinsic CRF is a complex trait that serves as a determinant of cardiometabolic health and a reproducible predictor of longevity. Beyond previous genetic association[33,34] and skeletal muscle transcriptome studies[32,68,69], the molecular mechanisms underlying CRF are poorly understood. Here, we used the well-characterized HCR/LCR rat model to generate the first comprehensive skeletal muscle epigenomic map of CRF. By integrating multi-omic data and incorporating a validation study, we demonstrate that CRF-divergent genetic variants converge on skeletal muscle epigenomic loci that, in turn, shape transcriptomic signatures. These molecular signatures highlight remodeled skeletal muscle optimized for oxygen and nutrient delivery and expanded mitochondrial capacity of ATP generation to support contraction.

Focusing on understanding the underlying biology of CRF, we linked genetic variants to epigenomic features, associated epigenomic features to genes, nominated transcription factor binding events through motif enrichment, and explored how selected variants may influence complex regulatory networks (Figure S11). Our analyses discovered several notable pathways associated with CRF, including enhanced mitochondrial metabolism, fatty acid oxidation, and vascular remodeling, reflecting the phenotypes of humans with higher CRF[70–75]. To validate these discoveries, we performed a large-scale independent and orthogonal F2 population study with multi-omics profiling, and found concordant, allele-driven regulatory effects. In addition to providing a mechanistic understanding of how genetic variation shapes the skeletal muscle CRF-associated regulatory landscapes, this framework more broadly demonstrates how similar strategies can be applied to the study of other complex traits. Using this approach, we show that ATAC-Seq and H3K4me1 marked enhancer regions associated with high CRF are enriched for ETS-family TF motifs, which we propose increase the expression of genes that encode for fatty acid oxidation and mitochondrial inner membrane proteins. While ETS TF regulation of mitochondrial gene expression in muscle has only been reported for a single ATP synthase beta-subunit[59], our results demonstrate broader involvement in mitochondrial programs. Notably, ETS TF expression differences are largely insignificant between HCR and LCR, indicating that genetically encoded differences in distal enhancer sites, rather than TF abundance, underlie CRF-related transcriptional differences.

Selective breeding of HCR and LCR rats gradually amplified phenotypes directly relevant to human CRF[17], with the greatest impact on skeletal muscle as observed by fewer mitochondrial differences in heart[76], liver[68], and adipose tissue[77] between the lines. The selection reshaped the skeletal muscle epigenome, specifically through ETS TF-driven enhancer networks regulating mitochondrial metabolism and angiogenesis. We propose that ETS TF-enriched regulatory elements enhance mitochondrial function in a coordinated fashion by increasing expression of trifunctional protein subunits, acyl-CoA dehydrogenases, fumarate hydratase, and genes encoding subunits of ETC complexes I-V in HCR. These epigenomic and transcriptomic patterns support streamlined electron entry and flow through Complex II, in which acyl-CoA dehydrogenases may indirectly increase electron flux, succinate dehydrogenase converts succinate into fumarate, which is subsequently metabolized to malate by fumarate hydratase ($\log_2$FC = 0.68, FDR = 1.47×10^{-257}, rank = 3). Together, these complementary regulatory mechanisms prime high CRF muscle for efficient oxidative metabolism (Figure 7).

Reduced mitochondrial fatty acid metabolism is consistent with mitochondrial ‘dysfunction’ observed in individuals predisposed to type 2 diabetes[78–80], where overload leads to the accumulation of diacylglycerols, ceramides, and other lipid species that promote insulin resistance and oxidative stress[81]. Additionally, short chain acyl-CoAs accumulation can influence protein modifications and impair mitochondrial function[82]. 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[83,84], suggesting a deeply conserved central role in biology. The fine-tuning of gene expression by CRF-divergent variants in non-coding regions and preservation of coding sequences suggest that fundamental enzymatic functions of these genes have been conserved, while regulatory adaptation to selection has optimized mitochondrial metabolism. Notably, specific acyl-CoA dehydrogenase isoforms also metabolize peroxidized fatty acid metabolism, contributing to the removal of toxic lipid species[85,86]. 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[87], obesity, and insulin resistance[88,89].

Selection for mitochondrial metabolism in HCR/LCR is accompanied by epigenomic regulation of angiogenic genes, also under control of ETS TFs. The parallel ETS control of angiogenesis supports both oxygen delivery and energy production under increased demand. While enhanced angiogenesis-related signal in HCR likely reflects epigenomically primed endothelial cells responding to myofibers-derived cues (*Vegfa* and *Vegfb*), several angiogenesis related genes (*Slc1a1*, *Parp14*, *Mmrn2*, and *Atp8b1*) are also linked to regulatory variants validated in the F2 cohort, suggesting allele-specific effects that prime the epigenome to regulate gene

expression in a dose-response manner (Figure 6C). This metabolic-vascular coupling likely reflects both a developmental program and an adaptive response to increased skeletal muscle energy demand, with shared transcriptional regulatory control at its core.

The HCR/LCR model provides a simple and physiologically relevant system to study CRF. We identified coordinated epigenomic tuning of skeletal muscle fatty acid metabolism, BCAA catabolism, and vascular remodeling, which are directly relevant to human cardiometabolic health. The regulatory signatures of noncoding variation that shape the chromatin landscape, TF binding, and metabolic gene expression establish a mechanistic framework for enhancing CRF through targeted intervention. As low intrinsic CRF is a strong predictor of cardiometabolic disease and reduced mortality regardless of exercise training, molecular strategies to manipulate CRF hold significant promise for improving health span, particularly in genetically susceptible individuals. Beyond CRF, our approach illustrates a generalizable strategy for dissecting the molecular architecture of complex traits through integrated multi-omic mapping and cross-population validation.

Resource Availability

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Charles Burant (burantc@umich.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- 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.17245661).
- 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.

Methods

HCR/LCR Study Design

A total of 128 rats from generation 39 of the HCR/LCR line were included in this study, including equal numbers of females and males (Figure S1A). 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 on a 15% incline, configured as previously described[27]. 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[27]. 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), 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 actual 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 (Figure 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[90]. 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[38] following the “CUT&Tag-direct with CUTAC V.3” protocol[91]. To reduce batch effect across experimental contrasts, samples were randomly assigned to five batches to carry out 360 CUT&Tag reactions at the bench (Figure 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 (Figure 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[92], GenomicRanges[93], plyranges[94], and ComplexHeatmap[95,96]. Snakemake[97] was used to create pipelines used throughout the analysis. Sequencing data was aligned to the ensembl release (v109) of the rn7 genome[98].

Processing of RNA-Seq Data

Fastq files were trimmed using the TrimmomaticPE function (Trimmomatic[99] v0.36) with the following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10:2:keepBothReads SLIDINGWINDOW:5:20. Reads were aligned using STAR[100] (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[101] *mem* version 0.7.15 parameters "-I 200,200,5000" and then sorted using samtools[102] 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[102]. Duplicates were marked using picard[103] *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[102] *view* with parameters "-h -f 3 -F 4 -F 8 -F 256 -F 1024 -F 2048 -q 30". Bedtools[104] 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[105] 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[106] 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 (Figure 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[94] *reduce_ranges* was used create the master peak set for each modality and chromVar[107] *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[108]. 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[104] version 2.30.0 *makewindows* and the number of reads present in each bin were counted. To detect erroneous signal, quantile normalization (broman::normalize[109]) 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[110] 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 (Figure S4A) and observed variability in PCA when added to the starting model (Figure S4C-D), while ensuring that the latent variable(s) show a relationship with the ataqv[106] QC attributes and minimal relationship with the binarized experimental variables (Figure 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[111] 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 Figure 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[40] 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[93] 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 (Figure 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 Figure 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[112]. Missing Ensembl or Entrez ID matches were filled in where possible using the *biomaRt*[113] package to retrieve gene information from the *rnorvegicus_gene_ensembl* v109[98] 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[114]. Unless stated otherwise, categories returning an FDR < 0.05 were considered significant.

RNA-Enrich[115] and ChIP-Enrich[116] were used for pathway enrichment analysis on differential transcriptomic and epigenomic features, respectively. Pathway enrichment analysis was performed independently on all data modalities. Pathway enrichment analysis of differential gene expression data included all genes, ranked by *p*-value with direction of change provided. Custom ChIP-Enrich locus definitions were created for the rn7 genome using code published on GitHub (https://github.com/weitzela/custom-chipenrich-locus-defs, commit e6ebc23). Epigenomic pathway enrichment analysis included peaks that passed a more inclusive significance threshold to allow for an adequate number of peaks to be included in the analysis. Sequencing-based epigenomic peaks with a DESeq *p*-value < 0.05 and chromatin state regions with a linear regression *p*-value < 0.01 were included as the input for ChIP-Enrich, using the *broadenrich* test. The number of input features ranged from 5k-60k sequencing-based epigenomic peaks and 26k-63k chromatin state regions. Input features were annotated to genes with the nearest transcription start site. The number of genes annotated by each epigenomic dataset following the Broad-Enrich annotation protocol is as follows: 8,320 H3K27me3 peaks annotated to 5,591 genes, 6,747 H3K27ac peaks annotated to 4,763 genes, 27,769 H3K4me1 peaks annotated to 10,037 genes, 12,978 H3K4me3 peaks annotated to 7,526 genes, 4,960 H3K36me3 peaks annotated to 3,706 genes, 59,528 ATAC-Seq peaks annotated to 15,682 genes, 27,123 EnhS (E1) regions annotated to 8,064 genes, 45,612 EnhW (E3) regions annotated to 7,355 genes, 46,603 EnhS (E4) regions annotated to 8,488 genes, 41,584 PromWk (E5) regions annotated to 8,828 genes, 39,132 TxTrans (E6) regions annotated to 8,200 genes, 26,159 PromA (E7) regions annotated to 8,181 genes, 30,797 PromPois (E8) regions annotated to 8,042 genes, 32,917 ReprPC (E10) regions annotated to 6,690 genes, 38,313 TxElon (E12) regions annotated to 7,208 genes, and 63,960 QuiesSum (EX) regions annotated to 9,231 genes.

To assess the processes that Pk2G Link target genes were involved in, *Line* differential peaks (*p*-value < 0.05) were queried. The list of input genes was separated by direction of peak and direction of relationship between the peak and target genes. These lenient criteria were chosen to be consistent with the ChIP-Enrich analysis conducted previously. The number of input genes for each trial is as follows: HCR up peaks positively linked to 4,762 genes, HCR up peaks negatively linked to 4,849 genes, LCR up peaks positively linked to 5,039 genes, and LCR up peaks negatively linked to 4,595 genes. The fgsea function from the fgsea[117] 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[118] 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[119] 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.[120], detailed further here.

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[121] version 4.3.0.0. Recalibration detection referred to the known variants found in the 8 HCR/LCR founder strains[122]. SNPs were called using bcftools[123,124] 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[122].

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.[32]. 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)[103] 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[121] 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[124] *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[124] *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.[125] 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[126] version 1.2.0 and SHAPEIT5[127] version 5.1.1 were used to carry out genotype imputation. The reference panel published by Hermsen et al.[128], was prephased using IMPUTE5 *xcftools_static view* (-m 0.03125). Target files were prephased using SHAPEIT5[127] *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 (Figure 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})[129] 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[129]. 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[130] 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[131] = 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 Figure 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)[132] 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)[133] 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 (Methods 2.2.14.3), retaining those with Δ FIMO $-\log_{10}$(*p*-value) > 1.

Sample Swap Detection for F2 Analysis

QTLtools[134] (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[111]-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)[135]. 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[136] (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[114] 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).

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. Library prep and next-generation sequencing was carried out in the Advanced Genomics Core at the University of Michigan. Funding for the collection of HCR/LCR skeletal muscle samples was provided by a pilot award to C.R.E. from the Michigan Nutrition Obesity Research Center, supported by NIH grant P30 DK089503. Additional funding sources that supported this research: A. Alfred Taubman Medical Research Institute NIH grant U24 DK112342, R01 DK117960, and Systems and Integrative Biology Training Grant T32 GM008322. Johanna Y Fleischman, Martin J Walsh, and Alexander J Weitzel for reviewing the manuscript and providing advice as it was being prepared.

Author Contributions

Conceptualization: A.M.W., C.R.E., J.Z.L., S.C.J.P, and C.F.B.; Methodology: A.M.W., P.O., C.R.E, S.L.B., L.K., S.C.J.P., and C.F.B.; Validation: A.M.W., P.O., S.C.J.P., and C.F.B.; Formal analysis: A.M.W. and P.O.; Investigation: A.M.W., C.R.E., N.M., and M.K.T.; Resources: J.Z.L. and C.F.B.; Data curation: A.M.W. and P.O.; Writing – original draft: A.M.W., S.C.J.P., and C.F.B.; Writing – review & editing: all authors; Visualization: A.M.W.; Supervision: P.O., J.Z.L., S.C.J.P., and C.F.B.; Project Administration: M.K.T., C.R.E., S.C.J.P, and C.F.B.; Funding Acquisition: J.Z.L., S.C.J.P, and C.F.B.

Declaration of Interests

The authors declare no competing interests.

References

1 Lee I-M, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT, *et al.* Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. *Lancet Lond Engl* 2012;**380**:219–29. https://doi.org/10.1016/S0140-6736(12)61031-9.

2 Physical Activity Guidelines for Americans 2023.

3 Bouchard C, Blair SN, Katzmarzyk PT. Less Sitting, More Physical Activity, or Higher Fitness? *Mayo Clin Proc* 2015;**90**:1533–40. https://doi.org/10.1016/j.mayocp.2015.08.005.

4 Lang JJ, Prince SA, Merucci K, Cadenas-Sanchez C, Chaput J-P, Fraser BJ, *et al.* 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* 2024;**58**:556–66. https://doi.org/10.1136/bjsports-2023-107849.

5 Blair SN, Kohl HW, Paffenbarger RS, Clark DG, Cooper KH, Gibbons LW. Physical fitness and all-cause mortality. A prospective study of healthy men and women. *JAMA* 1989;**262**:2395–401. https://doi.org/10.1001/jama.262.17.2395.

6 Ross Robert, Blair Steven N., Arena Ross, Church Timothy S., Després Jean-Pierre, Franklin Barry A., *et al.* 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* 2016;**134**:e653–99. https://doi.org/10.1161/CIR.0000000000000461.

7 Franklin BA, Wedig IJ, Sallis RE, Lavie CJ, Elmer SJ. Physical Activity and Cardiorespiratory Fitness as Modulators of Health Outcomes. *Mayo Clin Proc* 2023;**98**:316–31. https://doi.org/10.1016/j.mayocp.2022.09.011.

8 Schutte NM, Nederend I, Hudziak JJ, Bartels M, De Geus EJC. Twin-sibling study and meta-analysis on the heritability of maximal oxygen consumption. *Physiol Genomics* 2016;**48**:210–9. https://doi.org/10.1152/physiolgenomics.00117.2015.

9 Sarzynski MA, Rice TK, Després J-P, Pérusse L, Tremblay A, Stanforth PR, *et al.* The HERITAGE Family Study: A Review of the Effects of Exercise Training on Cardiometabolic Health, with Insights into Molecular Transducers. *Med Sci Sports Exerc* 2022;**54**:S1–43. https://doi.org/10.1249/MSS.0000000000002859.

10 Bouchard C, An P, Rice T, Skinner JS, Wilmore JH, Gagnon J, *et al.* Familial aggregation of VO(2max) response to exercise training: results from the HERITAGE Family Study. *J Appl Physiol Bethesda Md 1985* 1999;**87**:1003–8. https://doi.org/10.1152/jappl.1999.87.3.1003.

11 Blair SN, Cheng Y, Holder JS. Is physical activity or physical fitness more important in defining health benefits? *Med Sci Sports Exerc* 2001;**33**:S379-399; discussion S419-420. https://doi.org/10.1097/00005768-200106001-00007.

12 Celis-Morales CA, Lyall DM, Anderson J, Iliodromiti S, Fan Y, Ntuk UE, *et al.* 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* 2016:ehw249. https://doi.org/10.1093/eurheartj/ehw249.

13 Ross R, Goodpaster BH, Koch LG, Sarzynski MA, Kohrt WM, Johannsen NM, *et al.* Precision exercise medicine: understanding exercise response variability. *Br J Sports Med* 2019;**53**:1141–53. https://doi.org/10.1136/bjsports-2018-100328.

14 Hanscombe KB, Persyn E, Traylor M, Glanville KP, Hamer M, Coleman JRI, *et al.* The genetic case for cardiorespiratory fitness as a clinical vital sign and the routine prescription of physical activity in healthcare. *Genome Med* 2021;**13**:180. https://doi.org/10.1186/s13073-021-00994-9.

15 Koch LG, Britton SL. Artificial selection for intrinsic aerobic endurance running capacity in rats. *Physiol Genomics* 2001;**5**:45–52. https://doi.org/10.1152/physiolgenomics.2001.5.1.45.

16 Koch LG, Kemi OJ, Qi N, Leng SX, Bijma P, Gilligan LJ, *et al.* Intrinsic aerobic capacity sets a divide for aging and longevity. *Circ Res* 2011;**109**:1162–72. https://doi.org/10.1161/CIRCRESAHA.111.253807.

17 Wisløff U, Najjar SM, Ellingsen O, Haram PM, Swoap S, Al-Share Q, *et al.* Cardiovascular risk factors emerge after artificial selection for low aerobic capacity. *Science* 2005;**307**:418–20. https://doi.org/10.1126/science.1108177.

18 Howlett RA, Kirkton SD, Gonzalez NC, Wagner HE, Britton SL, Koch LG, *et al.* Peripheral oxygen transport and utilization in rats following continued selective breeding for endurance running capacity. *J Appl Physiol* 2009;**106**:1819–25. https://doi.org/10.1152/japplphysiol.00914.2007.
19 Gonzalez NC, Kirkton SD, Howlett RA, Britton SL, Koch LG, Wagner HE, *et al.* Continued divergence in $\dot{V}O_{2\,max}$ of rats artificially selected for running endurance is mediated by greater convective blood O_2 delivery. *J Appl Physiol* 2006;**101**:1288–96. https://doi.org/10.1152/japplphysiol.01527.2005.
20 Souza RWA, Alves CRR, Medeiros A, Rolim N, Silva GJJ, Moreira JBN, *et al.* Differential regulation of cysteine oxidative post-translational modifications in high and low aerobic capacity. *Sci Rep* 2018;**8**:17772. https://doi.org/10.1038/s41598-018-35728-2.
21 Seifert EL, Bastianelli M, Aguer C, Moffat C, Estey C, Koch LG, *et al.* Intrinsic aerobic capacity correlates with greater inherent mitochondrial oxidative and H_2O_2 emission capacities without major shifts in myosin heavy chain isoform. *J Appl Physiol* 2012;**113**:1624–34. https://doi.org/10.1152/japplphysiol.01475.2011.
22 Rivas DA, Lessard SJ, Saito M, Friedhuber AM, Koch LG, Britton SL, *et al.* 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* 2011;**300**:R835–43. https://doi.org/10.1152/ajpregu.00659.2010.
23 Kivelä R, Silvennoinen M, Lehti M, Rinnankoski-Tuikka, R, Purhonen T, Ketola T, *et al.* Gene expression centroids that link with low intrinsic aerobic exercise capacity and complex disease risk. *FASEB J* 2010;**24**:4565–74. https://doi.org/10.1096/fj.10-157313.
24 Gonzalez NC, Howlett RA, Henderson KK, Koch LG, Britton SL, Wagner HE, *et al.* Systemic oxygen transport in rats artificially selected for running endurance. *Respir Physiol Neurobiol* 2006;**151**:141–50. https://doi.org/10.1016/j.resp.2005.09.012.
25 Howlett RA, Gonzalez NC, Wagner HE, Fu Z, Britton SL, Koch LG, *et al.* Selected Contribution: Skeletal muscle capillarity and enzyme activity in rats selectively bred for running endurance. *J Appl Physiol* 2003;**94**:1682–8. https://doi.org/10.1152/japplphysiol.00556.2002.
26 Rogers RS, Morris EM, Wheatley JL, Archer AE, McCoin CS, White KS, *et al.* Deficiency in the Heat Stress Response Could Underlie Susceptibility to Metabolic Disease. *Diabetes* 2016;**65**:3341–51. https://doi.org/10.2337/db16-0292.
27 Overmyer KA, Evans CR, Qi NR, Minogue CE, Carson JJ, Chermside-Scabbo CJ, *et al.* Maximal Oxidative Capacity during Exercise Is Associated with Skeletal Muscle Fuel Selection and Dynamic Changes in Mitochondrial Protein Acetylation. *Cell Metab* 2015;**21**:468–78. https://doi.org/10.1016/j.cmet.2015.02.007.
28 Fleischman JY, Van Den Bergh F, Collins NL, Bowers M, Beard DA, Burant CF. Higher mitochondrial oxidative capacity is the primary molecular differentiator in muscle of rats with high and low intrinsic cardiorespiratory fitness. *Mol Metab* 2023;**76**:101793. https://doi.org/10.1016/j.molmet.2023.101793.
29 Aon MA, Cortassa S, Juhaszova M, González-Reyes JA, Calvo-Rubio M, Villalba JM, *et al.* Mitochondrial health is enhanced in rats with higher vs. lower intrinsic exercise capacity and extended lifespan. *NPJ Aging Mech Dis* 2021;**7**:1. https://doi.org/10.1038/s41514-020-00054-3.
30 Heyne E, Zeeb S, Junker C, Petzinna A, Schrepper A, Doenst T, *et al.* Exercise Training Differentially Affects Skeletal Muscle Mitochondria in Rats with Inherited High or Low Exercise Capacity. *Cells* 2024;**13**:393. https://doi.org/10.3390/cells13050393.
31 Pinto AJ, Bergouignan A, Dempsey PC, Roschel H, Owen N, Gualano B, *et al.* Physiology of sedentary behavior. *Physiol Rev* 2023;**103**:2561–622. https://doi.org/10.1152/physrev.00022.2022.
32 Ren Y, Koch LG, Britton SL, Qi NR, Treutelaar MK, Burant CF, *et al.* Selection-, age-, and exercise-dependence of skeletal muscle gene expression patterns in a rat model of metabolic fitness. *Physiol Genomics* 2016;**48**:816–25. https://doi.org/10.1152/physiolgenomics.00118.2015.
33 Klevjer M, Nordeidet AN, Hansen AF, Madssen E, Wisløff U, Brumpton BM, *et al.* Genome-Wide Association Study Identifies New Genetic Determinants of Cardiorespiratory Fitness: The Trøndelag Health Study. *Med Sci Sports Exerc* 2022;**54**:1534–45. https://doi.org/10.1249/MSS.0000000000002951.
34 Bye A, Klevjer M, Ryeng E, Silva GJJD, Moreira JBN, Stensvold D, *et al.* Identification of novel genetic variants associated with cardiorespiratory fitness. *Prog Cardiovasc Dis* 2020;**63**:341–9. https://doi.org/10.1016/j.pcad.2020.02.001.
35 Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, *et al.* Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* 2012;**337**:1190–5. https://doi.org/10.1126/science.1222794.

