**Multiomic identification of key transcriptional regulatory programs during endurance exercise training**

Gregory R. Smith[1,9*], Bingqing Zhao[2,9*], Malene E. Lindholm[3], Archana Raja[3], Mark Viggars[4], Hanna Pincas[1], Nicole R. Gay[2], Yifei Sun[6], Yongchao Ge[1], Venugopalan D. Nair[1], James A. Sanford[7], Mary Anne S. Amper[1], Mital Vasoya[1], Kevin S. Smith[2,5], Stephen Montgomery[2,5], Elena Zaslavsky[1], Sue C. Bodine[8], Karyn A. Esser[4], Martin J. Walsh[6], Michael P. Snyder[2*] , Stuart C. Sealfon[1*] for the MoTrPAC Study Group

[1]Department of Neurology, Center for Advanced Research on Diagnostic Assays, Icahn School of Medicine at Mount Sinai, New York, NY 10029

[2]Department of Genetics, Stanford University, Stanford, CA 94305

[3]Division of Cardiovascular Medicine, Stanford University, Stanford, CA 94305

[4]Department of Physiology and Aging, University of Florida, Gainesville, Florida 32610

[5]Department of Pathology, Stanford University, Stanford, CA 94305

[6]Department of Pharmacological Sciences, Icahn School of Medicine at Mount Sinai, New York, NY 10029

[7]Pacific Northwest National Laboratory, Richland, WA 99354, USA

[8]Division of Endocrinology and Metabolism, Department of Internal Medicine, Carver College of Medicine, University of Iowa, Iowa City, IA 52242, USA

[9]These authors contributed equally

* corresponding authors

# Abstract

Transcription factors (TFs) play a key role in regulating gene expression and responses to stimuli. We conducted an integrated analysis of chromatin accessibility and RNA expression across various rat tissues following endurance exercise training (EET) to map epigenomic changes to transcriptional changes and determine key TFs involved. We uncovered tissue-specific changes across both omic layers, including highly correlated differentially accessible regions (DARs) and differentially expressed genes (DEGs). We identified open chromatin regions associated with DEGs (DEGaPs) and found tissue-specific and genomic feature-specific TF motif enrichment patterns among both DARs and DEGaPs. Accessible promoters of up- vs. down-regulated DEGs per tissue showed distinct TF enrichment patterns. Further, some EET-induced TFs in skeletal muscle were either validated at the proteomic level (MEF2C and NUR77) or correlated with exercise-related phenotypic changes. We provide an in-depth analysis of the epigenetic and trans-factor-dependent processes governing gene expression during EET.

# Abbreviations

| Abbreviation | Definition |
|---|---|
| HEART | Heart |
| SKM-GN | Skeletal muscle (Gastrocnemius) |
| WAT-SC | Subcutaneous white adipose tissue |
| BAT | Brown adipose tissue |
| LIVER | Liver |
| LUNG | Lung |
| KIDNEY | Kidney |
| HIPPOC | Hippocampus |
| EET | Endurance Exercise Training |
| DAR | Differentially accessible regions |
| DEG | Differentially expressed genes |
| DEGaP | Differentially expressed gene associated peaks |
| TSS | Transcription start site |
| TES | Transcription end site |
| TF | Transcription factor |

# Introduction

Regular exercise has a considerable impact on general health, as it helps to maintain muscle function, improves cardiovascular wellness and cognitive performance, and lowers the risk of cardiovascular disease and other disorders, including neurological diseases and cancer[1]. While a number of studies have focused on deciphering the molecular responses to acute exercise, the molecular changes induced by exercise training across tissues are poorly characterized.

As the main regulators of gene transcription, transcription factors (TFs) act via the recruitment of other factors, co-activators, or co-repressors, to cis-regulatory elements at the promoter or distal regions of target genes. The access of TFs to cis-motifs partly depends on chromatin structure. Hence, along with changes in chromatin accessibility and other epigenetic modifications, TFs govern gene expression in tissues as well as gene responses to stimuli. TFs are critical exercise-response mediators[2,3] and, in skeletal muscle, exercise training-induced transcriptomic changes have been associated with different TFs than those induced by acute exercise[4].