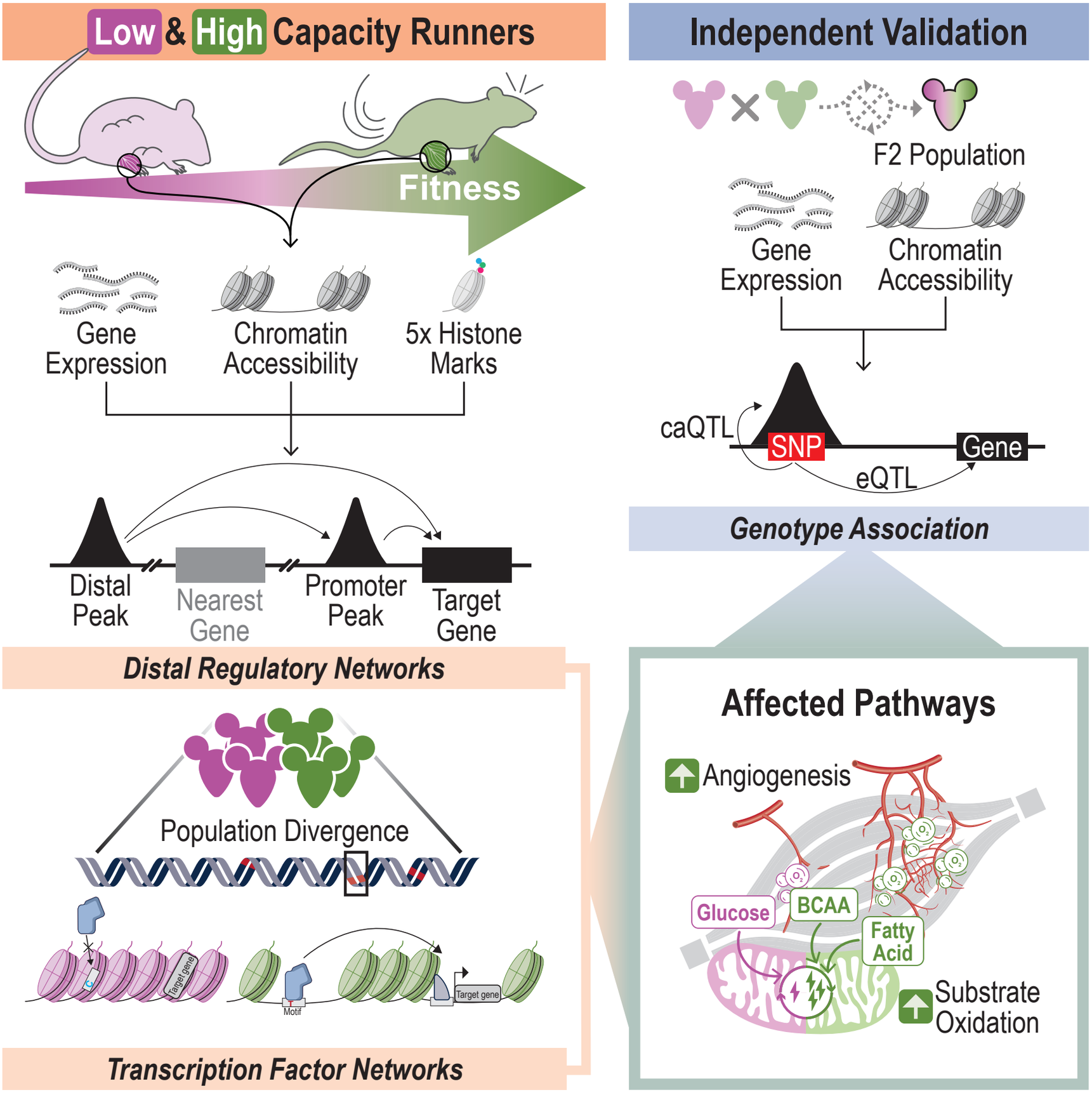
36 Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, *et al.* Extensive Variation in Chromatin States Across Humans. *Science* 2013;**342**:750–2. https://doi.org/10.1126/science.1242510.
37 Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* 2015;**109**:. https://doi.org/10.1002/0471142727.mb2129s109.
38 Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* 2019;**10**:1930. https://doi.org/10.1038/s41467-019-09982-5.
39 Ren Y, Overmyer KA, Qi NR, Treutelaar MK, Heckenkamp L, Kalahar M, *et al.* Genetic Analysis of a Rat Model of Aerobic Capacity and Metabolic Fitness. *PLoS ONE* 2013;**8**:e77588. https://doi.org/10.1371/journal.pone.0077588.
40 Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 2012;**9**:215–6. https://doi.org/10.1038/nmeth.1906.
41 Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. *Nature* 2012;**489**:109–13. https://doi.org/10.1038/nature11279.
42 Fulco CP, Munschauer M, Anyoha R, Munson G, Grossman SR, Perez EM, *et al.* Systematic mapping of functional enhancer-promoter connections with CRISPR interference. *Science* 2016;**354**:769–73. https://doi.org/10.1126/science.aag2445.
43 Nasser J, Bergman DT, Fulco CP, Guckelberger P, Doughty BR, Patwardhan TA, *et al.* Genome-wide enhancer maps link risk variants to disease genes. *Nature* 2021;**593**:238–43. https://doi.org/10.1038/s41586-021-03446-x.
44 Eder M, Moene CJI, Dauban L, Magnitov M, Drayton J, de Haas M, *et al.* Functional maps of a genomic locus reveal confinement of an enhancer by its target gene. *Science* 2025;**389**:eads6552. https://doi.org/10.1126/science.ads6552.
45 Miinalainen IJ, Schmitz W, Huotari A, Autio KJ, Soininen R, Ver Loren van Themaat E, *et al.* Mitochondrial 2,4-dienoyl-CoA reductase deficiency in mice results in severe hypoglycemia with stress intolerance and unimpaired ketogenesis. *PLoS Genet* 2009;**5**:e1000543. https://doi.org/10.1371/journal.pgen.1000543.
46 Hsieh P-L, Chen M-L, Huang C-M, Chen W-C, Li C-H, Chang L-C. Physical activity, body mass index, and cardiorespiratory fitness among school children in Taiwan: a cross-sectional study. *Int J Environ Res Public Health* 2014;**11**:7275–85. https://doi.org/10.3390/ijerph110707275.
47 Sterkowicz S, Lech G, Pałka T, Tyka A, Sterkowicz-Przybycień KL, Szyguła Z, *et al.* BODY BUILD AND BODY COMPOSITION VS. PHYSICAL CAPACITY IN YOUNG JUDO CONTESTANTS COMPARED TO UNTRAINED SUBJECTS. *Biol Sport* 2011;**28**:271–7. https://doi.org/10.5604/965486.
48 Setty P, Padmanabha B, Doddamani B. Correlation between obesity and cardio respiratory fitness. *Int J Med Sci Public Health* 2013;**2**:300. https://doi.org/10.5455/ijmsph.2013.2.298-302.
49 Cc L, Ib U. Effect of body mass index on cardiorespiratory fitness in young healthy males. *Int J Sci Res Publ* 2014;**4**:.
50 Mondal H, Mishra SP. Effect of BMI, Body Fat Percentage and Fat Free Mass on Maximal Oxygen Consumption in Healthy Young Adults. *J Clin Diagn Res JCDR* 2017;**11**:CC17–20. https://doi.org/10.7860/JCDR/2017/25465.10039.
51 Kang S-J, Ko K-J. Association between resting heart rate, VO2max and carotid intima-media thickness in middle-aged men. *IJC Heart Vasc* 2019;**23**:100347. https://doi.org/10.1016/j.ijcha.2019.100347.
52 Ren R, Fan Y, Peng Z, Wang S, Jiang Y, Fu L, *et al.* Characterization and perturbation of CTCF-mediated chromatin interactions for enhancing myogenic transdifferentiation. *Cell Rep* 2022;**40**:111206. https://doi.org/10.1016/j.celrep.2022.111206.
53 Liu Y, Wan X, Li H, Chen Y, Hu X, Chen H, *et al.* CTCF coordinates cell fate specification via orchestrating regulatory hubs with pioneer transcription factors. *Cell Rep* 2023;**42**:113259. https://doi.org/10.1016/j.celrep.2023.113259.
54 Zhang W, Tong H, Zhang Z, Shao S, Liu D, Li S, *et al.* Transcription factor EGR1 promotes differentiation of bovine skeletal muscle satellite cells by regulating *MyoG* gene expression. *J Cell Physiol* 2018;**233**:350–62. https://doi.org/10.1002/jcp.25883.
55 Xu M, Chen X, Chen D, Yu B, Huang Z. FoxO1: a novel insight into its molecular mechanisms in the regulation of skeletal muscle differentiation and fiber type specification. *Oncotarget* 2017;**8**:10662–74. https://doi.org/10.18632/oncotarget.12891.

56 Sanchez AMJ, Candau RB, Bernardi H. FoxO transcription factors: their roles in the maintenance of skeletal muscle homeostasis. *Cell Mol Life Sci* 2014;**71**:1657–71. https://doi.org/10.1007/s00018-013-1513-z.
57 Zammit PS. Function of the myogenic regulatory factors Myf5, MyoD, Myogenin and MRF4 in skeletal muscle, satellite cells and regenerative myogenesis. *Semin Cell Dev Biol* 2017;**72**:19–32. https://doi.org/10.1016/j.semcdb.2017.11.011.
58 Ustanina S, Carvajal J, Rigby P, Braun T. The Myogenic Factor Myf5 Supports Efficient Skeletal Muscle Regeneration by Enabling Transient Myoblast Amplification. *Stem Cells* 2007;**25**:2006–16. https://doi.org/10.1634/stemcells.2006-0736.
59 Villena JA, Martin I, Viñas O, Cormand B, Iglesias R, Mampel T, *et al.* ETS transcription factors regulate the expression of the gene for the human mitochondrial ATP synthase beta-subunit. *J Biol Chem* 1994;**269**:32649–54.
60 Fan L, Sweet DR, Prosdocimo DA, Vinayachandran V, Chan ER, Zhang R, *et al.* Muscle Krüppel-like factor 15 regulates lipid flux and systemic metabolic homeostasis. *J Clin Invest* 2021;**131**:e139496. https://doi.org/10.1172/JCI139496.
61 Liao X, Zhang R, Lu Y, Prosdocimo DA, Sangwung P, Zhang L, *et al.* Kruppel-like factor 4 is critical for transcriptional control of cardiac mitochondrial homeostasis. *J Clin Invest* 2015;**125**:3461–76. https://doi.org/10.1172/JCI79964.
62 Wilson BC, Boehme L, Annibali A, Hodgkinson A, Carroll TS, Oakey RJ, *et al.* Intellectual disability-associated factor Zbtb11 cooperates with NRF-2/GABP to control mitochondrial function. *Nat Commun* 2020;**11**:5469. https://doi.org/10.1038/s41467-020-19205-x.
63 Cheng Z. FoxO transcription factors in mitochondrial homeostasis. *Biochem J* 2022;**479**:525–36. https://doi.org/10.1042/BCJ20210777.
64 Peng S, Li W, Hou N, Huang N. A Review of FoxO1-Regulated Metabolic Diseases and Related Drug Discoveries. *Cells* 2020;**9**:184. https://doi.org/10.3390/cells9010184.
65 Shintaku J, Peterson JM, Talbert EE, Gu J-M, Ladner KJ, Williams DR, *et al.* MyoD Regulates Skeletal Muscle Oxidative Metabolism Cooperatively with Alternative NF-κB. *Cell Rep* 2016;**17**:514–26. https://doi.org/10.1016/j.celrep.2016.09.010.
66 He M, Pei Z, Mohsen A-W, Watkins P, Murdoch G, Van Veldhoven PP, *et al.* Identification and characterization of new long chain Acyl-CoA dehydrogenases. *Mol Genet Metab* 2011;**102**:418–29. https://doi.org/10.1016/j.ymgme.2010.12.005.
67 Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-Associated SNPs Are More Likely to Be eQTLs: Annotation to Enhance Discovery from GWAS. *PLoS Genet* 2010;**6**:e1000888. https://doi.org/10.1371/journal.pgen.1000888.
68 Fleischman JY, Qi NR, Treutelaar MK, Britton SL, Koch LG, Li JZ, *et al.* Intrinsic cardiorespiratory fitness modulates clinical and molecular response to caloric restriction. *Mol Metab* 2023;**68**:101668. https://doi.org/10.1016/j.molmet.2023.101668.
69 Kelahmetoglu Y. Comparative Analysis of Skeletal Muscle Transcriptional Signatures Associated With Aerobic Exercise Capacity or Response to Training in Humans and Rats. *Front Endocrinol* 2020;**11**:12.
70 Saltin B. Hemodynamic adaptations to exercise. *Am J Cardiol* 1985;**55**:42D-47D. https://doi.org/10.1016/0002-9149(85)91054-9.
71 Richardson RS, Grassi B, Gavin TP, Haseler LJ, Tagore K, Roca J, *et al.* Evidence of O2 supply-dependentV˙o 2 max in the exercise-trained human quadriceps. *J Appl Physiol* 1999;**86**:1048–53. https://doi.org/10.1152/jappl.1999.86.3.1048.
72 Hepple RT, Mackinnon SLM, Goodman JM, Thomas SG, Plyley MJ. Resistance and aerobic training in older men: effects onV˙ $O_{2\ peak}$ and the capillary supply to skeletal muscle. *J Appl Physiol* 1997;**82**:1305–10. https://doi.org/10.1152/jappl.1997.82.4.1305.
73 Venables MC, Achten J, Jeukendrup AE. Determinants of fat oxidation during exercise in healthy men and women: a cross-sectional study. *J Appl Physiol* 2005;**98**:160–7. https://doi.org/10.1152/japplphysiol.00662.2003.
74 Achten J, Jeukendrup AE. Maximal Fat Oxidation During Exercise in Trained Men. *Int J Sports Med* 2003;**24**:603–8. https://doi.org/10.1055/s-2003-43265.
75 Chávez-Guevara IA, Hernández-Torres RP, González-Rodríguez E, Ramos-Jiménez A, Amaro-Gahete FJ. Biomarkers and genetic polymorphisms associated with maximal fat oxidation during physical exercise: implications for metabolic health and sports performance. *Eur J Appl Physiol* 2022;**122**:1773–95. https://doi.org/10.1007/s00421-022-04936-0.

76 Bye A, Langaas M, Høydal MA, Kemi OJ, Heinrich G, Koch LG, *et al.* Aerobic capacity-dependent differences in cardiac gene expression. *Physiol Genomics* 2008;**33**:100–9. https://doi.org/10.1152/physiolgenomics.00269.2007.
77 Bowden-Davies K, Connolly J, Burghardt P, Koch LG, Britton SL, Burniston JG. Label-free profiling of white adipose tissue of rats exhibiting high or low levels of intrinsic exercise capacity. *Proteomics* 2015;**15**:2342–9. https://doi.org/10.1002/pmic.201400537.
78 Petersen KF, Dufour S, Befroy D, Garcia R, Shulman GI. Impaired mitochondrial activity in the insulin-resistant offspring of patients with type 2 diabetes. *N Engl J Med* 2004;**350**:664–71. https://doi.org/10.1056/NEJMoa031314.
79 Mihalik SJ, Goodpaster BH, Kelley DE, Chace DH, Vockley J, Toledo FGS, *et al.* Increased levels of plasma acylcarnitines in obesity and type 2 diabetes and identification of a marker of glucolipotoxicity. *Obes Silver Spring Md* 2010;**18**:1695–700. https://doi.org/10.1038/oby.2009.510.
80 Koves TR, Ussher JR, Noland RC, Slentz D, Mosedale M, Ilkayeva O, *et al.* Mitochondrial overload and incomplete fatty acid oxidation contribute to skeletal muscle insulin resistance. *Cell Metab* 2008;**7**:45–56. https://doi.org/10.1016/j.cmet.2007.10.013.
81 Summers SA, Goodpaster BH. CrossTalk proposal: Intramyocellular ceramide accumulation does modulate insulin resistance. *J Physiol* 2016;**594**:3167–70. https://doi.org/10.1113/JP271676.
82 Davies MN, Kjalarsdottir L, Thompson JW, Dubois LG, Stevens RD, Ilkayeva OR, *et al.* The Acetyl Group Buffering Action of Carnitine Acetyltransferase Offsets Macronutrient-Induced Lysine Acetylation of Mitochondrial Proteins. *Cell Rep* 2016;**14**:243–54. https://doi.org/10.1016/j.celrep.2015.12.030.
83 Timmis JN, Ayliffe MA, Huang CY, Martin W. Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat Rev Genet* 2004;**5**:123–35. https://doi.org/10.1038/nrg1271.
84 Swigonová Z, Mohsen A-W, Vockley J. Acyl-CoA dehydrogenases: Dynamic history of protein family evolution. *J Mol Evol* 2009;**69**:176–93. https://doi.org/10.1007/s00239-009-9263-0.
85 Paquay S, Duraffourd J, Bury M, Heremans IP, Caligiore F, Gerin I, *et al.* ACAD10 and ACAD11 allow entry of 4-hydroxy fatty acids into β-oxidation. *Cell Mol Life Sci CMLS* 2024;**81**:367. https://doi.org/10.1007/s00018-024-05397-8.
86 Rashan EH, Bartlett AK, Khana DB, Zhang J, Jain R, Wade G, *et al.* ACAD10 and ACAD11 enable mammalian 4-hydroxy acid lipid catabolism. *Nat Struct Mol Biol* 2025. https://doi.org/10.1038/s41594-025-01596-4.
87 Kamarulzaman NT, Makpol S. The link between Mitochondria and Sarcopenia. *J Physiol Biochem* 2025;**81**:1–20. https://doi.org/10.1007/s13105-024-01062-7.
88 Holloway GP, Bonen A, Spriet LL. Regulation of skeletal muscle mitochondrial fatty acid metabolism in lean and obese individuals. *Am J Clin Nutr* 2009;**89**:455S-462S. https://doi.org/10.3945/ajcn.2008.26717B.
89 Serra D, Mera P, Malandrino MI, Mir JF, Herrero L. Mitochondrial Fatty Acid Oxidation in Obesity. *Antioxid Redox Signal* 2013;**19**:269–84. https://doi.org/10.1089/ars.2012.4875.
90 Orchard P, Manickam N, Ventresca C, Vadlamudi S, Varshney A, Rai V, *et al.* Human and rat skeletal muscle single-nuclei multi-omic integrative analyses nominate causal cell types, regulatory elements, and SNPs for complex traits. *Genome Res* 2021;**31**:2258–75. https://doi.org/10.1101/gr.268482.120.
91 Henikoff S, Henikoff J, Ahmad K. CUT&Tag-direct with CUTAC v3 2020. https://doi.org/10.17504/protocols.io.bqwvmxe6.
92 Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, *et al.* Welcome to the Tidyverse. *J Open Source Softw* 2019;**4**:1686. https://doi.org/10.21105/joss.01686.
93 Lawrence M, Huber W, Pagès H, Aboyoun P, Carlson M, Gentleman R, *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput Biol* 2013;**9**:e1003118. https://doi.org/10.1371/journal.pcbi.1003118.
94 Lee S, Cook D, Lawrence M. plyranges: a grammar of genomic data transformation. *Genome Biol* 2019;**20**:4. https://doi.org/10.1186/s13059-018-1597-8.
95 Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* 2016;**32**:2847–9. https://doi.org/10.1093/bioinformatics/btw313.
96 Gu Z. Complex heatmap visualization. *iMeta* 2022;**1**:e43. https://doi.org/10.1002/imt2.43.
97 Mölder F, Jablonski KP, Letcher B, Hall MB, Tomkins-Tinch CH, Sochat V, *et al.* Sustainable data analysis with Snakemake. *F1000Research* 2021;**10**:33. https://doi.org/10.12688/f1000research.29032.2.
98 Martin FJ, Amode MR, Aneja A, Austine-Orimoloye O, Azov AG, Barnes I, *et al.* Ensembl 2023. *Nucleic Acids Res* 2023;**51**:D933–41. https://doi.org/10.1093/nar/gkac958.

99 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20. https://doi.org/10.1093/bioinformatics/btu170.
100 Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;**29**:15–21. https://doi.org/10.1093/bioinformatics/bts635.
101 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* 2013.
102 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;**25**:2078–9. https://doi.org/10.1093/bioinformatics/btp352.
103 *Picard Tools - By Broad Institute*. n.d. URL: http://broadinstitute.github.io/picard/ (Accessed 12 June 2021).
104 Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–2. https://doi.org/10.1093/bioinformatics/btq033.
105 Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 2008;**9**:R137. https://doi.org/10.1186/gb-2008-9-9-r137.
106 Orchard P, Kyono Y, Hensley J, Kitzman JO, Parker SCJ. Quantification, Dynamic Visualization, and Validation of Bias in ATAC-Seq Data with ataqv. *Cell Syst* 2020;**10**:298-306.e4. https://doi.org/10.1016/j.cels.2020.02.009.
107 Schep AN, Wu B, Buenrostro JD, Greenleaf WJ. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* 2017;**14**:975–8. https://doi.org/10.1038/nmeth.4401.
108 Amemiya HM, Kundaje A, Boyle AP. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci Rep* 2019;**9**:9354. https://doi.org/10.1038/s41598-019-45839-z.
109 Broman K. broman: Karl Broman's R Code 2022.
110 Davide Risso [Aut C. RUVSeq 2017. https://doi.org/10.18129/B9.BIOC.RUVSEQ.
111 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550. https://doi.org/10.1186/s13059-014-0550-8.
112 Dolgalev I. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format 2022.
113 Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, *et al.* BioMart and Bioconductor: a powerful link between biological databases and microarray data analysis. *Bioinformatics* 2005;**21**:3439–40. https://doi.org/10.1093/bioinformatics/bti525.
114 Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Stat Methodol* 1995;**57**:289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
115 Lee C, Patil S, Sartor MA. RNA-Enrich: a cut-off free functional enrichment testing method for RNA-seq with improved detection power. *Bioinformatics* 2016;**32**:1100–2. https://doi.org/10.1093/bioinformatics/btv694.
116 Welch RP, Lee C, Imbriano PM, Patil S, Weymouth TE, Smith RA, *et al.* ChIP-Enrich: gene set enrichment testing for ChIP-seq data. *Nucleic Acids Res* 2014;**42**:e105–e105. https://doi.org/10.1093/nar/gku463.
117 Korotkevich G, Sukhov V, Budin N, Shpak B, Artyomov MN, Sergushichev A. Fast gene set enrichment analysis 2016. https://doi.org/10.1101/060012.
118 Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, *et al.* clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *The Innovation* 2021;**2**:100141. https://doi.org/10.1016/j.xinn.2021.100141.
119 Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, *et al.* Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 2019;**10**:1523. https://doi.org/10.1038/s41467-019-09234-6.
120 Currin KW, Erdos MR, Narisu N, Rai V, Vadlamudi S, Perrin HJ, *et al.* Genetic effects on liver chromatin accessibility identify disease regulatory variants. *Am J Hum Genet* 2021;**108**:1169–89. https://doi.org/10.1016/j.ajhg.2021.05.001.
121 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;**43**:491–8. https://doi.org/10.1038/ng.806.
122 Ramdas S, Ozel AB, Treutelaar MK, Holl K, Mandel M, Woods LCS, *et al.* Extended regions of suspected mis-assembly in the rat reference genome. *Sci Data* 2019;**6**:39. https://doi.org/10.1038/s41597-019-0041-6.
123 Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 2011;**27**:2987–93. https://doi.org/10.1093/bioinformatics/btr509.