The recent multi-tissue analysis of the molecular responses to endurance training points to a majority of tissue-specific differentially regulated genes, along with a smaller proportion of shared training-response genes[5]. Thus, gene responses to training are likely mediated through the combinatorial function of tissue-enriched and shared transcriptional regulators. Shared exercise-induced TF regulation can elicit tissue-specific functions, as seen with PPARγ, which is implicated in PGC1α-stimulated mitochondrial biogenesis[6], regulation of adipogenesis[7], and hippocampal BDNF activity and its cognitive effects[8]. As complex regulatory patterns drive tissue-specific gene regulation[9], they are likely to be involved in mediating the diverse effects of exercise training on tissues. This highlights the importance of identifying the TFs that coordinate gene responses to training in multiple tissues and inferring their underlying mechanisms. However, few studies have evaluated training-induced genome-wide changes in RNA expression and chromatin accessibility[10], and most have concentrated on a few tissues, mainly skeletal muscle.

We leveraged the study design of the Molecular Transducers of Physical Activity Consortium (MoTrPAC) endurance exercise training (EET) study in rats[5] to characterize the TFs mediating gene responses to training across multiple tissues. During 8 weeks of EET, genome-wide transcriptome and chromatin accessibility were assayed in 8 tissues from male and female rats. By connecting both omes within the same tissues, we establish a map of the regulatory transcriptional responses to training across tissues. We investigate highly correlated omic responses to exercise and identify key TFs and pathways involved in those responses in each tissue.

# Results

**Characterization of epigenetic and transcriptional responses to endurance training**

To analyze the epigenetic and transcriptional changes during eight weeks of EET, we analyzed ATAC-seq and RNA-seq datasets generated in skeletal muscle (gastrocnemius; SKM-GN), heart, hippocampus (HIPPOC), kidney, liver, lung, brown adipose tissue (BAT), and subcutaneous white adipose tissue (WAT-SC) from rats subjected to 1, 2, 4, and 8 weeks of training (Fig 1a) and untrained controls. Each of these 5 groups consisted of 5 males and 5 females.

We identified differentially accessible regions (DARs; F test adjusted p value < 0.1), and differentially expressed genes (DEGs; F test adjusted p value < 0.1) between EET and control groups (Fig 1b). To characterize the transcriptional and epigenomic changes induced by EET across tissues, we evaluated the tissue-specificity of DARs and DEGs. Although most expressed genes and open chromatin sites were detectable in multiple tissues, the majority of DARs (90%) and DEGs (66%) were identified in only one tissue (Fig 1c, Fig. S1). This suggested that gene regulatory responses to EET were largely confined to individual tissues, which was in line with another MoTrPAC manuscript (Nair et al., manuscript in preparation).

We then examined the distribution patterns of log2 fold change (L2FC) in gene expression and L2FC in chromatin accessibility across time points and sexes (Fig. S2 and S3). While the ratio of up- to down-regulated analytes (i.e. DEGs or DARs) was similar across sexes and time points in the majority of tissues, distribution patterns differed between tissues. Additionally, we observed sex differences in the expression changes in HIPPOC, BAT, and WAT-SC. We observed sex differences in the chromatin

accessibility changes in BAT as well. Overall, the proportion of DEGs showing concordant changes (L2FC) across the 8 timepoint and sex groups was higher than that of DARs (Fig. 1d). Notably, heart, SKM-GN, and kidney exhibited the most consistency in expression changes across groups, whereas WAT-SC showed the least. In BAT, we detected pronounced variations in DEG as well as in DAR profiles between earlier and late time points in both sexes (Fig. S2g and S3g). Gene set enrichment analysis among week 1 or week 8 DEGs in males and females highlighted varying pathway enrichment patterns that were, in most cases, consistent across sexes and time points (Fig S4). SKM-GN and heart shared enrichment for oxidative phosphorylation and cardiac muscle contraction pathways, as well as markers for Parkinson's, Huntington's, and Alzheimer's diseases. WAT-SC DEGs were enriched for the chemokine signaling pathway and immune-related diseases including systemic lupus erythematosus, asthma, and primary immunodeficiency.