124 Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, *et al.* Twelve years of SAMtools and BCFtools. *GigaScience* 2021;**10**:giab008. https://doi.org/10.1093/gigascience/giab008.
125 De Jong TV, Pan Y, Rastas P, Munro D, Tutaj M, Akil H, *et al.* A revamped rat reference genome improves the discovery of genetic diversity in laboratory rats 2023. https://doi.org/10.1101/2023.04.13.536694.
126 Rubinacci S, Delaneau O, Marchini J. Genotype imputation using the Positional Burrows Wheeler Transform. *PLOS Genet* 2020;**16**:e1009049. https://doi.org/10.1371/journal.pgen.1009049.
127 Hofmeister RJ, Ribeiro DM, Rubinacci S, Delaneau O. Accurate rare variant phasing of whole-genome and whole-exome sequencing data in the UK Biobank. *Nat Genet* 2023;**55**:1243–9. https://doi.org/10.1038/s41588-023-01415-w.
128 Hermsen R, De Ligt J, Spee W, Blokzijl F, Schäfer S, Adami E, *et al.* Genomic landscape of rat strain and substrain variation. *BMC Genomics* 2015;**16**:357. https://doi.org/10.1186/s12864-015-1594-1.
129 Wright S. The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. *Evolution* 1965;**19**:395. https://doi.org/10.2307/2406450.
130 Paul Shannon MR. MotifDb 2017. https://doi.org/10.18129/B9.BIOC.MOTIFDB.
131 Tan G, Lenhard B. TFBSTools: an R/bioconductor package for transcription factor binding site analysis. *Bioinformatics* 2016;**32**:1555–6. https://doi.org/10.1093/bioinformatics/btw024.
132 Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 2011;**27**:1017–8. https://doi.org/10.1093/bioinformatics/btr064.
133 McLeay RC, Bailey TL. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. *BMC Bioinformatics* 2010;**11**:165. https://doi.org/10.1186/1471-2105-11-165.
134 Fort A, Panousis NI, Garieri M, Antonarakis SE, Lappalainen T, Dermitzakis ET, *et al. MBV* : a method to solve sample mislabeling and detect technical bias in large combined genotype and sequencing assay datasets. *Bioinformatics* 2017;**33**:1895–7. https://doi.org/10.1093/bioinformatics/btx074.
135 Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* 2012;**44**:821–4. https://doi.org/10.1038/ng.2310.
136 Davis JR, Fresard L, Knowles DA, Pala M, Bustamante CD, Battle A, *et al.* An Efficient Multiple-Testing Adjustment for eQTL Studies that Accounts for Linkage Disequilibrium between Variants. *Am J Hum Genet* 2016;**98**:216–24. https://doi.org/10.1016/j.ajhg.2015.11.021.


Low & High Capacity Runners
Fitness
Gene Expression
Chromatin Accessibility
5x Histone Marks
Distal Peak
Nearest Gene
Promoter Peak
Target Gene
Distal Regulatory Networks
Population Divergence
Target gene
Motif
Target gene
Transcription Factor Networks
Independent Validation
F2 Population
Gene Expression
Chromatin Accessibility
caQTL
SNP
Gene
eQTL
Genotype Association
Affected Pathways
Angiogenesis
Glucose
BCAA
Fatty Acid
Substrate Oxidation

Graphical Abstract

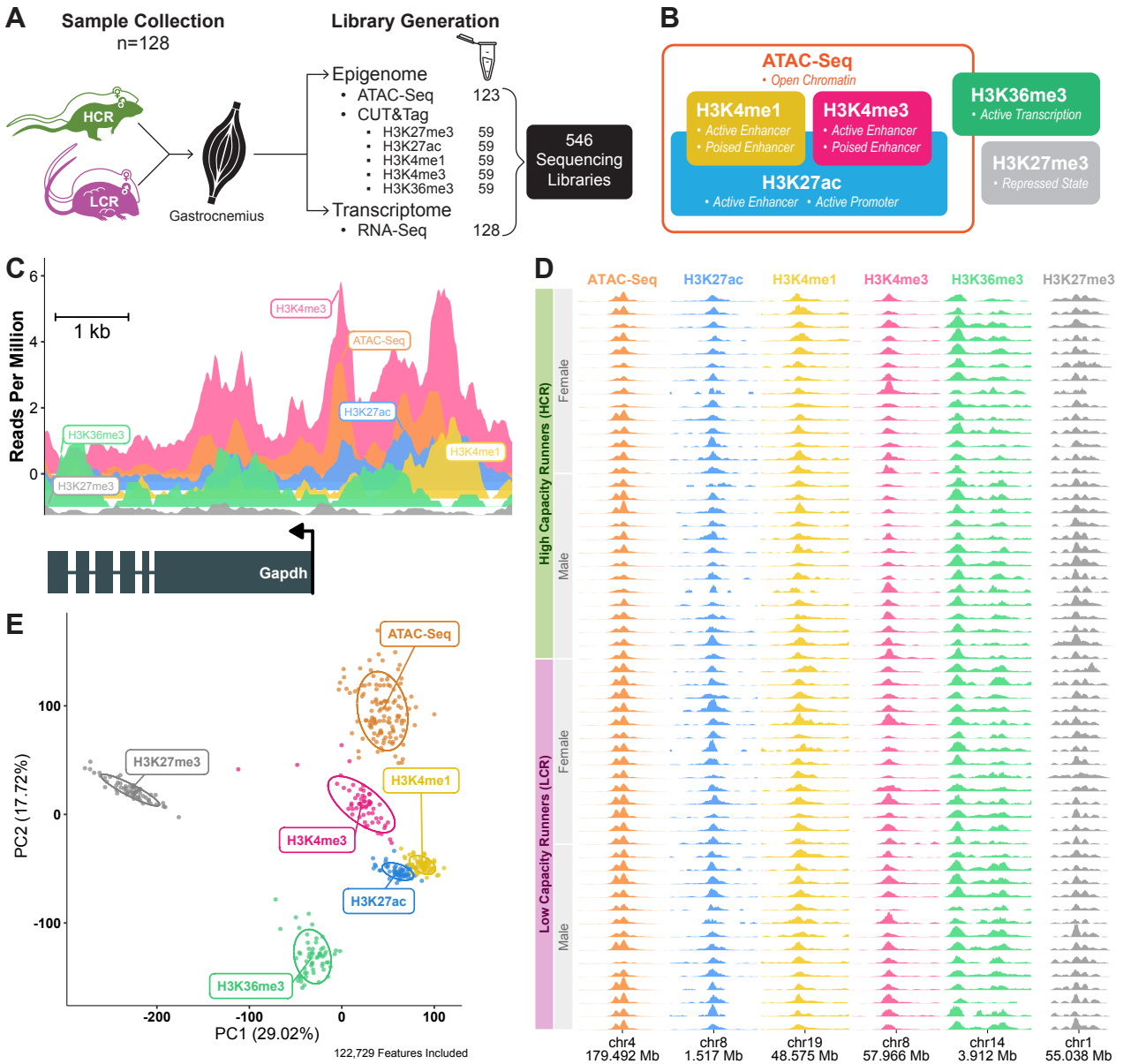

A
Sample Collection
n=128
HCR
LCR
Gastrocnemius
Library Generation
Epigenome
ATAC-Seq 123
CUT&Tag
H3K27me3 59
H3K27ac 59
H3K4me1 59
H3K4me3 59
H3K36me3 59
Transcriptome
RNA-Seq 128
546 Sequencing Libraries
B
ATAC-Seq
Open Chromatin
H3K4me1
Active Enhancer
Poised Enhancer
H3K4me3
Active Enhancer
Poised Enhancer
H3K27ac
Active Enhancer
Active Promoter
H3K36me3
Active Transcription
H3K27me3
Repressed State
C
Reads Per Million
1 kb
H3K4me3
ATAC-Seq
H3K27ac
H3K36me3
H3K4me1
H3K27me3
Gapdh
D
ATAC-Seq
H3K27ac
H3K4me1
H3K4me3
H3K36me3
H3K27me3
High Capacity Runners (HCR)
Low Capacity Runners (LCR)
Female
Male
Female
Male
chr4 179.492 Mb
chr8 1.517 Mb
chr19 48.575 Mb
chr8 57.966 Mb
chr14 3.912 Mb
chr1 55.038 Mb
E
ATAC-Seq
H3K27me3
H3K4me1
H3K4me3
H3K27ac
H3K36me3
PC2 (17.72%)
PC1 (29.02%)
122,729 Features Included

Figure 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 Figure 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 Figure 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.

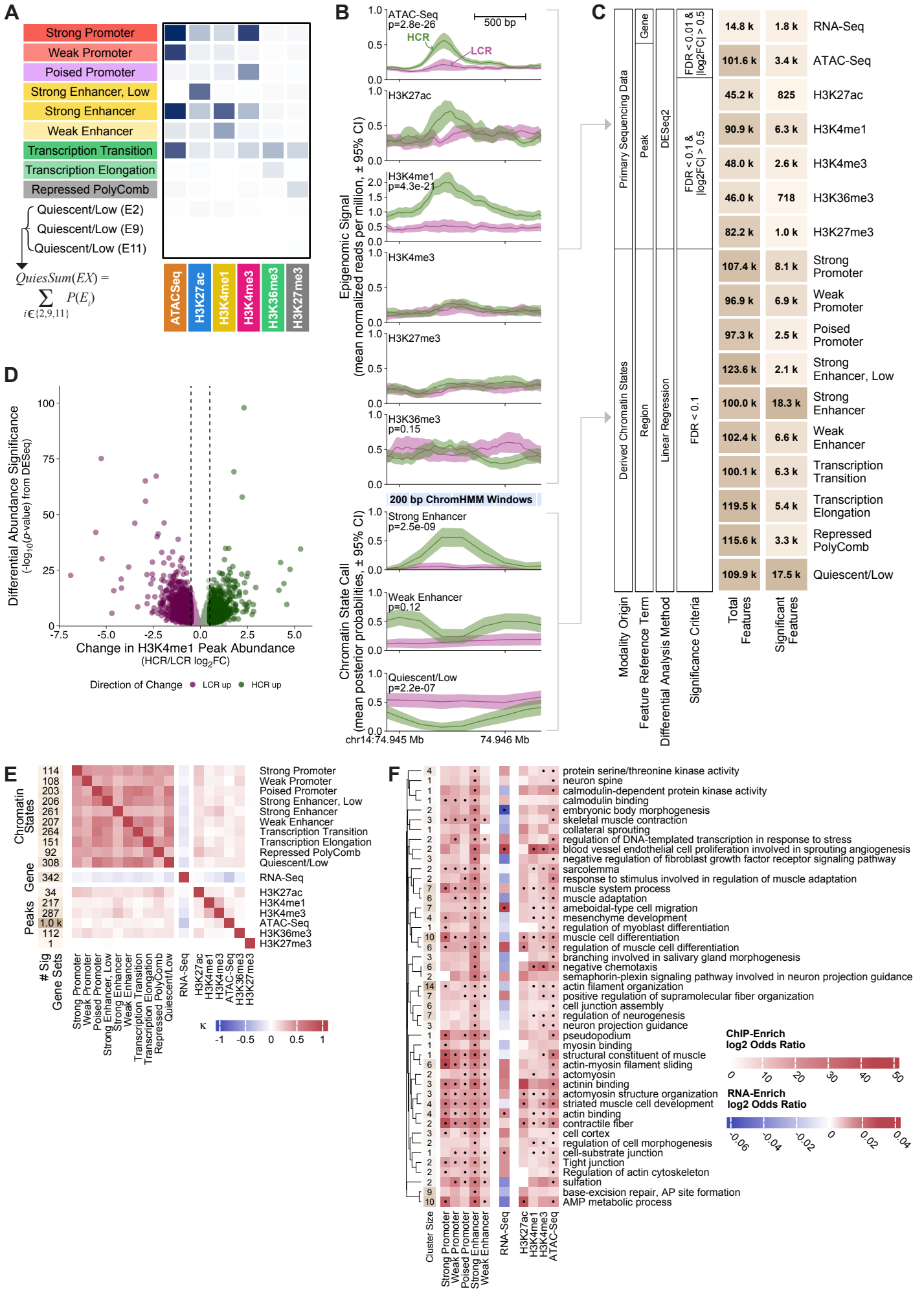

A
Strong Promoter
Weak Promoter
Poised Promoter
Strong Enhancer, Low
Strong Enhancer
Weak Enhancer
Transcription Transition
Transcription Elongation
Repressed PolyComb
Quiescent/Low (E2)
Quiescent/Low (E9)
Quiescent/Low (E11)
$QuiesSum(EX) = \sum_{i \in \{2,9,11\}} P(E_i)$
ATACSeq
H3K27ac
H3K4me1
H3K4me3
H3K36me3
H3K27me3
B
ATAC-Seq
p=2.8e-26
500 bp
HCR
LCR
H3K27ac
H3K4me1
p=4.3e-21
H3K4me3
H3K27me3
H3K36me3
p=0.15
Epigenomic Signal
(mean normalized reads per million, ± 95% CI)
200 bp ChromHMM Windows
Strong Enhancer
p=2.5e-09
Weak Enhancer
p=0.12
Quiescent/Low
p=2.2e-07
Chromatin State Call
(mean posterior probabilities, ± 95% CI)
chr14:74.945 Mb
74.946 Mb
C
Primary Sequencing Data
Derived Chromatin States
Gene
Peak
Region
DESeq2
Linear Regression
FDR < 0.01 & |log2FC| > 0.5
FDR < 0.1 & |log2FC| > 0.5
FDR < 0.1
Modality Origin
Feature Reference Term
Differential Analysis Method
Significance Criteria
Total Features
Significant Features
14.8 k 1.8 k RNA-Seq
101.6 k 3.4 k ATAC-Seq
45.2 k 825 H3K27ac
90.9 k 6.3 k H3K4me1
48.0 k 2.6 k H3K4me3
46.0 k 718 H3K36me3
82.2 k 1.0 k H3K27me3
107.4 k 8.1 k Strong Promoter
96.9 k 6.9 k Weak Promoter
97.3 k 2.5 k Poised Promoter
123.6 k 2.1 k Strong Enhancer, Low
100.0 k 18.3 k Strong Enhancer
102.4 k 6.6 k Weak Enhancer
100.1 k 6.3 k Transcription Transition
119.5 k 5.4 k Transcription Elongation
115.6 k 3.3 k Repressed PolyComb
109.9 k 17.5 k Quiescent/Low
D
Differential Abundance Significance
(-log10(p-value) from DESeq)
Change in H3K4me1 Peak Abundance
(HCR/LCR log2FC)
Direction of Change
LCR up
HCR up
E
Chromatin States
Gene
Peaks
Sig Gene Sets
Strong Promoter
Weak Promoter
Poised Promoter
Strong Enhancer, Low
Strong Enhancer
Weak Enhancer
Transcription Transition
Transcription Elongation
Repressed PolyComb
Quiescent/Low
RNA-Seq
H3K27ac
H3K4me1
H3K4me3
ATAC-Seq
H3K36me3
H3K27me3
κ
F
protein serine/threonine kinase activity
neuron spine
calmodulin-dependent protein kinase activity
calmodulin binding
embryonic body morphogenesis
skeletal muscle contraction
collateral sprouting
regulation of DNA-templated transcription in response to stress
blood vessel endothelial cell proliferation involved in sprouting angiogenesis
negative regulation of fibroblast growth factor receptor signaling pathway
sarcolemma
response to stimulus involved in regulation of muscle adaptation
muscle system process
muscle adaptation
ameboidal-type cell migration
mesenchyme development
regulation of myoblast differentiation
muscle cell differentiation
regulation of muscle cell differentiation
branching involved in salivary gland morphogenesis
negative chemotaxis
semaphorin-plexin signaling pathway involved in neuron projection guidance
actin filament organization
positive regulation of supramolecular fiber organization
cell junction assembly
regulation of neurogenesis
neuron projection guidance
pseudopodium
myosin binding
structural constituent of muscle
actin-myosin filament sliding
actomyosin
actinin binding
actomyosin structure organization
striated muscle cell development
actin binding
contractile fiber
cell cortex
regulation of cell morphogenesis
cell-substrate junction
Tight junction
Regulation of actin cytoskeleton
sulfation
base-excision repair, AP site formation
AMP metabolic process
Cluster Size
ChIP-Enrich
log2 Odds Ratio
RNA-Enrich
log2 Odds Ratio

Figure 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. 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 Figure S6.

(E) Cross-modality comparison of significant enrichment. Heatmap of pairwise kappa coefficients measuring the similarity of significantly enriched gene sets compared between all data modalities. A number greater than 0 represents the number of shared gene sets is greater than expected by chance, while a value less than 0 indicates that the overlap is lower than expected by chance. The annotation on the left side of the heatmap presents the count of significant gene sets (FDR < 0.05) returned by RNA-Enrich and ChIP-Enrich when epigenomic features were annotated to genes with the nearest TSS.

(F) Strong Enhancer significantly enriched gene sets ($\log_2$ odds ratio > 0 & FDR < 0.05) were collapsed into parent terms via kappa clustering of the genes (Cluster Size) and compared to additional modalities. Log_2 odds ratios of the most significant gene set within the collapsed term are plotted for the top 50% of the clusters, with significant clusters annotated with ‘•’. Rows are further hierarchically clustered based on kappa similarity of all genes collapsed into the parent term.

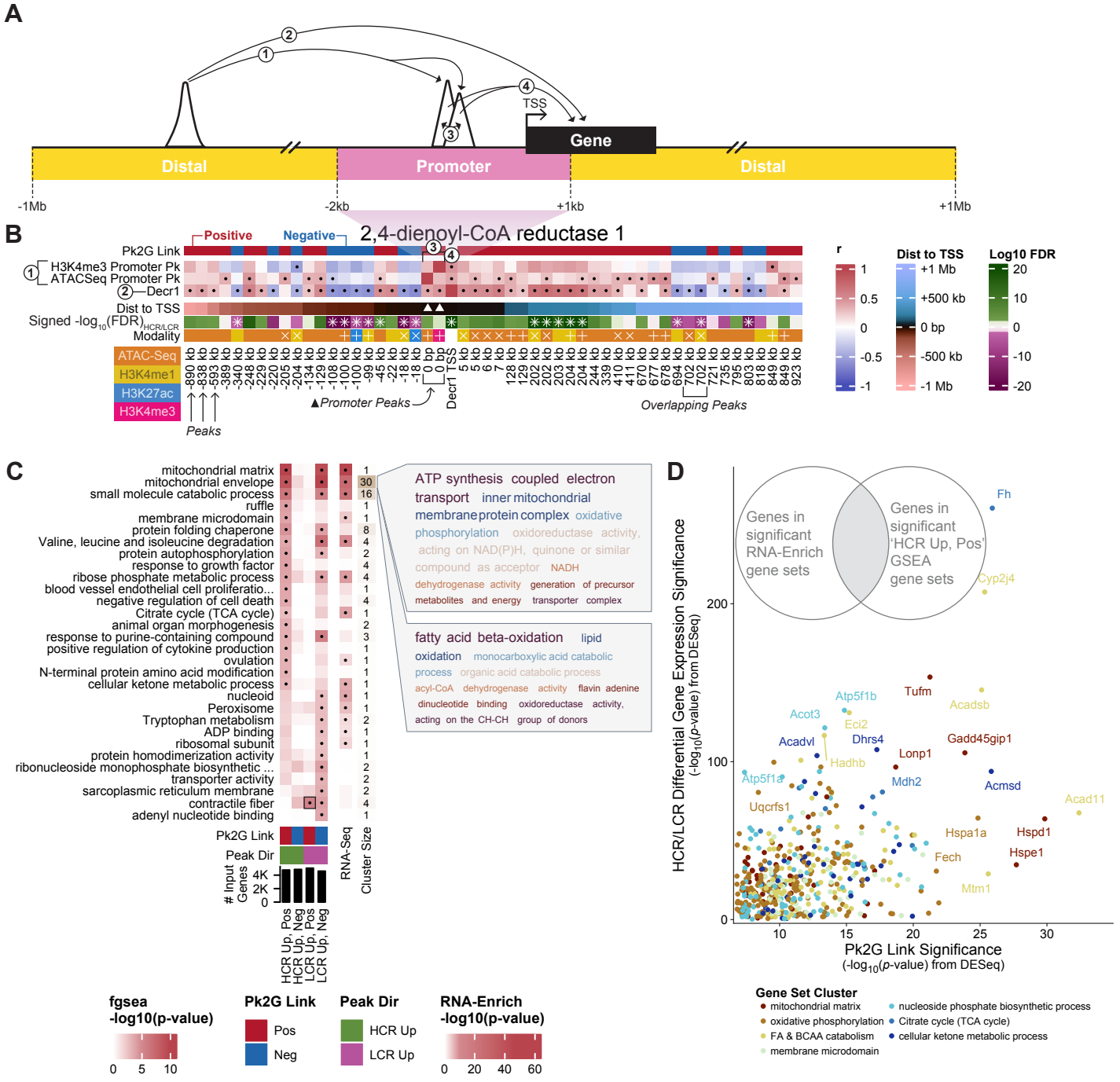

A
TSS
Gene
Distal
Promoter
Distal
-1Mb
-2kb
+1kb
+1Mb
B
2,4-dienoyl-CoA reductase 1
Positive
Negative
Pk2G Link
H3K4me3 Promoter Pk
ATACSeq Promoter Pk
Decr1
Dist to TSS
Signed -log10(FDR)HCR/LCR
Modality
ATAC-Seq
H3K4me1
H3K27ac
H3K4me3
Peaks
Promoter Peaks
Overlapping Peaks
r
Dist to TSS
Log10 FDR
C
mitochondrial matrix
mitochondrial envelope
small molecule catabolic process
ruffle
membrane microdomain
protein folding chaperone
Valine, leucine and isoleucine degradation
protein autophosphorylation
response to growth factor
ribose phosphate metabolic process
blood vessel endothelial cell proliferatio...
negative regulation of cell death
Citrate cycle (TCA cycle)
animal organ morphogenesis
response to purine-containing compound
positive regulation of cytokine production
ovulation
N-terminal protein amino acid modification
cellular ketone metabolic process
nucleoid
Peroxisome
Tryptophan metabolism
ADP binding
ribosomal subunit
protein homodimerization activity
ribonucleoside monophosphate biosynthetic ...
transporter activity
sarcoplasmic reticulum membrane
contractile fiber
adenyl nucleotide binding
ATP synthesis coupled electron transport inner mitochondrial membrane protein complex oxidative phosphorylation oxidoreductase activity, acting on NAD(P)H, quinone or similar compound as acceptor NADH dehydrogenase activity generation of precursor metabolites and energy transporter complex
fatty acid beta-oxidation lipid oxidation monocarboxylic acid catabolic process organic acid catabolic process acyl-CoA dehydrogenase activity flavin adenine dinucleotide binding oxidoreductase activity, acting on the CH-CH group of donors
Pk2G Link
Peak Dir
Input Genes
RNA-Seq
Cluster Size
HCR Up, Pos
HCR Up, Neg
LCR Up, Pos
LCR Up, Neg
fgsea -log10(p-value)
Pk2G Link
Pos
Neg
Peak Dir
HCR Up
LCR Up
RNA-Enrich -log10(p-value)
D
Genes in significant RNA-Enrich gene sets
Genes in significant 'HCR Up, Pos' GSEA gene sets
HCR/LCR Differential Gene Expression Significance (-log10(p-value) from DESeq)
Pk2G Link Significance (-log10(p-value) from DESeq)
Fh
Cyp2j4
Tufm
Acadsb
Atp5f1b
Acot3
Eci2
Acadvl
Dhrs4
Gadd45gip1
Lonp1
Hadhb
Atp5f1a
Mdh2
Acmsd
Acad11
Uqcrfs1
Hspa1a
Hspd1
Fech
Hspe1
Mtm1
Gene Set Cluster
mitochondrial matrix
oxidative phosphorylation
FA & BCAA catabolism
membrane microdomain
nucleoside phosphate biosynthetic process
Citrate cycle (TCA cycle)
cellular ketone metabolic process

Figure 3. Linking Epigenomic Features to Gene Expression

(A) Correlation framework epigenomic peaks to gene expression. Distal regulatory elements (yellow) reside within 1 Mb of the gene's TSS and are correlated with promoter peaks (1, pink) and gene expression (2). Promoter peaks are correlated with one another (3) and gene expression (4).

(B) Correlation of distal (columns) and promoter (rows and ▲-marked columns) elements surrounding Decr1, with the top annotation bar indicating the directionality of the Pk2G link (red=positive, blue=negative). The subsequent heatmap plots the spearman's correlation coefficient with significant relationships (FDR < 0.05) annotated by '•'. Columns are labeled and ordered by the distance of the peak from the TSS (Dist to TSS). The signed $-\log_{10}(\text{FDR})_{\text{HCR/LCR}}$ bar provides feature-specific differential results for HCR up (green) regions and LCR up (magenta) features, with an asterisk annotating those that pass the significance threshold provided in Figure 2C. The bottom-most 'Modality' annotation bar is colored based on data origin. Nearby and overlapping peaks are indicated by '+' and 'X', alternating between distinct instances of clustered regions. See additional examples in Figure S8.

(C) Ranked gene set enrichment analysis of Pk2G targets linked to differential (*p*-value < 0.1) 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.

(D) 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 502 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.

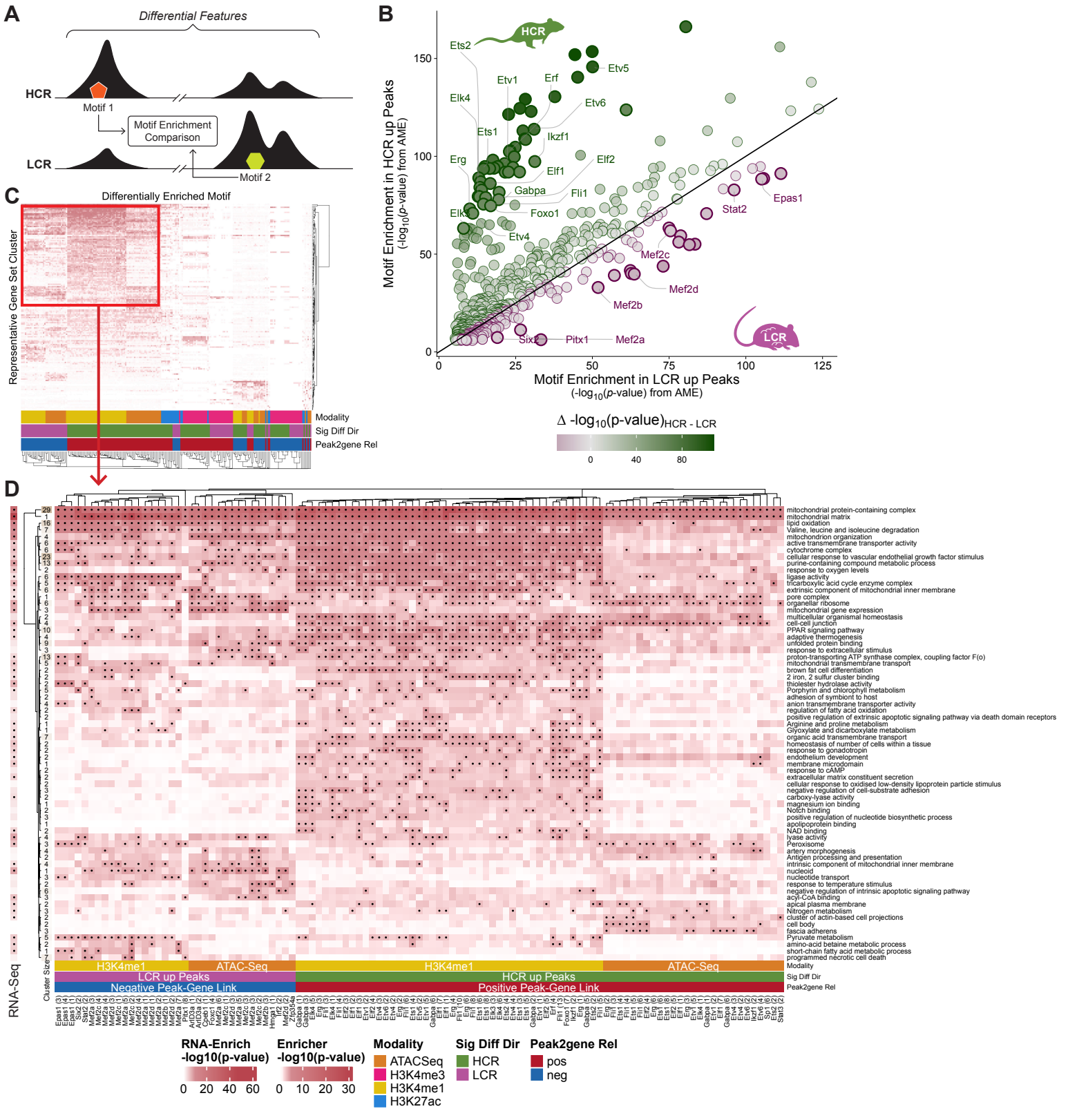

A
Differential Features
HCR
Motif 1
Motif Enrichment Comparison
LCR
Motif 2
B
HCR
LCR
Ets2
Etv5
Etv1
Erf
Elk4
Etv6
Ets1
Ikzf1
Erg
Elf2
Elf1
Gabpa
Fli1
Elk3
Foxo1
Etv4
Epas1
Stat2
Mef2c
Mef2d
Mef2b
Six2
Pitx1
Mef2a
Motif Enrichment in HCR up Peaks
(-log10(p-value) from AME)
Motif Enrichment in LCR up Peaks
(-log10(p-value) from AME)
Δ -log10(p-value)HCR - LCR
C
Differentially Enriched Motif
Representative Gene Set Cluster
Modality
Sig Diff Dir
Peak2gene Rel
D
mitochondrial protein-containing complex
mitochondrial matrix
lipid oxidation
Valine, leucine and isoleucine degradation
mitochondrion organization
active transmembrane transporter activity
cytochrome complex
cellular response to vascular endothelial growth factor stimulus
purine-containing compound metabolic process
response to oxygen levels
ligase activity
tricarboxylic acid cycle enzyme complex
extrinsic component of mitochondrial inner membrane
pore complex
organellar ribosome
mitochondrial gene expression
multicellular organismal homeostasis
cell-cell junction
PPAR signaling pathway
adaptive thermogenesis
unfolded protein binding
response to extracellular stimulus
proton-transporting ATP synthase complex, coupling factor F(o)
mitochondrial transmembrane transport
brown fat cell differentiation
2 iron, 2 sulfur cluster binding
thiolester hydrolase activity
Porphyrin and chlorophyll metabolism
adhesion of symbiont to host
anion transmembrane transporter activity
regulation of fatty acid oxidation
positive regulation of extrinsic apoptotic signaling pathway via death domain receptors
Arginine and proline metabolism
Glyoxylate and dicarboxylate metabolism
organic acid transmembrane transport
homeostasis of number of cells within a tissue
response to gonadotropin
endothelium development
membrane microdomain
response to cAMP
extracellular matrix constituent secretion
cellular response to oxidised low-density lipoprotein particle stimulus
negative regulation of cell-substrate adhesion
carboxy-lyase activity
magnesium ion binding
Notch binding
positive regulation of nucleotide biosynthetic process
apolipoprotein binding
NAD binding
lyase activity
Peroxisome
artery morphogenesis
Antigen processing and presentation
intrinsic component of mitochondrial inner membrane
nucleoid
nucleotide transport
response to temperature stimulus
negative regulation of intrinsic apoptotic signaling pathway
acyl-CoA binding
apical plasma membrane
Nitrogen metabolism
cluster of actin-based cell projections
cell body
fascia adherens
Pyruvate metabolism
amino-acid betaine metabolic process
short-chain fatty acid metabolic process
programmed necrotic cell death
RNA-Seq
Cluster Size
H3K4me1
ATAC-Seq
H3K4me1
ATAC-Seq
Modality
LCR up Peaks
HCR up Peaks
Sig Diff Dir
Negative Peak-Gene Link
Positive Peak-Gene Link
Peak2gene Rel
RNA-Enrich -log10(p-value)
Enricher -log10(p-value)
Modality
ATACSeq
H3K4me3
H3K4me1
H3K27ac
Sig Diff Dir
HCR
LCR
Peak2gene Rel
pos
neg

Figure 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.

(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 Figure S9.

(C) Over-representation analysis of genes linked to differentially enriched motifs outlined in panel B and Figure 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. 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. 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 hierarchically clustered; columns are kappa clustered by significant gene inclusion. 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.

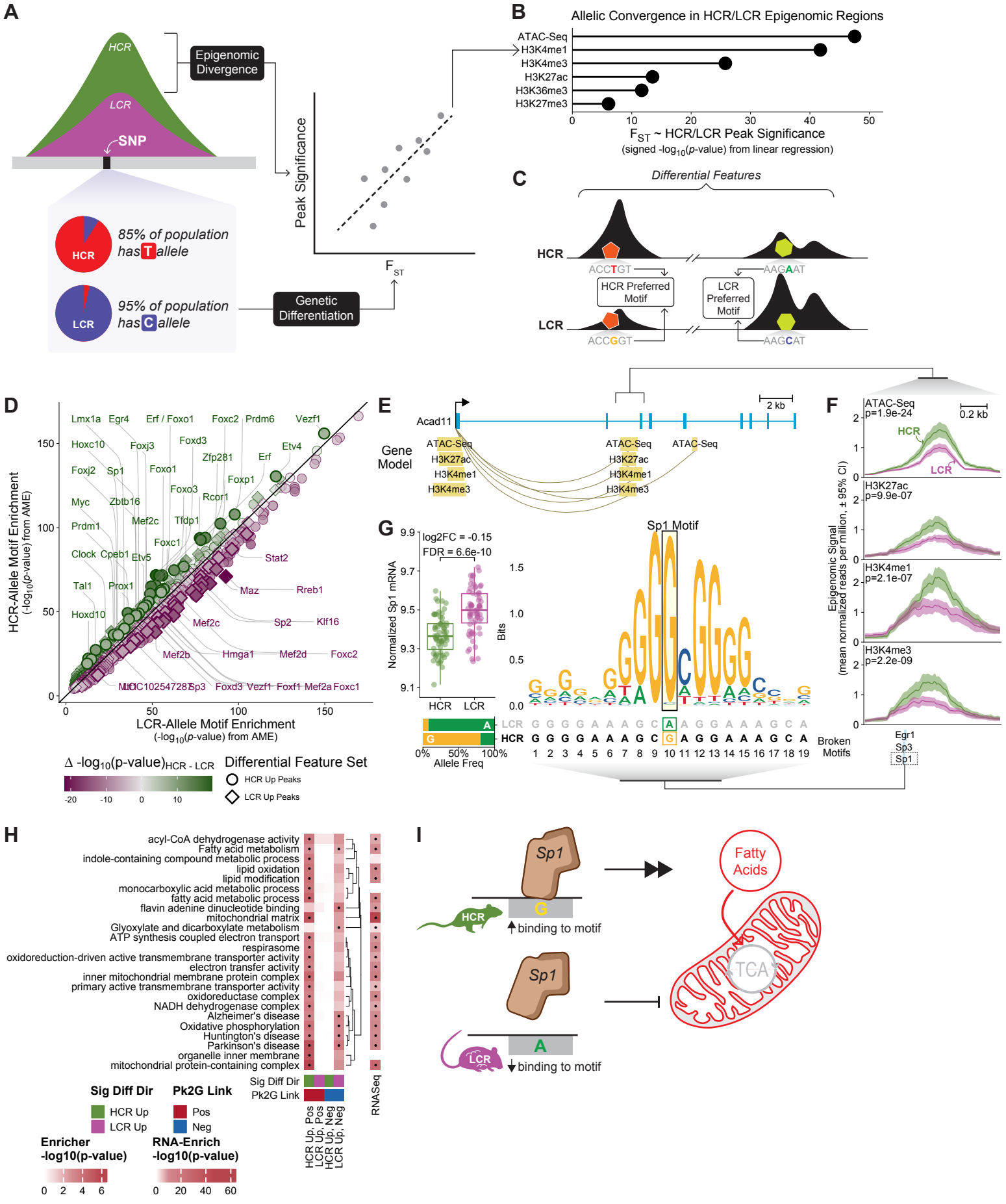

A
HCR
LCR
SNP
Epigenomic Divergence
Peak Significance
F_ST
85% of population has T allele
95% of population has C allele
Genetic Differentiation
B
Allelic Convergence in HCR/LCR Epigenomic Regions
ATAC-Seq
H3K4me1
H3K4me3
H3K27ac
H3K36me3
H3K27me3
F_ST ~ HCR/LCR Peak Significance
(signed -log10(p-value) from linear regression)
C
Differential Features
HCR
LCR
ACCTGT
AAGAAT
ACCGGT
AAGCAT
HCR Preferred Motif
LCR Preferred Motif
D
HCR-Allele Motif Enrichment (-log10(p-value) from AME)
LCR-Allele Motif Enrichment (-log10(p-value) from AME)
Lmx1a
Egr4
Erf / Foxo1
Foxc2
Prdm6
Vezf1
Hoxc10
Foxj3
Foxd3
Etv4
Zfp281
Erf
Foxj2
Sp1
Foxo1
Foxp1
Foxo3
Rcor1
Myc
Zbtb16
Mef2c
Tfdp1
Prdm1
Foxc1
Clock
Cpeb1
Etv5
Stat2
Tal1
Prox1
Maz
Rreb1
Hoxd10
Mef2c
Sp2
Klf16
Mef2b
Hmga1
Mef2d
Foxc2
LOC102547287
Sp3
Foxd3
Vezf1
Foxf1
Mef2a
Foxc1
Δ -log10(p-value)HCR - LCR
Differential Feature Set
HCR Up Peaks
LCR Up Peaks
E
Acad11
2 kb
Gene Model
ATAC-Seq
H3K27ac
H3K4me1
H3K4me3
F
ATAC-Seq p=1.9e-24
0.2 kb
HCR
LCR
H3K27ac p=9.9e-07
H3K4me1 p=2.1e-07
H3K4me3 p=2.2e-09
Epigenomic Signal (mean normalized reads per million, ± 95% CI)
Egr1
Sp3
Sp1
G
log2FC = -0.15
FDR = 6.6e-10
Normalized Sp1 mRNA
HCR
LCR
Sp1 Motif
Bits
Allele Freq
Broken Motifs
H
acyl-CoA dehydrogenase activity
Fatty acid metabolism
indole-containing compound metabolic process
lipid oxidation
lipid modification
monocarboxylic acid metabolic process
fatty acid metabolic process
flavin adenine dinucleotide binding
mitochondrial matrix
Glyoxylate and dicarboxylate metabolism
ATP synthesis coupled electron transport
respirasome
oxidoreduction-driven active transmembrane transporter activity
electron transfer activity
inner mitochondrial membrane protein complex
primary active transmembrane transporter activity
oxidoreductase complex
NADH dehydrogenase complex
Alzheimer's disease
Oxidative phosphorylation
Huntington's disease
Parkinson's disease
organelle inner membrane
mitochondrial protein-containing complex
Sig Diff Dir
Pk2G Link
RNASeq
HCR Up, Pos
LCR Up, Pos
HCR Up, Neg
LCR Up, Neg
Sig Diff Dir
HCR Up
LCR Up
Pk2G Link
Pos
Neg
Enricher -log10(p-value)
RNA-Enrich -log10(p-value)
I
Sp1
HCR
binding to motif
Sp1
LCR
binding to motif
Fatty Acids
TCA

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

(A) Schematic showing comparison of population differentiation (F_{ST}) and HCR/LCR epigenomic divergence carried out for each modality.

(B) 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 (y-axis), while controlling for technical variables and genomic location (coding vs noncoding). The signed -$\log_{10}$(*p*-value) for the F_{ST} term in the model is plotted on the x-axis.

(C) 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).

(D) 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 Figure S9.

(E) 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.

(F) 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.

(G) 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.

(H) 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).

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

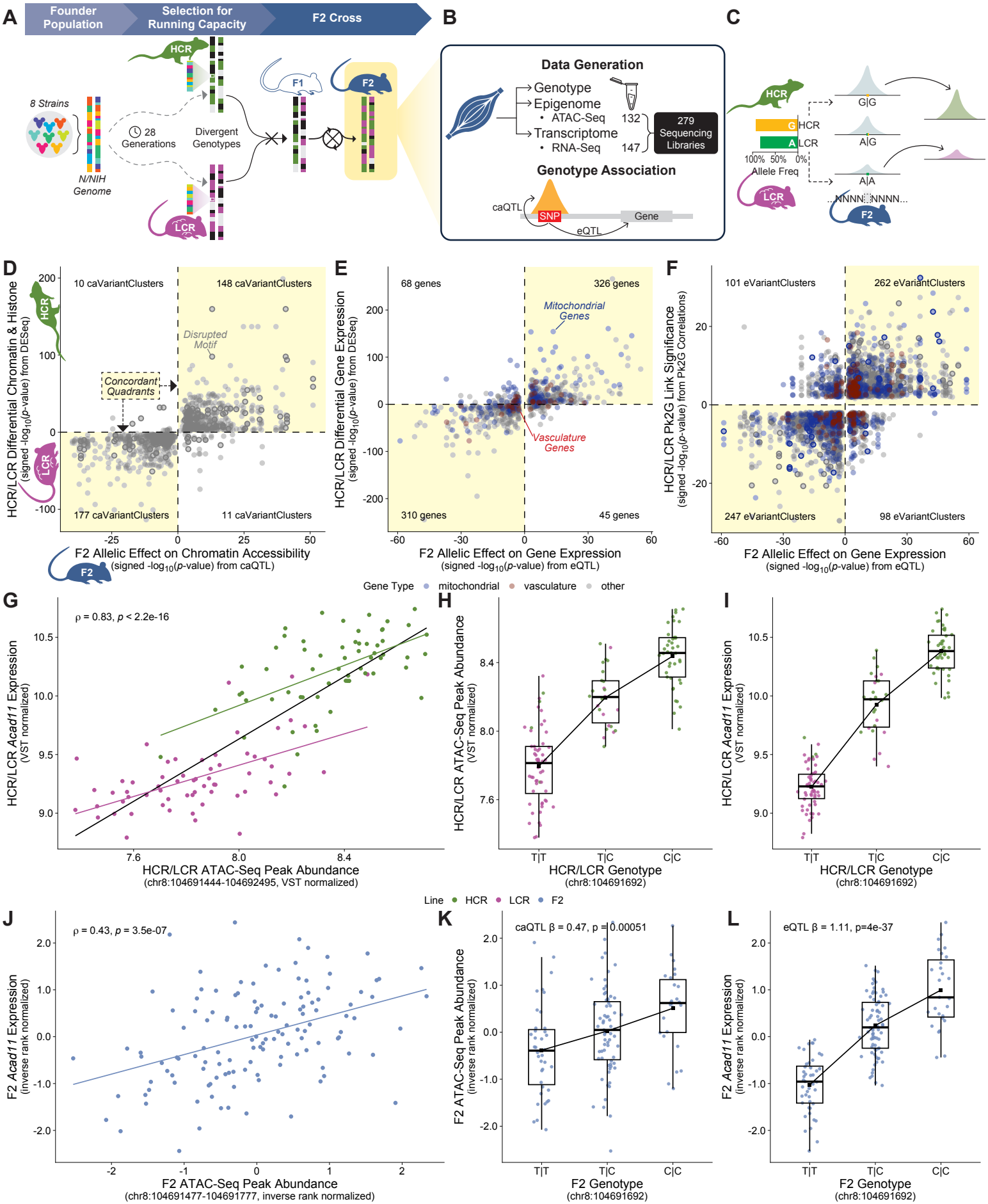

A
Founder Population
Selection for Running Capacity
F2 Cross
8 Strains
N/NIH Genome
28 Generations
Divergent Genotypes
HCR
LCR
F1
F2
B
Data Generation
Genotype
Epigenome
• ATAC-Seq 132
Transcriptome
• RNA-Seq 147
279 Sequencing Libraries
Genotype Association
caQTL
SNP
Gene
eQTL
C
HCR
LCR
G HCR
A LCR
100% 50% 0%
Allele Freq
G|G
A|G
A|A
...NNNN NNNN...
F2
D
HCR/LCR Differential Chromatin & Histone (signed -log10(p-value) from DESeq)
10 caVariantClusters
148 caVariantClusters
Disrupted Motif
Concordant Quadrants
177 caVariantClusters
11 caVariantClusters
F2 Allelic Effect on Chromatin Accessibility (signed -log10(p-value) from caQTL)
E
HCR/LCR Differential Gene Expression (signed -log10(p-value) from DESeq)
68 genes
326 genes
Mitochondrial Genes
Vasculature Genes
310 genes
45 genes
F2 Allelic Effect on Gene Expression (signed -log10(p-value) from eQTL)
F
HCR/LCR Pk2G Link Significance (signed -log10(p-value) from Pk2G Correlations)
101 eVariantClusters
262 eVariantClusters
247 eVariantClusters
98 eVariantClusters
F2 Allelic Effect on Gene Expression (signed -log10(p-value) from eQTL)
Gene Type mitochondrial vasculature other
G
ρ = 0.83, p < 2.2e-16
HCR/LCR Acad11 Expression (VST normalized)
HCR/LCR ATAC-Seq Peak Abundance (chr8:104691444-104692495, VST normalized)
H
HCR/LCR ATAC-Seq Peak Abundance (VST normalized)
T|T T|C C|C
HCR/LCR Genotype (chr8:104691692)
I
HCR/LCR Acad11 Expression (VST normalized)
T|T T|C C|C
HCR/LCR Genotype (chr8:104691692)
Line HCR LCR F2
J
ρ = 0.43, p = 3.5e-07
F2 Acad11 Expression (inverse rank normalized)
F2 ATAC-Seq Peak Abundance (chr8:104691477-104691777, inverse rank normalized)
K
caQTL β = 0.47, p = 0.00051
F2 ATAC-Seq Peak Abundance (inverse rank normalized)
T|T T|C C|C
F2 Genotype (chr8:104691692)
L
eQTL β = 1.11, p=4e-37
F2 Acad11 Expression (inverse rank normalized)
T|T T|C C|C
F2 Genotype (chr8:104691692)

Figure 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 shown by 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.

(C) F2 results compared with HCR/LCR findings to validate proposed regulatory mechanisms driven by allelic convergence.

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

(D) HCR/LCR peaks 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, where a positive value indicates greater peak signal in HCRs than in LCRs. Points are outlined if the peak compared contains a differential motif that is disrupted based on the presence of an HCR/LCR fixed allele.

(E) Comparison of the direction of allelic effect and the difference of gene expression between HCR/LCR lines. Genes with HCR/LCR DESeq FDR < 0.01 and variants with LD $r^2 > 0.9$ with the top eVariant were included for consideration. Points are colored by the biological relevance of the gene (blue = mitochondrial related, red = vasculature related, gray = other).

(F) Comparison of -log10(*p*-values) signed by beta (eQTL) and direction of relationship between peak and gene in the HCR/LCR analysis. The eQTL $-\log_{10}$(*p*-values) are signed by the beta value. The HCR/LCR $-\log_{10}$(*p*-values) are signed by the direction of the Pk2G link. Outlined points indicate disrupted motif presence within the peak. Points are colored by the biological relevance of the gene (blue = mitochondrial related, red = vasculature related, 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.

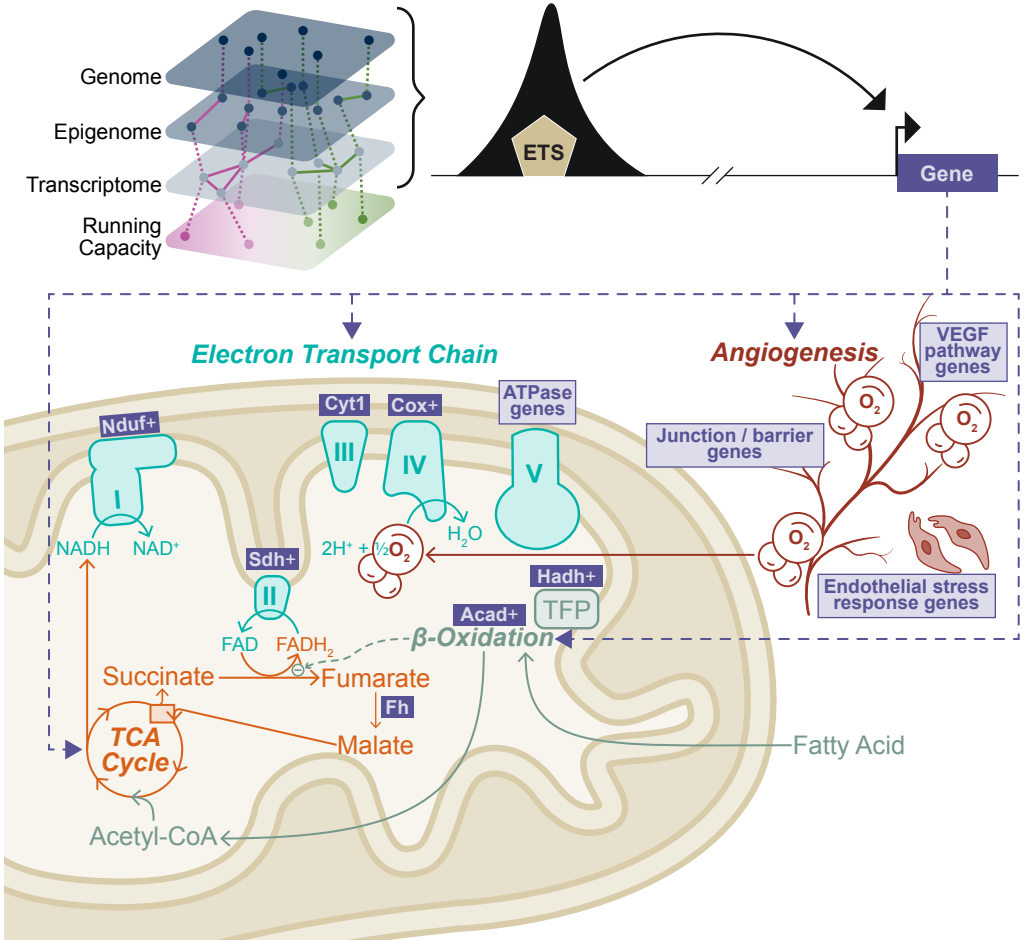
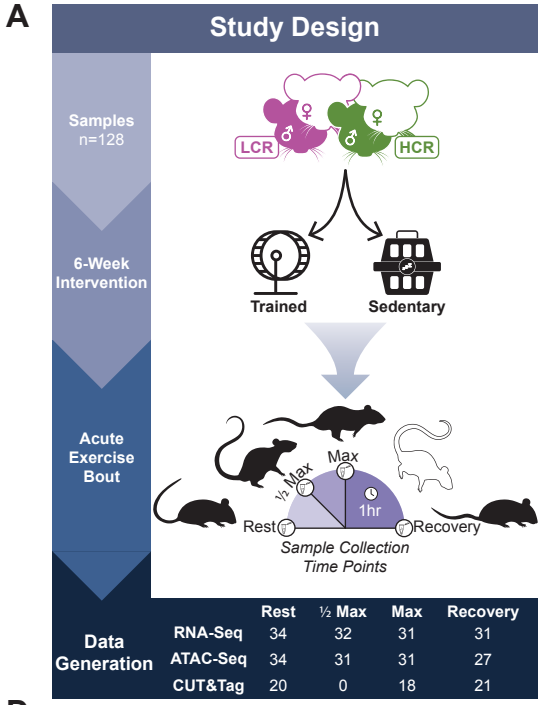
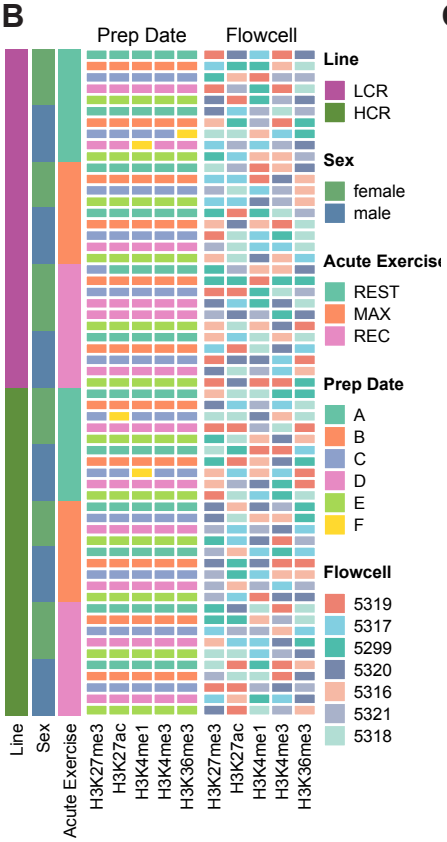
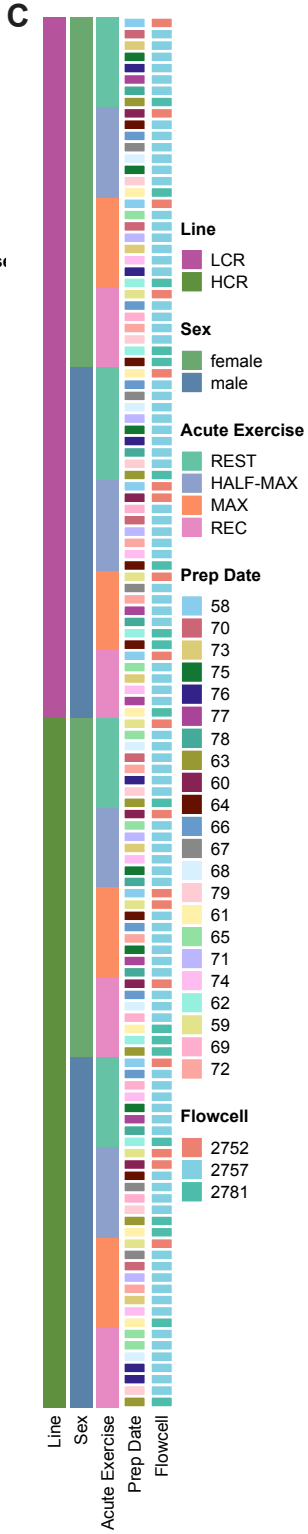

Genome
Epigenome
Transcriptome
Running Capacity
ETS
Gene
Electron Transport Chain
Angiogenesis
VEGF pathway genes
Junction / barrier genes
Endothelial stress response genes
Nduf+
Cyt1
Cox+
ATPase genes
I
III
IV
V
NADH
NAD+
2H+ + ½O2
H2O
Sdh+
II
FAD
FADH2
Hadh+
TFP
Acad+
β-Oxidation
Succinate
Fumarate
Fh
Malate
TCA Cycle
Fatty Acid
Acetyl-CoA

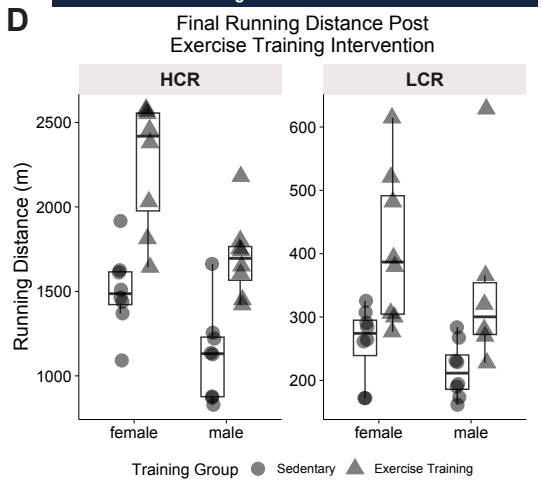
Figure 7. ETS-mediated regulation of mitochondrial energy metabolism genes supports oxidative efficiency in high CRF muscle

Cartoon representation of ETS TF motifs in enhancer regions upregulating genes involved in fatty acid oxidation (ACADs, Hadh), TCA flux (Fh), 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.


A
Study Design
Samples
n=128
6-Week
Intervention
Acute
Exercise
Bout
Data
Generation
LCR
HCR
Trained
Sedentary
½ Max
Max
1hr
Rest
Recovery
Sample Collection
Time Points
Rest ½ Max Max Recovery
RNA-Seq 34 32 31 31
ATAC-Seq 34 31 31 27
CUT&Tag 20 0 18 21


B
Prep Date
Flowcell
Line
LCR
HCR
Sex
female
male
Acute Exercise
REST
MAX
REC
Prep Date
A
B
C
D
E
F
Flowcell
5319
5317
5299
5320
5316
5321
5318
Line
Sex
Acute Exercise
H3K27me3
H3K27ac
H3K4me1
H3K4me3
H3K36me3
H3K27me3
H3K27ac
H3K4me1
H3K4me3
H3K36me3


C
Line
LCR
HCR
Sex
female
male
Acute Exercise
REST
HALF-MAX
MAX
REC
Prep Date
58
70
73
75
76
77
78
63
60
64
66
67
68
79
61
65
71
74
62
59
69
72
Flowcell
2752
2757
2781
Line
Sex
Acute Exercise
Prep Date
Flowcell


D
Final Running Distance Post
Exercise Training Intervention
HCR
LCR
Running Distance (m)
2500
2000
1500
1000
600
500
400
300
200
female
male
female
male
Training Group
Sedentary
Exercise Training

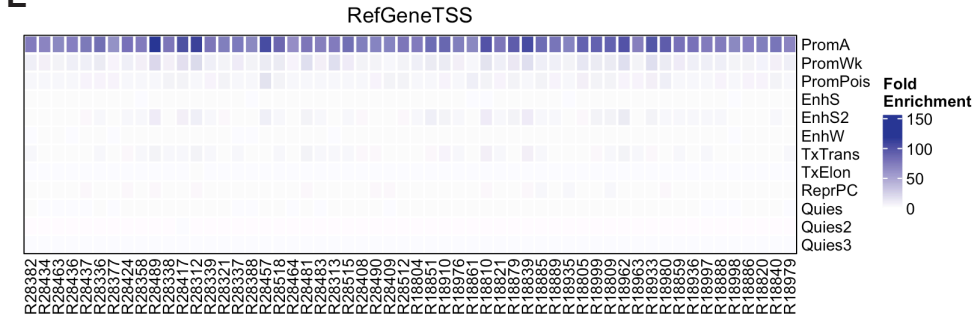

E
RefGeneTSS
PromA
PromWk
PromPois
EnhS
EnhS2
EnhW
TxTrans
TxElon
ReprPC
Quies
Quies2
Quies3
Fold
Enrichment
150
100
50
0
R28382
R28434
R28463
R28436
R28437
R28336
R28377
R28424
R28358
R28489
R28338
R28417
R28312
R28339
R28321
R28337
R28397
R28457
R28464
R28481
R28483
R28313
R28515
R28408
R28490
R28409
R28512
R18804
R18851
R18910
R18976
R18861
R18810
R18821
R18879
R18839
R18885
R18889
R18935
R18805
R18809
R18962
R18963
R18933
R18980
R18859
R18936
R18997
R18888
R18998
R18886
R18820
R18840
R18979

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

(A) Study design that included 128 HCR and LCR rats, balanced for sex. Half of the rats underwent a six-week training intervention 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. 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.

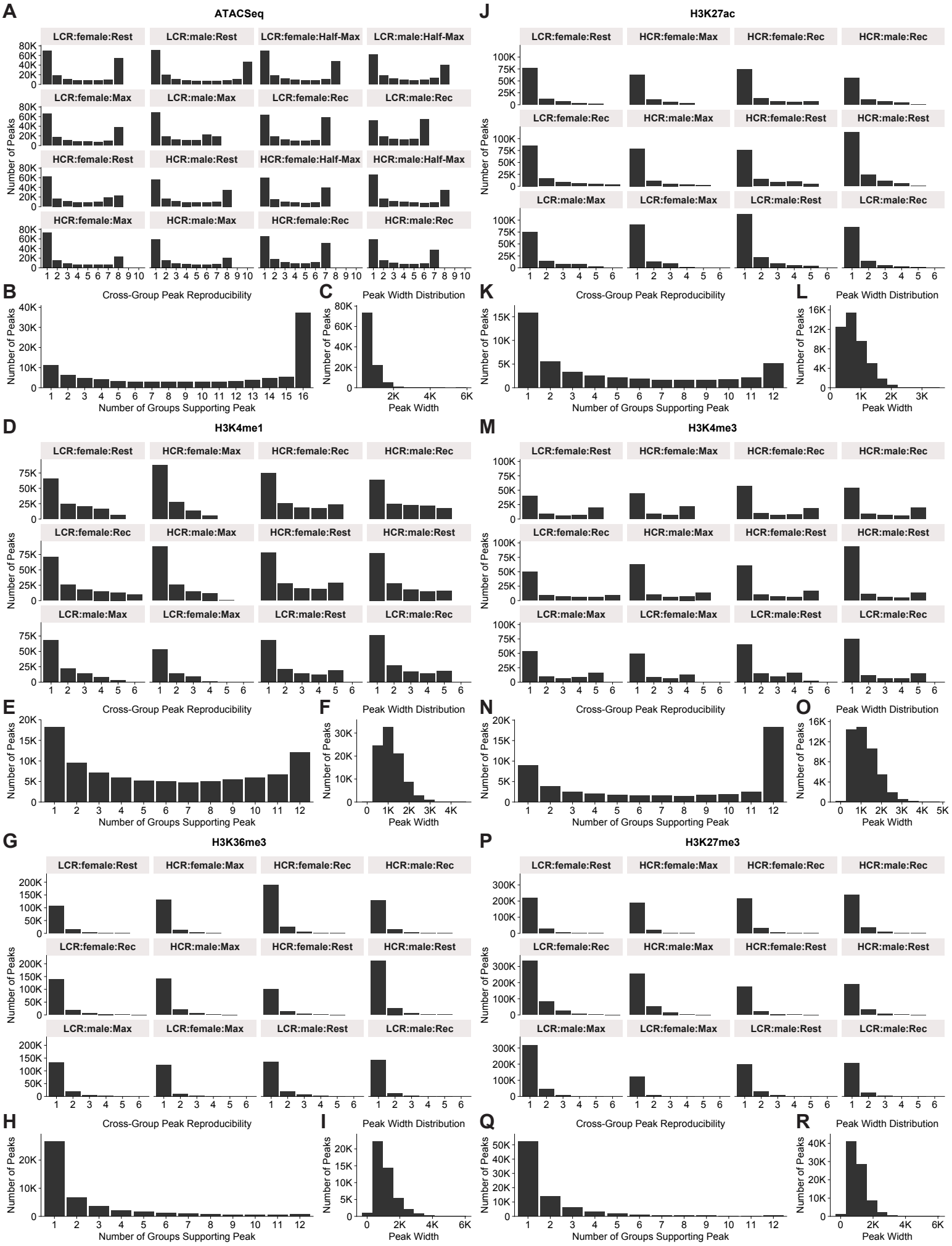

A
ATACSeq
LCR:female:Rest
LCR:male:Rest
LCR:female:Half-Max
LCR:male:Half-Max
LCR:female:Max
LCR:male:Max
LCR:female:Rec
LCR:male:Rec
HCR:female:Rest
HCR:male:Rest
HCR:female:Half-Max
HCR:male:Half-Max
HCR:female:Max
HCR:male:Max
HCR:female:Rec
HCR:male:Rec
Number of Peaks
B
Cross-Group Peak Reproducibility
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

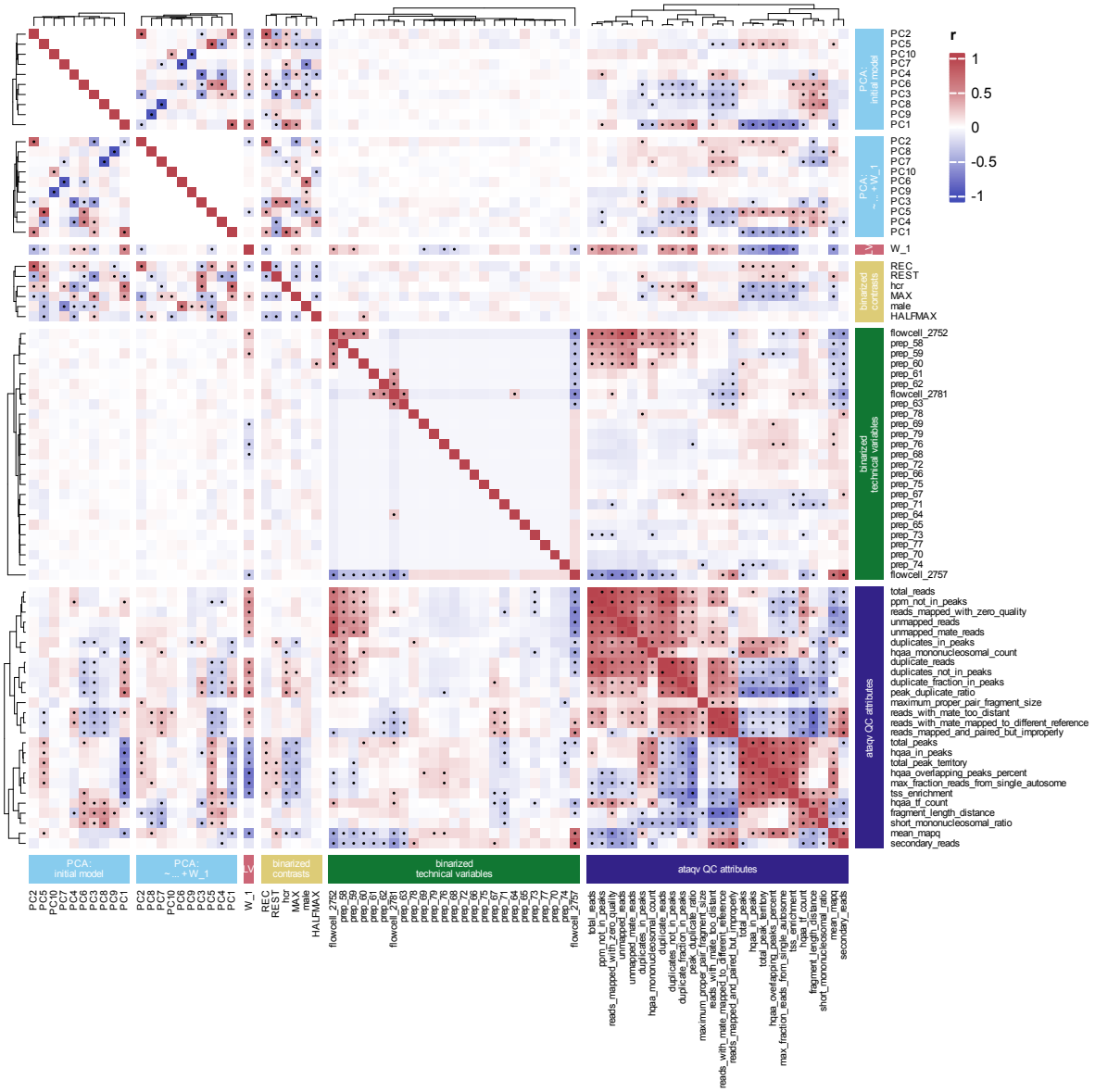
Figure S2. QC Statistics for the Creation of Master Peak Matrices, related to Figure 2

(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

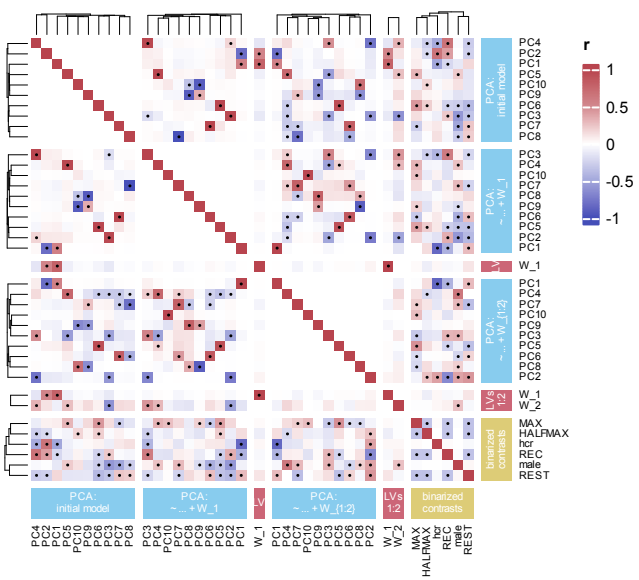
A

ATAC-Seq

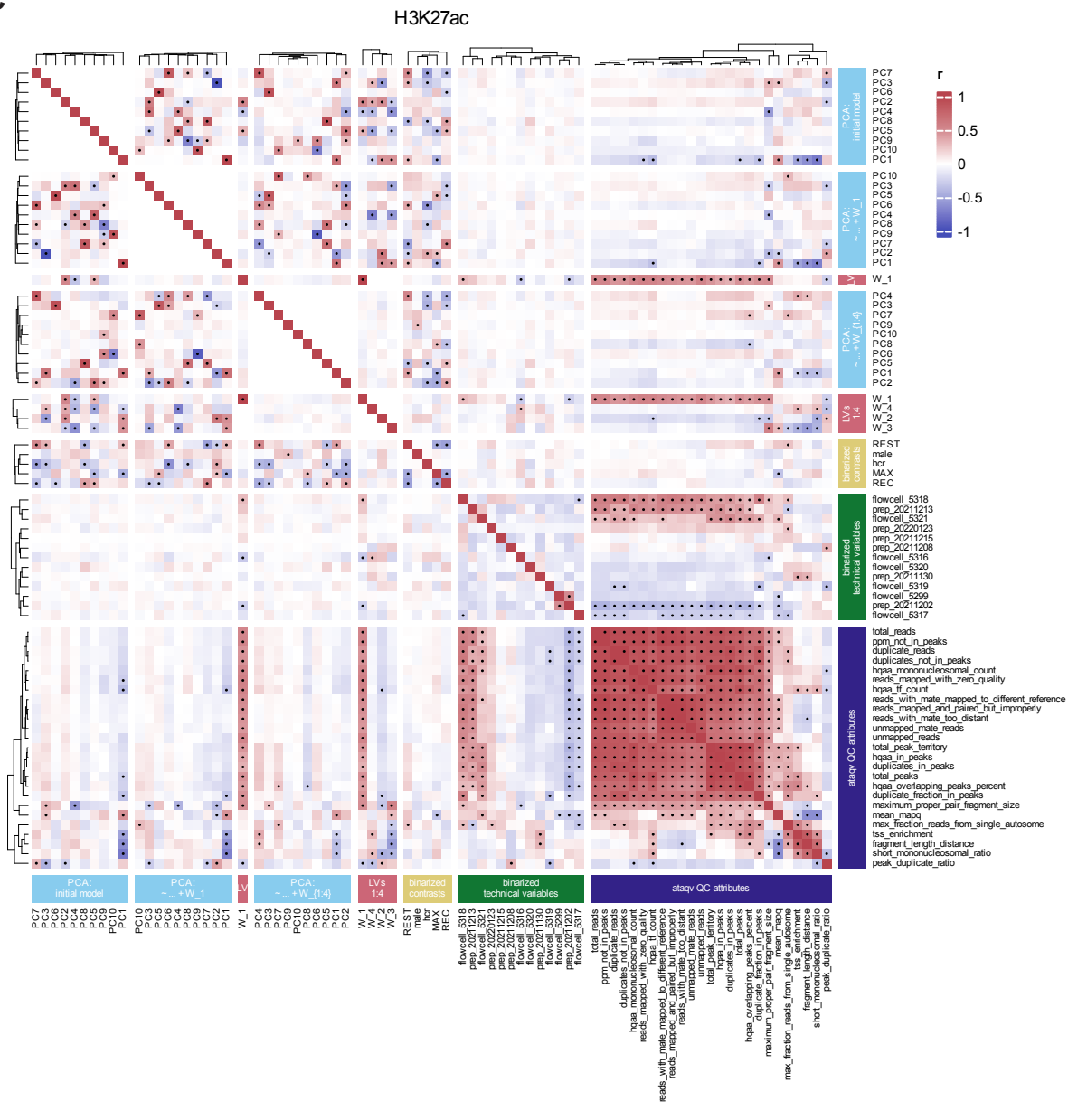


B

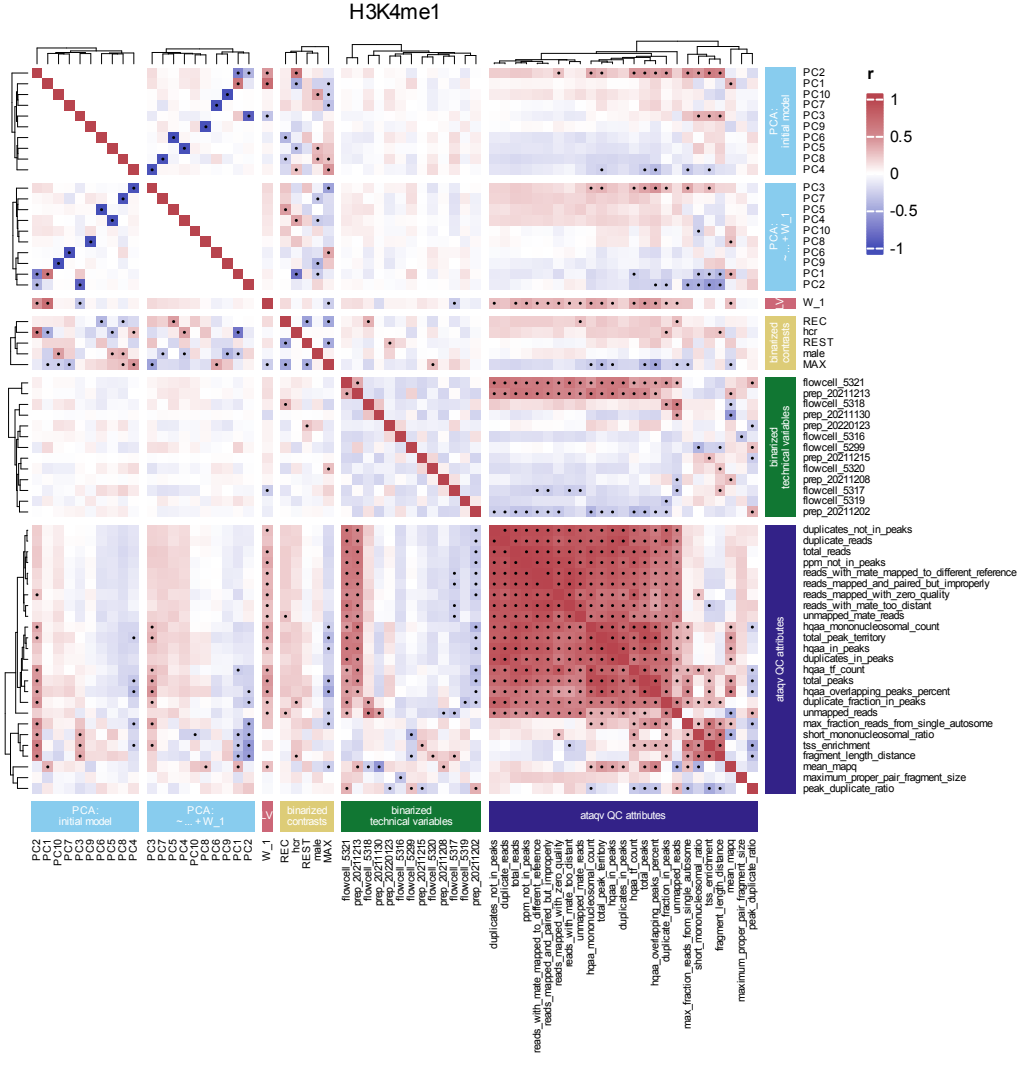
RNA-Seq



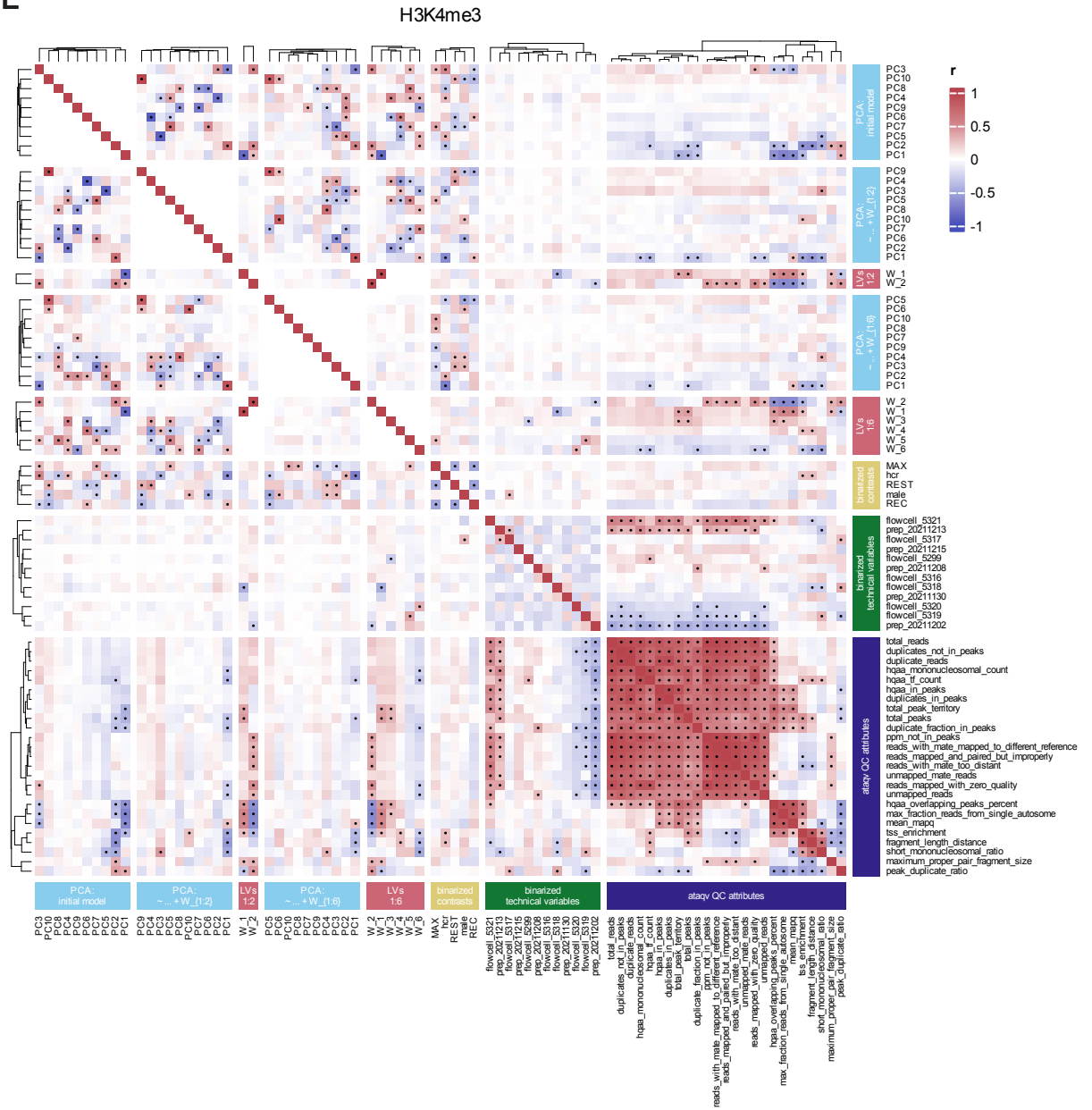
C


H3K27ac
r
1
0.5
0
-0.5
-1
PCA: initial model
PCA: ~ ... + W_1
LV
PCA: ~ ... + W_{1:4}
LVs 1:4
binarized contrasts
binarized technical variables
ataqv QC attributes
PC7
PC3
PC6
PC2
PC4
PC8
PC5
PC9
PC10
PC1
W_1
W_4
W_2
W_3
REST
male
hcr
MAX
REC
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
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

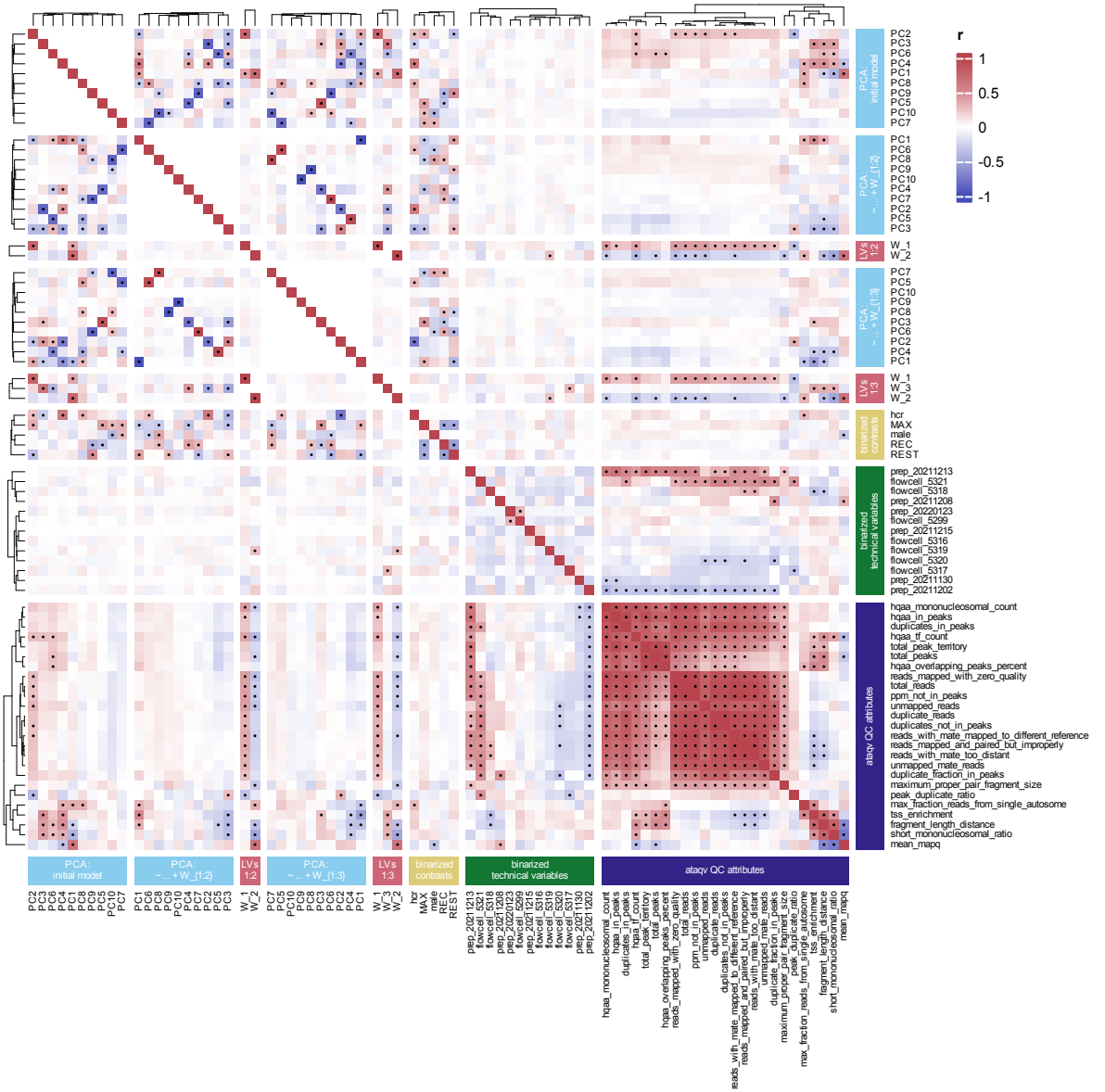

H3K4me1
r
1
0.5
0
-0.5
-1
PCA: initial model
PCA: ~ .. + W_1
W_1
binarized contrasts
binarized technical variables
ataqv QC attributes
PC2
PC1
PC10
PC7
PC3
PC9
PC6
PC5
PC8
PC4
REC
hcr
REST
male
MAX
flowcell_5321
prep_20211213
flowcell_5318
prep_20211130
prep_20220123
flowcell_5316
flowcell_5299
prep_20211215
flowcell_5320
prep_20211208
flowcell_5317
flowcell_5319
prep_20211202
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

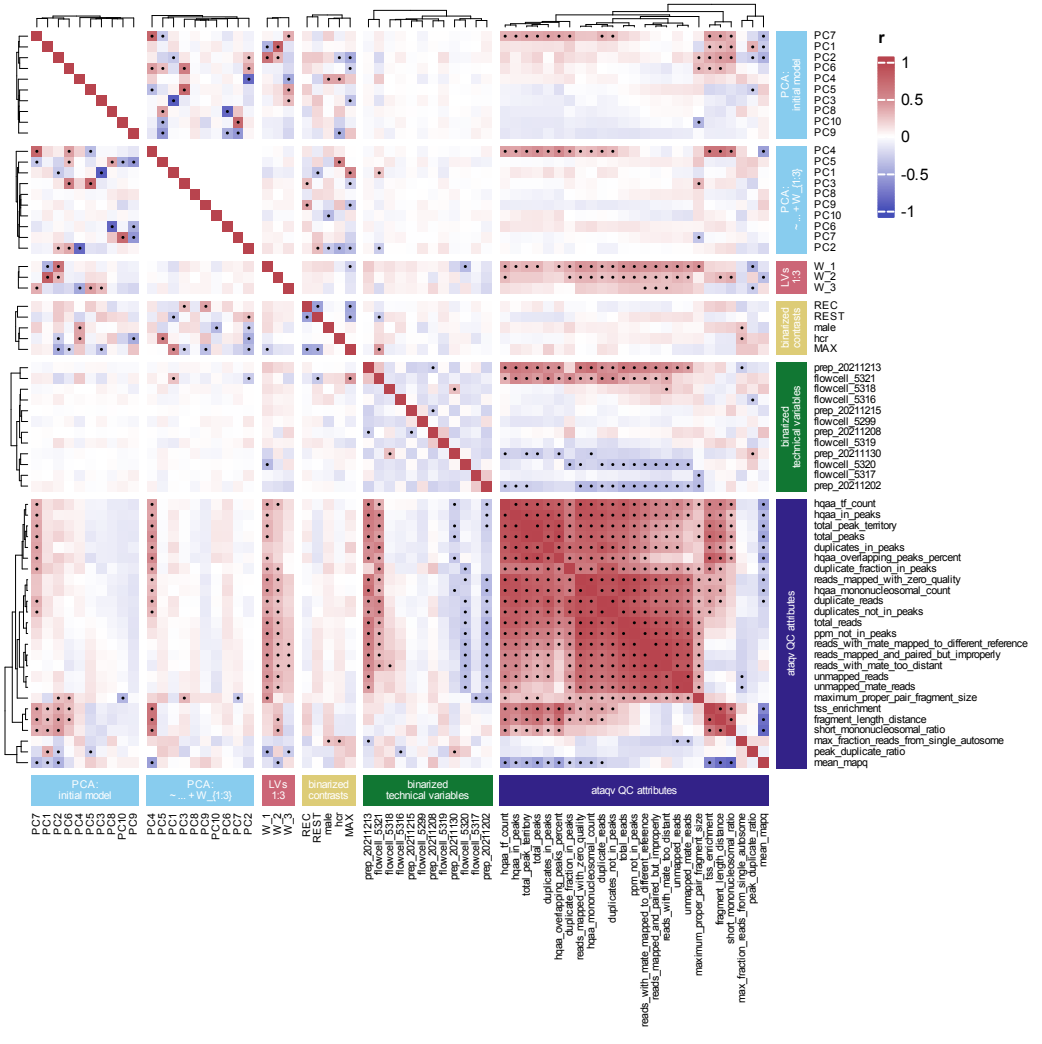

H3K4me3
r
1
0.5
0
-0.5
-1
PCA: initial model
PCA: ~ ... + W_{1:2}
LVs 1:2
PCA: ~ ... + W_{1:6}
LVs 1:6
binarized contrasts
binarized technical variables
ataqv QC attributes
PC3
PC10
PC8
PC4
PC9
PC6
PC7
PC5
PC2
PC1
W_1
W_2
W_3
W_4
W_5
W_6
MAX
hcr
REST
male
REC
flowcell_5321
prep_20211213
flowcell_5317
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

G



H3K27me3
r
1
0.5
0
-0.5
-1
PCA: initial model
PCA: ~ ... + W_{1:3}
LVs 1:3
binarized contrasts
binarized technical variables
ataqv QC attributes
PC7
PC1
PC2
PC6
PC4
PC5
PC3
PC8
PC10
PC9
PC4
PC5
PC1
PC3
PC8
PC9
PC10
PC6
PC7
PC2
W_1
W_2
W_3
REC
REST
male
hcr
MAX
prep_20211213
flowcell_5321
flowcell_5318
flowcell_5316
prep_20211215
flowcell_5299
prep_20211208
flowcell_5319
prep_20211130
flowcell_5320
flowcell_5317
prep_20211202
hqaa_tf_count
hqaa_in_peaks
total_peak_territory
total_peaks
duplicates_in_peaks
hqaa_overlapping_peaks_percent
duplicate_fraction_in_peaks
reads_mapped_with_zero_quality
hqaa_mononucleosomal_count
duplicate_reads
duplicates_not_in_peaks
total_reads
ppm_not_in_peaks
reads_with_mate_mapped_to_different_reference
reads_mapped_and_paired_but_improperly
reads_with_mate_too_distant
unmapped_reads
unmapped_mate_reads
maximum_proper_pair_fragment_size
tss_enrichment
fragment_length_distance
short_mononucleosomal_ratio
max_fraction_reads_from_single_autosome
peak_duplicate_ratio
mean_mapq

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

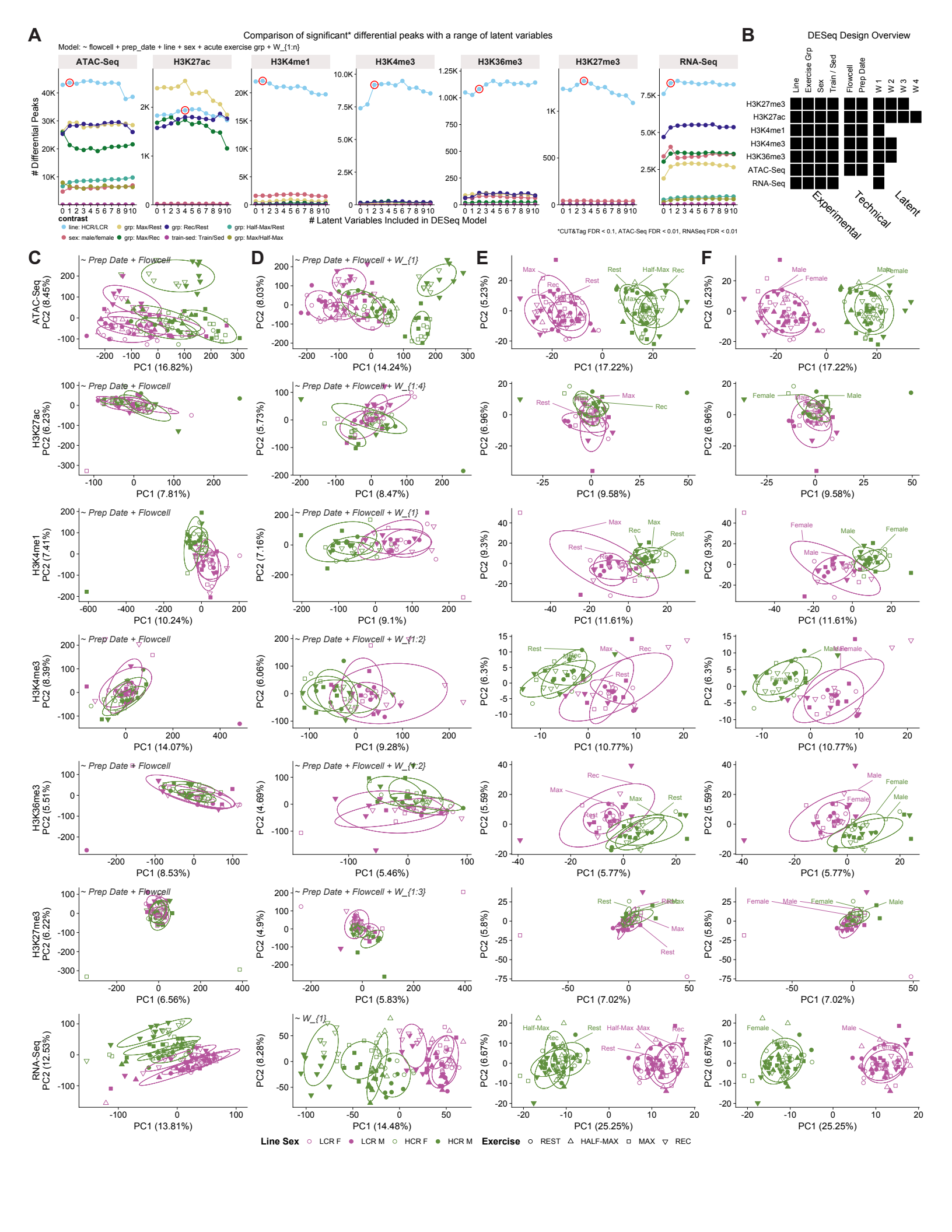

A
Comparison of significant* differential peaks with a range of latent variables
Model: ~ flowcell + prep_date + line + sex + acute exercise grp + W_{1:n}
ATAC-Seq
H3K27ac
H3K4me1
H3K4me3
H3K36me3
H3K27me3
RNA-Seq
Differential Peaks
Latent Variables Included in DESeq Model
contrast
line: HCR/LCR
sex: male/female
grp: Max/Rest
grp: Max/Rec
grp: Rec/Rest
train-sed: Train/Sed
grp: Half-Max/Rest
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
H3K27me3
H3K27ac
H3K4me1
H3K4me3
H3K36me3
ATAC-Seq
RNA-Seq
Experimental
Technical
Latent
C
D
E
F
~ Prep Date + Flowcell
~ Prep Date + Flowcell + W_{1}
~ Prep Date + Flowcell + W_{1:4}
~ Prep Date + Flowcell + W_{1:2}
~ Prep Date + Flowcell + W_{1:3}
~ W_{1}
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%)
Max
Rest
Rec
Half-Max
Male
Female
Line Sex
LCR F
LCR M
HCR F
HCR M
Exercise
REST
HALF-MAX
MAX
REC

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

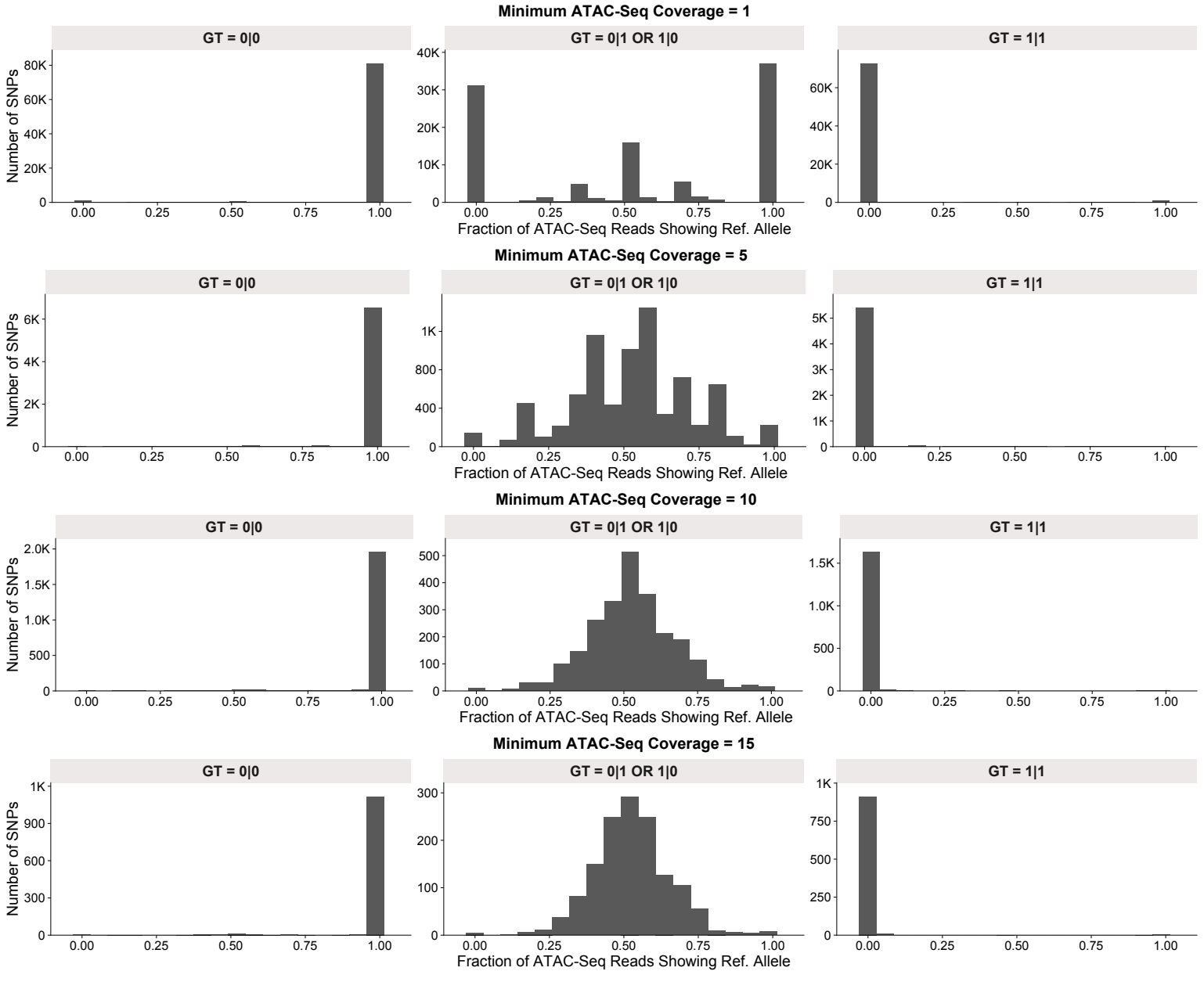

Minimum ATAC-Seq Coverage = 1
GT = 0|0
GT = 0|1 OR 1|0
GT = 1|1
Number of SNPs
Fraction of ATAC-Seq Reads Showing Ref. Allele
Minimum ATAC-Seq Coverage = 5
GT = 0|0
GT = 0|1 OR 1|0
GT = 1|1
Number of SNPs
Fraction of ATAC-Seq Reads Showing Ref. Allele
Minimum ATAC-Seq Coverage = 10
GT = 0|0
GT = 0|1 OR 1|0
GT = 1|1
Number of SNPs
Fraction of ATAC-Seq Reads Showing Ref. Allele
Minimum ATAC-Seq Coverage = 15
GT = 0|0
GT = 0|1 OR 1|0
GT = 1|1
Number of SNPs
Fraction of ATAC-Seq Reads Showing Ref. Allele

Figure S5. Genotype Imputation QC

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.

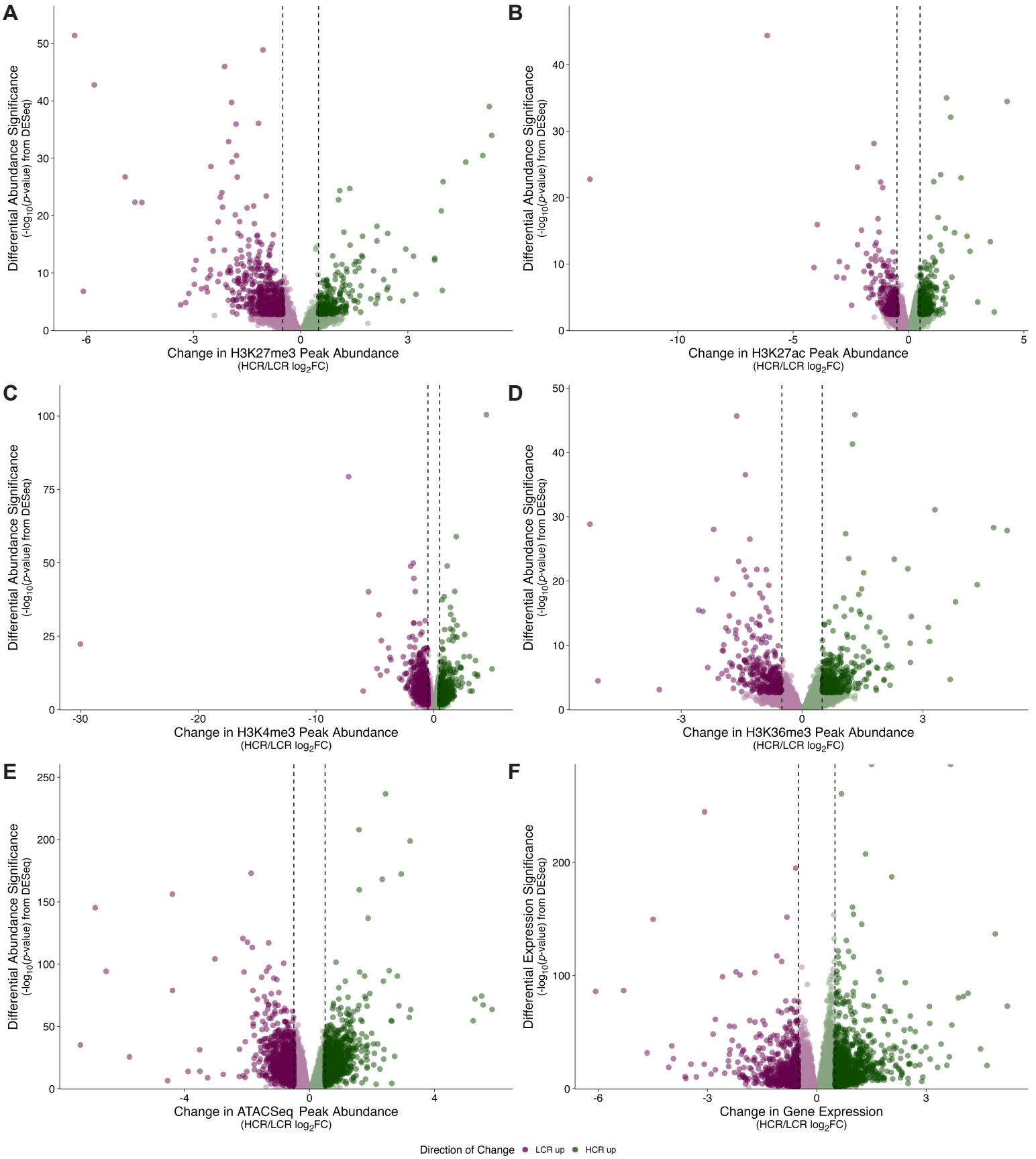

A
Differential Abundance Significance
(-log10(p-value) from DESeq)
Change in H3K27me3 Peak Abundance
(HCR/LCR log2FC)
B
Differential Abundance Significance
(-log10(p-value) from DESeq)
Change in H3K27ac Peak Abundance
(HCR/LCR log2FC)
C
Differential Abundance Significance
(-log10(p-value) from DESeq)
Change in H3K4me3 Peak Abundance
(HCR/LCR log2FC)
D
Differential Abundance Significance
(-log10(p-value) from DESeq)
Change in H3K36me3 Peak Abundance
(HCR/LCR log2FC)
E
Differential Abundance Significance
(-log10(p-value) from DESeq)
Change in ATACSeq Peak Abundance
(HCR/LCR log2FC)
F
Differential Expression Significance
(-log10(p-value) from DESeq)
Change in Gene Expression
(HCR/LCR log2FC)
Direction of Change
LCR up
HCR up

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

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

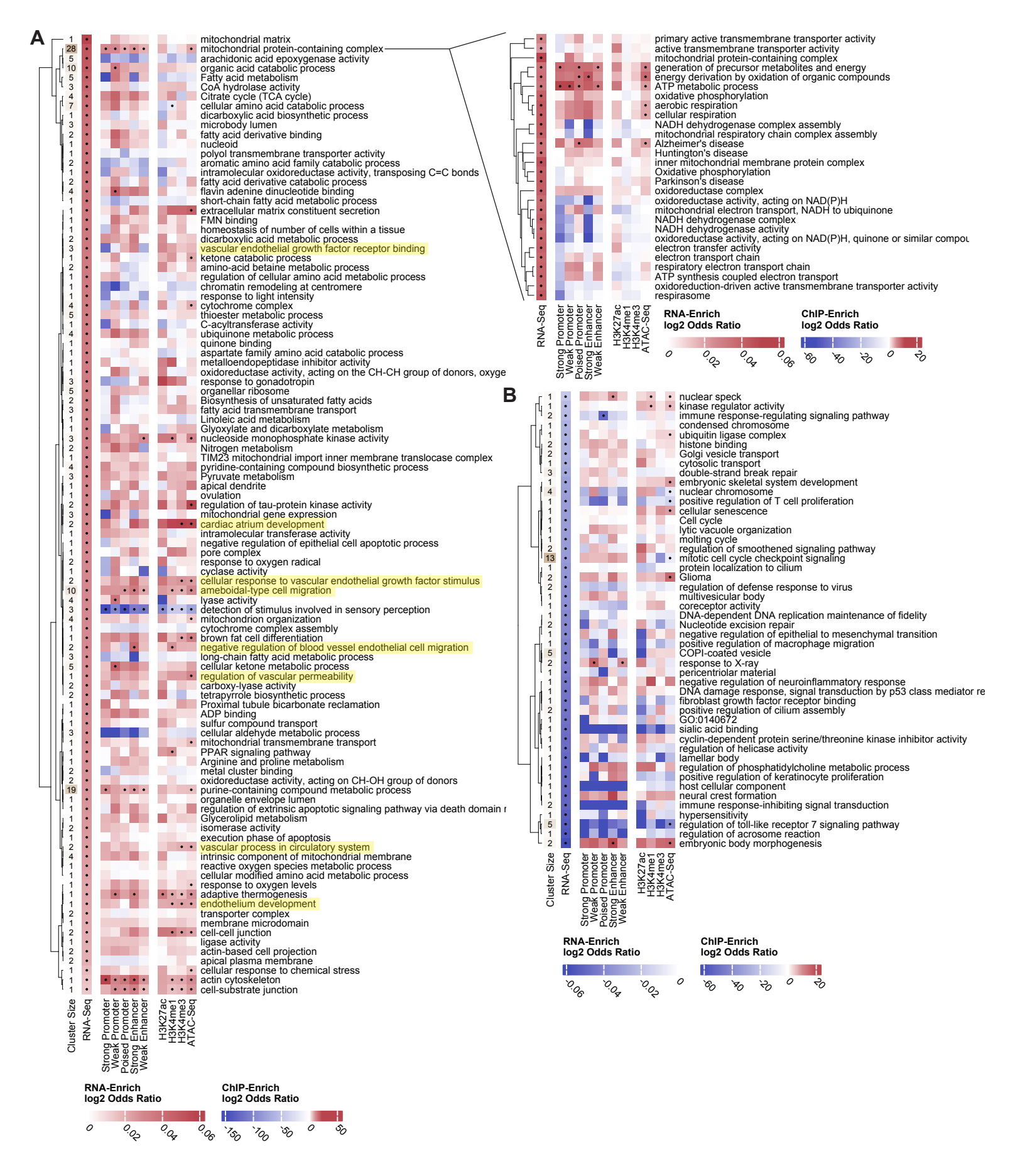

A
mitochondrial matrix
mitochondrial protein-containing complex
arachidonic acid epoxygenase activity
organic acid catabolic process
Fatty acid metabolism
CoA hydrolase activity
Citrate cycle (TCA cycle)
cellular amino acid catabolic process
dicarboxylic acid biosynthetic process
microbody lumen
fatty acid derivative binding
nucleoid
polyol transmembrane transporter activity
aromatic amino acid family catabolic process
intramolecular oxidoreductase activity, transposing C=C bonds
fatty acid derivative catabolic process
flavin adenine dinucleotide binding
short-chain fatty acid metabolic process
extracellular matrix constituent secretion
FMN binding
homeostasis of number of cells within a tissue
dicarboxylic acid metabolic process
vascular endothelial growth factor receptor binding
ketone catabolic process
amino-acid betaine metabolic process
regulation of cellular amino acid metabolic process
chromatin remodeling at centromere
response to light intensity
cytochrome complex
thioester metabolic process
C-acyltransferase activity
ubiquinone metabolic process
quinone binding
aspartate family amino acid catabolic process
metalloendopeptidase inhibitor activity
oxidoreductase activity, acting on the CH-CH group of donors, oxyge
response to gonadotropin
organellar ribosome
Biosynthesis of unsaturated fatty acids
fatty acid transmembrane transport
Linoleic acid metabolism
Glyoxylate and dicarboxylate metabolism
nucleoside monophosphate kinase activity
Nitrogen metabolism
TIM23 mitochondrial import inner membrane translocase complex
pyridine-containing compound biosynthetic process
Pyruvate metabolism
apical dendrite
ovulation
regulation of tau-protein kinase activity
mitochondrial gene expression
cardiac atrium development
intramolecular transferase activity
negative regulation of epithelial cell apoptotic process
pore complex
response to oxygen radical
cyclase activity
cellular response to vascular endothelial growth factor stimulus
ameboidal-type cell migration
lyase activity
detection of stimulus involved in sensory perception
mitochondrion organization
cytochrome complex assembly
brown fat cell differentiation
negative regulation of blood vessel endothelial cell migration
long-chain fatty acid metabolic process
cellular ketone metabolic process
regulation of vascular permeability
carboxy-lyase activity
tetrapyrrole biosynthetic process
Proximal tubule bicarbonate reclamation
ADP binding
sulfur compound transport
cellular aldehyde metabolic process
mitochondrial transmembrane transport
PPAR signaling pathway
Arginine and proline metabolism
metal cluster binding
oxidoreductase activity, acting on CH-OH group of donors
purine-containing compound metabolic process
organelle envelope lumen
regulation of extrinsic apoptotic signaling pathway via death domain r
Glycerolipid metabolism
isomerase activity
execution phase of apoptosis
vascular process in circulatory system
intrinsic component of mitochondrial membrane
reactive oxygen species metabolic process
cellular modified amino acid metabolic process
response to oxygen levels
adaptive thermogenesis
endothelium development
transporter complex
membrane microdomain
cell-cell junction
ligase activity
actin-based cell projection
apical plasma membrane
cellular response to chemical stress
actin cytoskeleton
cell-substrate junction
Cluster Size
RNA-Seq
Strong Promoter
Weak Promoter
Poised Promoter
Strong Enhancer
Weak Enhancer
H3K27ac
H3K4me1
H3K4me3
ATAC-Seq
RNA-Enrich
log2 Odds Ratio
0
0.02
0.04
0.06
ChIP-Enrich
log2 Odds Ratio
-150
-100
-50
0
50
primary active transmembrane transporter activity
active transmembrane transporter activity
mitochondrial protein-containing complex
generation of precursor metabolites and energy
energy derivation by oxidation of organic compounds
ATP metabolic process
oxidative phosphorylation
aerobic respiration
cellular respiration
NADH dehydrogenase complex assembly
mitochondrial respiratory chain complex assembly
Alzheimer's disease
Huntington's disease
inner mitochondrial membrane protein complex
Oxidative phosphorylation
Parkinson's disease
oxidoreductase complex
oxidoreductase activity, acting on NAD(P)H
mitochondrial electron transport, NADH to ubiquinone
NADH dehydrogenase complex
NADH dehydrogenase activity
oxidoreductase activity, acting on NAD(P)H, quinone or similar compou
electron transfer activity
electron transport chain
respiratory electron transport chain
ATP synthesis coupled electron transport
oxidoreduction-driven active transmembrane transporter activity
respirasome
RNA-Seq
Strong Promoter
Weak Promoter
Poised Promoter
Strong Enhancer
Weak Enhancer
H3K27ac
H3K4me1
H3K4me3
ATAC-Seq
RNA-Enrich
log2 Odds Ratio
0
0.02
0.04
0.06
ChIP-Enrich
log2 Odds Ratio
-60
-40
-20
0
20
B
nuclear speck
kinase regulator activity
immune response-regulating signaling pathway
condensed chromosome
ubiquitin ligase complex
histone binding
Golgi vesicle transport
cytosolic transport
double-strand break repair
embryonic skeletal system development
nuclear chromosome
positive regulation of T cell proliferation
cellular senescence
Cell cycle
lytic vacuole organization
molting cycle
regulation of smoothened signaling pathway
mitotic cell cycle checkpoint signaling
protein localization to cilium
Glioma
regulation of defense response to virus
multivesicular body
coreceptor activity
DNA-dependent DNA replication maintenance of fidelity
Nucleotide excision repair
negative regulation of epithelial to mesenchymal transition
positive regulation of macrophage migration
COPI-coated vesicle
response to X-ray
pericentriolar material
negative regulation of neuroinflammatory response
DNA damage response, signal transduction by p53 class mediator re
fibroblast growth factor receptor binding
positive regulation of cilium assembly
GO:0140672
sialic acid binding
cyclin-dependent protein serine/threonine kinase inhibitor activity
regulation of helicase activity
lamellar body
regulation of phosphatidylcholine metabolic process
positive regulation of keratinocyte proliferation
host cellular component
neural crest formation
immune response-inhibiting signal transduction
hypersensitivity
regulation of toll-like receptor 7 signaling pathway
regulation of acrosome reaction
embryonic body morphogenesis
Cluster Size
RNA-Seq
Strong Promoter
Weak Promoter
Poised Promoter
Strong Enhancer
Weak Enhancer
H3K27ac
H3K4me1
H3K4me3
ATAC-Seq
RNA-Enrich
log2 Odds Ratio
-0.06
-0.04
-0.02
0
ChIP-Enrich
log2 Odds Ratio
-60
-40
-20
0
20

Figure S7. Differential RNA-Seq Pathway Enrichment Results, related to Figure 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. These results are compared to additional epigenomic modalities, in which '•' annotates a significant gene set cluster, where at least one of the gene sets within the cluster returned FDR < 0.05. RNA-Enrich analysis provides the direction of enrichment (e.g., HCR up and LCR up), whereas ChIP-Enrich returns gene sets as enriched or depleted. The largest cluster in panel A that returns significant pathways from the epigenomic data, mitochondrial protein-containing complex, is expanded upon to the right to show that nearest-TSS annotated epigenomic signals are not enriched for the majority of the child terms within this cluster. The sparsity of significantly enriched epigenomic gene sets in comparison to RNA-Seq is apparent.

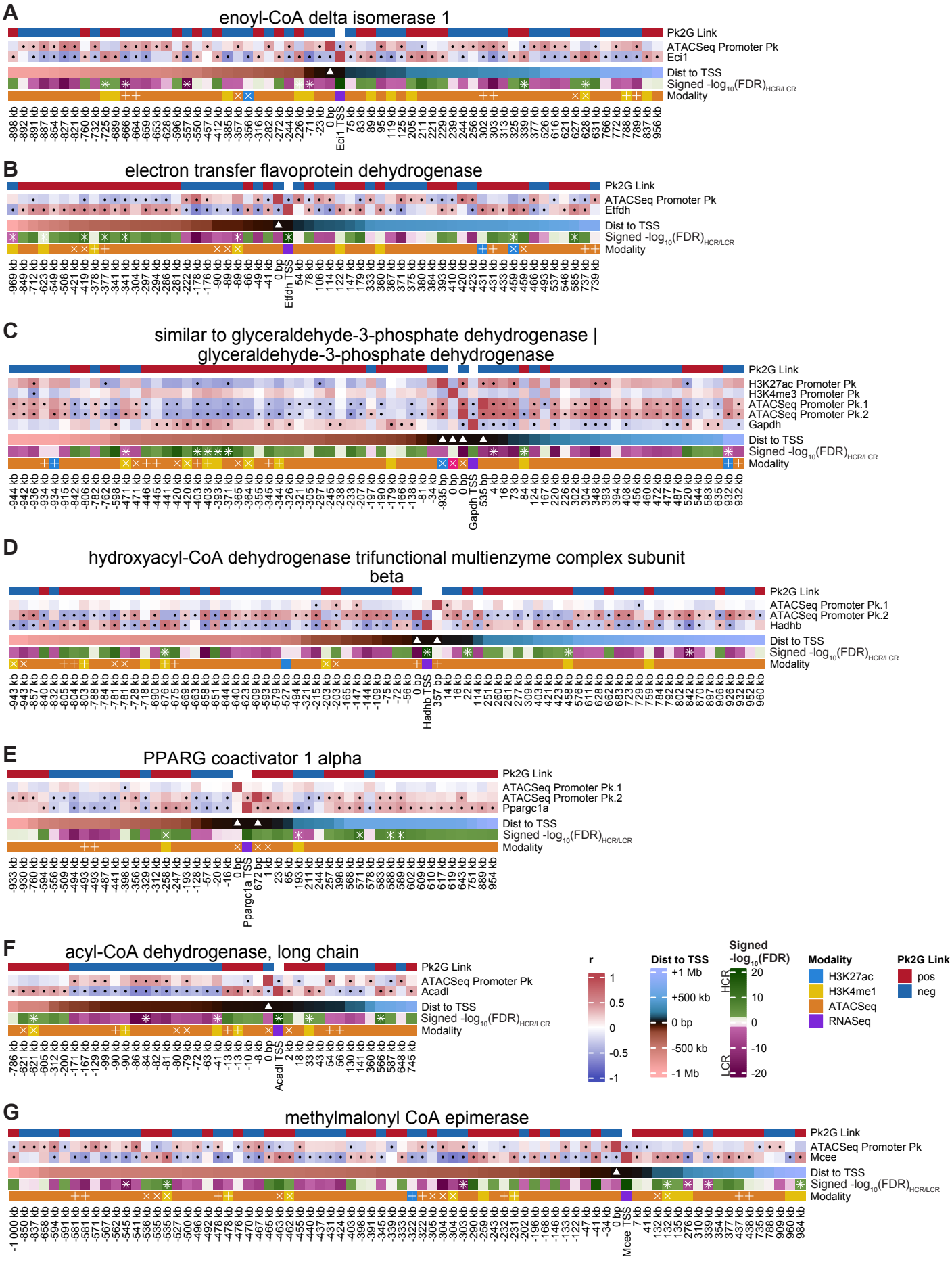

A
enoyl-CoA delta isomerase 1
Pk2G Link
ATACSeq Promoter Pk
Eci1
Dist to TSS
Signed -log10(FDR)HCR/LCR
Modality
B
electron transfer flavoprotein dehydrogenase
Pk2G Link
ATACSeq Promoter Pk
Etfdh
Dist to TSS
Signed -log10(FDR)HCR/LCR
Modality
C
similar to glyceraldehyde-3-phosphate dehydrogenase | glyceraldehyde-3-phosphate dehydrogenase
Pk2G Link
H3K27ac Promoter Pk
H3K4me3 Promoter Pk
ATACSeq Promoter Pk.1
ATACSeq Promoter Pk.2
Gapdh
Dist to TSS
Signed -log10(FDR)HCR/LCR
Modality
D
hydroxyacyl-CoA dehydrogenase trifunctional multienzyme complex subunit beta
Pk2G Link
ATACSeq Promoter Pk.1
ATACSeq Promoter Pk.2
Hadhb
Dist to TSS
Signed -log10(FDR)HCR/LCR
Modality
E
PPARG coactivator 1 alpha
Pk2G Link
ATACSeq Promoter Pk.1
ATACSeq Promoter Pk.2
Ppargc1a
Dist to TSS
Signed -log10(FDR)HCR/LCR
Modality
F
acyl-CoA dehydrogenase, long chain
Pk2G Link
ATACSeq Promoter Pk
Acadl
Dist to TSS
Signed -log10(FDR)HCR/LCR
Modality
r
Dist to TSS
Signed -log10(FDR)
HCR
LCR
Modality
H3K27ac
H3K4me1
ATACSeq
RNASeq
Pk2G Link
pos
neg
G
methylmalonyl CoA epimerase
Pk2G Link
ATACSeq Promoter Pk
Mcee
Dist to TSS
Signed -log10(FDR)HCR/LCR
Modality

Figure S8. Additional Pk2G Examples, related to Figure 3

Select examples of Pk2G relationships for target genes (A) Eci1, (B) Etfdh, (C) Gapdh, (D) Hadhb, (E) Ppargc1a, (F) Acadl, (G) and Mcee. Distal regions (present in columns only) are included if they have a significant correlation (FDR < 0.05) with either the gene of interest or peaks located in the gene's promoter region. Columns are ordered by peak distance to the gene's TSS. Cells displaying significant peak-to-peak and peak-to-gene correlations (FDR < 0.05) are annotated with a dot. 'Dist to TSS' annotation displays the distance of regions upstream or downstream of the TSS. Major color breaks occur at 20 kb, 100 kb, 200 kb, and 500 kb, with darker shades representing regions closer to the TSS. A white triangle is annotated in the 'Dist to TSS' bar if the region corresponds to a 'promoter' peak, which would also be included as a row. The 'Modality' annotation bar is color-coded to represent different modalities. Matching symbols within this bar indicate peaks residing within the same merged region (e.g., overlapping peaks from different modalities). The 'Log10 FDR' annotation bar is colored by the direction of change, highlighting regions with very low FDRs in darker shades, even if they do not meet the $\log_2$FC significance criteria. An asterisk is shown in this bar if the region is considered significant according to both FDR and $\log_2$FC thresholds in the differential analysis. The 'Pk2G Link' annotation bar shows the direction of relationship between the peak and the gene displayed in the heatmap. The 'Max F_{ST}' annotation bar shows the largest fixation index calculated for a variant residing in that peak.

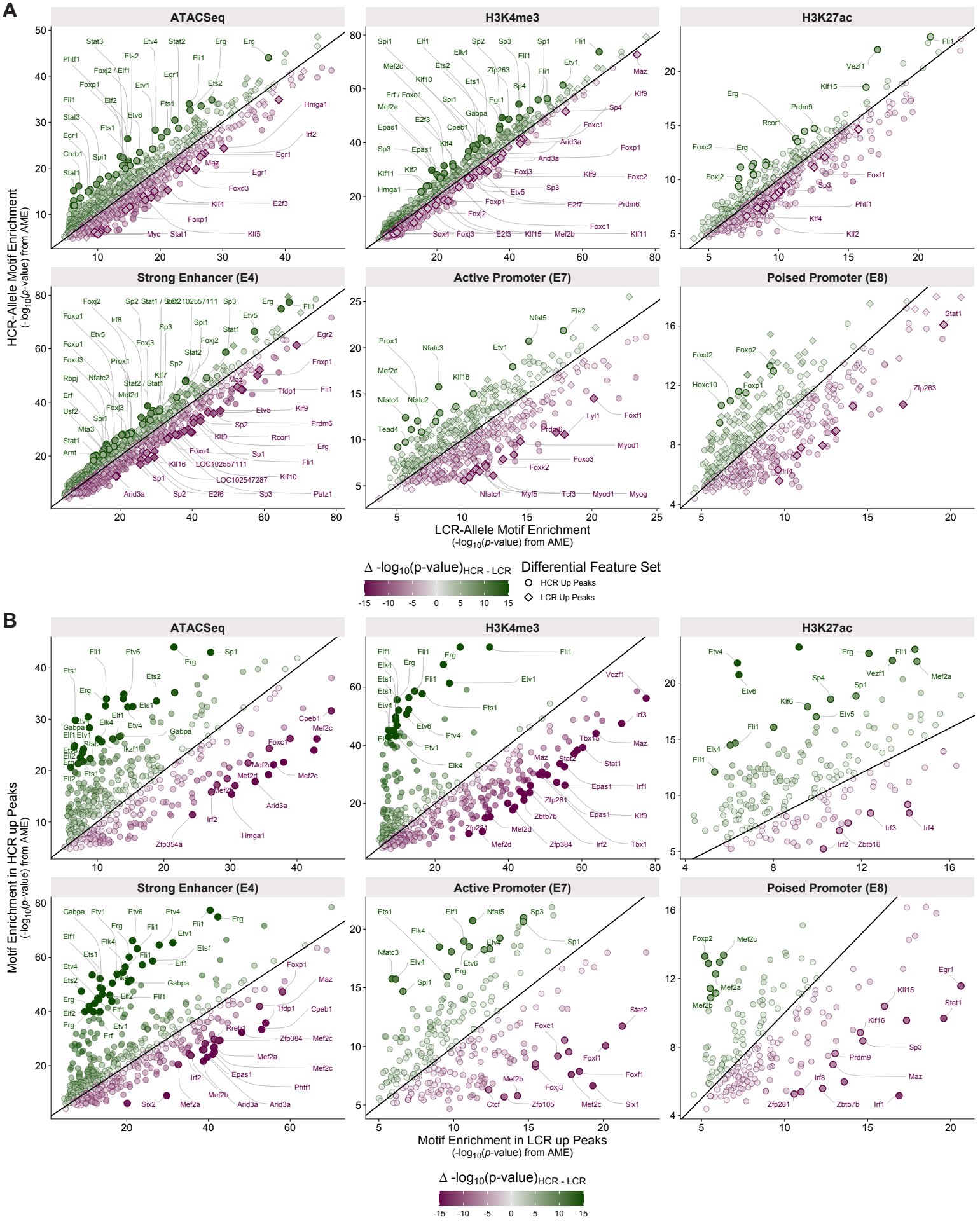
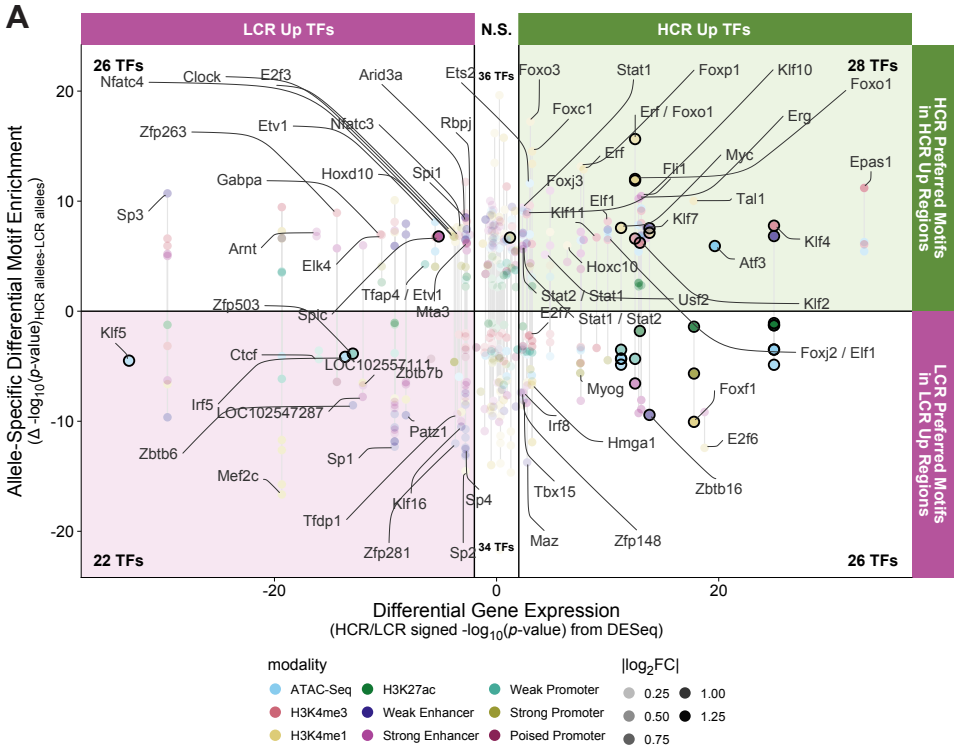

A
ATACSeq
H3K4me3
H3K27ac
Strong Enhancer (E4)
Active Promoter (E7)
Poised Promoter (E8)
HCR-Allele Motif Enrichment
(-log10(p-value) from AME)
LCR-Allele Motif Enrichment
(-log10(p-value) from AME)
Δ -log10(p-value)HCR - LCR
-15 -10 -5 0 5 10 15
Differential Feature Set
HCR Up Peaks
LCR Up Peaks
B
ATACSeq
H3K4me3
H3K27ac
Strong Enhancer (E4)
Active Promoter (E7)
Poised Promoter (E8)
Motif Enrichment in HCR up Peaks
(-log10(p-value) from AME)
Motif Enrichment in LCR up Peaks
(-log10(p-value) from AME)
Δ -log10(p-value)HCR - LCR
-15 -10 -5 0 5 10 15

Figure S9. Differential Motif Enrichment, related to Figure 4 and Figure 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.


A
LCR Up TFs
N.S.
HCR Up TFs
26 TFs
28 TFs
22 TFs
26 TFs
36 TFs
34 TFs
HCR Preferred Motifs in HCR Up Regions
LCR Preferred Motifs in LCR Up Regions
Allele-Specific Differential Motif Enrichment
(Δ -log10(p-value)HCR alleles-LCR alleles)
Differential Gene Expression
(HCR/LCR signed -log10(p-value) from DESeq)
Nfatc4
Clock
E2f3
Arid3a
Ets2
Foxo3
Stat1
Foxp1
Klf10
Foxo1
Foxc1
Erf / Foxo1
Erg
Zfp263
Etv1
Nfatc3
Rbpj
Erf
Myc
Epas1
Gabpa
Hoxd10
Spi1
Foxj3
Fli1
Tal1
Sp3
Elf1
Klf11
Klf7
Klf4
Arnt
Elk4
Hoxc10
Atf3
Tfap4 / Etv1
Stat2 / Stat1
Usf2
Klf2
Zfp503
Splc
Mta3
E2f7
Stat1 / Stat2
Klf5
Ctcf
LOC102557111
Zbtb7b
Foxj2 / Elf1
Myog
Foxf1
Irf5
LOC102547287
Irf8
Patz1
Hmga1
E2f6
Zbtb6
Sp1
Mef2c
Tbx15
Zbtb16
Klf16
Sp4
Tfdp1
Maz
Zfp148
Zfp281
Sp2
Differential Gene Expression
modality
ATAC-Seq
H3K4me3
H3K4me1
H3K27ac
Weak Enhancer
Strong Enhancer
Weak Promoter
Strong Promoter
Poised Promoter
|log2FC|
0.25
0.50
0.75
1.00
1.25

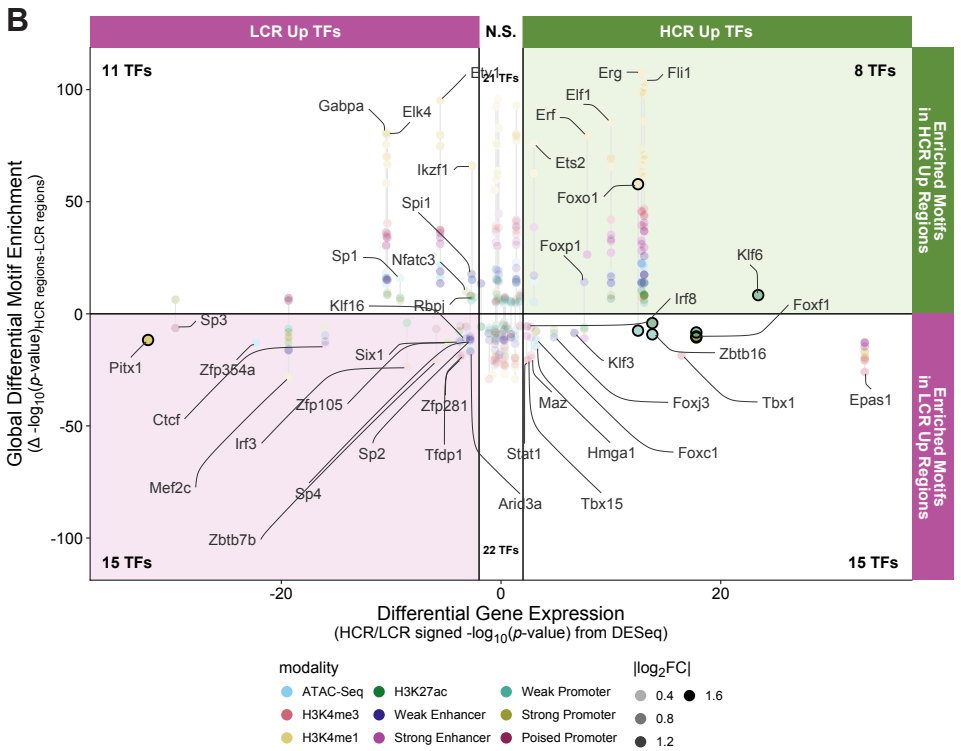

B
LCR Up TFs
N.S.
HCR Up TFs
11 TFs
8 TFs
15 TFs
15 TFs
21 TFs
22 TFs
Enriched Motifs in HCR Up Regions
Enriched Motifs in LCR Up Regions
Global Differential Motif Enrichment
(Δ -log10(p-value)HCR regions-LCR regions)
Differential Gene Expression
(HCR/LCR signed -log10(p-value) from DESeq)
Etv1
Erg
Fli1
Elf1
Gabpa
Elk4
Erf
Ets2
Ikzf1
Foxo1
Spi1
Foxp1
Klf6
Sp1
Nfatc3
Irf8
Foxf1
Klf16
Rbpj
Sp3
Pitx1
Zfp354a
Six1
Klf3
Zbtb16
Zfp105
Zfp281
Maz
Foxj3
Tbx1
Epas1
Ctcf
Irf3
Sp2
Tfdp1
Stat1
Hmga1
Foxc1
Mef2c
Sp4
Arid3a
Tbx15
Zbtb7b
modality
ATAC-Seq
H3K4me3
H3K4me1
H3K27ac
Weak Enhancer
Strong Enhancer
Weak Promoter
Strong Promoter
Poised Promoter
|log2FC|
0.4
0.8
1.2
1.6

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

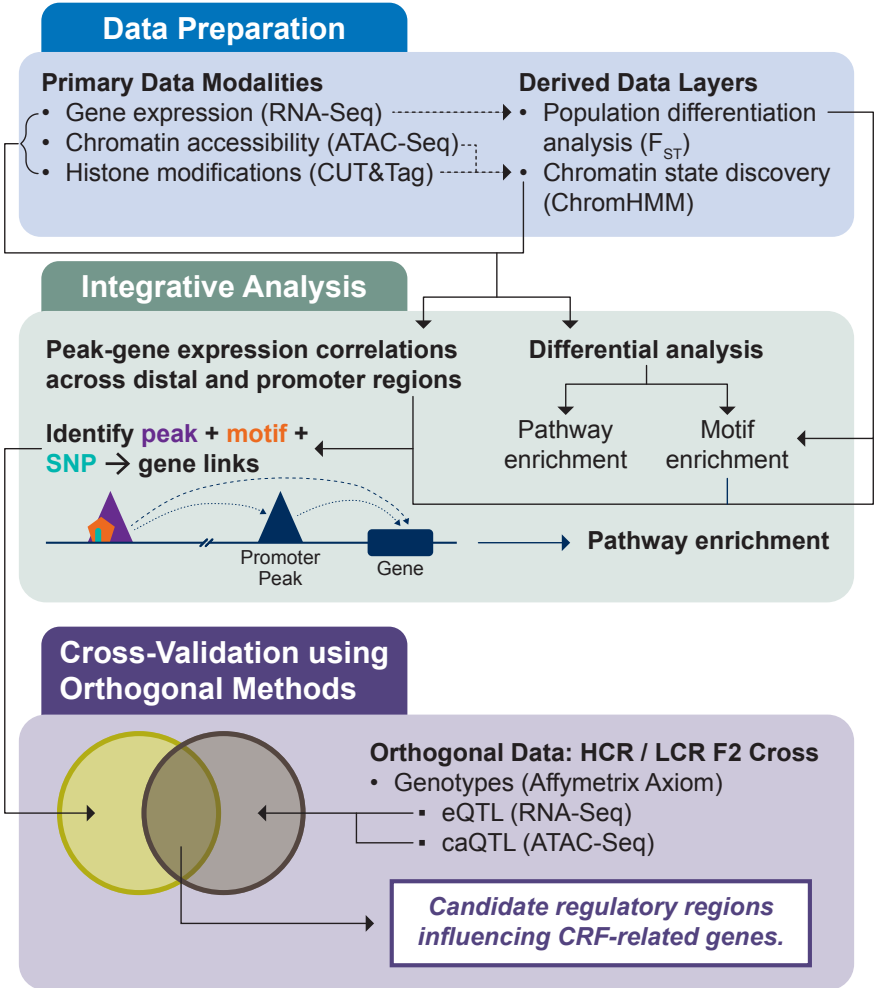

Data Preparation
Primary Data Modalities
Gene expression (RNA-Seq)
Chromatin accessibility (ATAC-Seq)
Histone modifications (CUT&Tag)
Derived Data Layers
Population differentiation analysis (F_{ST})
Chromatin state discovery (ChromHMM)
Integrative Analysis
Peak-gene expression correlations across distal and promoter regions
Differential analysis
Pathway enrichment
Motif enrichment
Identify peak + motif + SNP → gene links
Promoter Peak
Gene
Pathway enrichment
Cross-Validation using Orthogonal Methods
Orthogonal Data: HCR / LCR F2 Cross
Genotypes (Affymetrix Axiom)
eQTL (RNA-Seq)
caQTL (ATAC-Seq)
Candidate regulatory regions influencing CRF-related genes.

Figure 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.