We investigated whether alterations in cell type proportions contributed to the expression changes observed. Cell type deconvolution analysis (See Methods) identified changes in immune cell type proportions that were related to training duration in BAT (Fig. 1e and Fig. S5) and to sex in WAT (Fig. 1f and Fig. S5). Differential analysis was conducted in male and female samples separately, suggesting that DEGs in WAT-SC are independent of sex-specific cell type composition changes; conversely, in BAT, DEGs and DARs cannot be separated from EET-associated changes in cell type composition.

To further characterize the epigenomic changes induced by EET across tissues, we examined the genomic distribution of DARs vs. that of all open chromatin regions detected (Fig. 1g,h). Compared to all accessible regions, DARs were significantly enriched at proximal promoters across all tissues except WAT, which was excluded due to a scarcity of DARs identified (Fig 1h,k, Methods). Consistent with previous studies[11,12], open chromatin peaks across tissues were predominantly located in intronic and distal intergenic regions. Given the importance of the proximal promoter in the regulation of gene transcription[13], the enrichment of DARs in this region suggested that EET results in the transcriptional activation of target genes. Unlike DARs, open chromatin regions that mapped to DEGs, which we refer to as DEG-associated peaks (DEGaPs), shared a similar genomic distribution as the peaks associated with all expressed genes (Fig 1i,j,l). However, ATAC-seq peaks that mapped to either expressed genes or DEGs contained a higher proportion of intronic peaks compared to all open chromatin peaks in the dataset. This suggested that expressed genes tended to contribute intronic peaks, whereas non-expressed genes tended to contribute distal peaks but lack promoter peaks. Indeed, while 73% of genes expressed in SKM-GN contained both promoter and intronic accessible peaks, only 29% of non-expressed genes satisfied that metric. The distinct genomic distribution of DARs vs. DEGaPs suggests their involvement in different regulatory mechanisms.

**Linking epigenetic to transcriptomic responses: few DARs map to adjacent DEGs; identification of distal DARs correlating with DEGs**

We next sought out DAR-DEG associations by assessing the concordance between chromatin accessibility and gene expression changes. We assigned each DAR to the nearest gene, and determined the fraction of DARs that were annotated to DEGs. Applying a hypergeometric test, we found that BAT, SKM-GN, and liver showed a significant overlap between DARs and DEGs (Fig 2a). The substantial overlap between DARs and DEGs in BAT may be related to the EET-induced increase in immune cell populations (see Fig. 1e). SKM-GN and liver showed the highest count of DARs among all tissues (Fig. 1b). Despite hundreds of DARs in both kidney and lung, only a few of their nearest genes were DEGs (Fig 2a). Reciprocally, we identified the DAR(s) that mapped to each DEG, and determined the fraction of

DEGs containing DAR(s) (Fig 2b). While BAT, SKM-GN, and liver showed a large set of DEGs containing DAR(s), only BAT and SKM-GN reached significance.

The binding of TFs to distal open chromatin regions can regulate gene transcription[14,15]. Given the modest proportion of DARs mapped to adjacent DEGs, we extended the search window and sought relationships outside the closest gene for a given DAR. With respect to the location of DARs relative to the TSS of the nearest DEG, in all tissues, the majority of nearest DAR-DEG pairs reflected a normal distribution with a median centered approximately 1 Mb away from the nearest DEG, and a substantial left tail representing closer pairs (Fig 2c). BAT, SKM-GN, and liver contributed most of the DARs adjacent to a DEG, confirming our earlier observations (Fig 2a). When labeling DARs by their genomic region (promoter, intron, etc), we found that proximal promoter-located DARs were enriched among the DARs closest to DEGs (Fig 2d). DARs that were adjacent to DEGs tended to be more highly correlated with gene expression changes (Fig 2e, Fig S6). We focused on DARs and DEGs located within 500 kb of each other, and isolated those that were either positively or negatively correlated across time points and sexes (Pearson correlation coefficient > 0.5 or <-0.5, respectively; Fig. 2f, Supplementary Table 1). Thus, the identified DAR-DEG pairs comprised DARs that mapped to nearby DEGs along with distal, within-500 kb DARs that correlated with DEGs.

**DAR-DEG pairs are associated with distinct pathways in each tissue and MAZ and SMAD3 represent key regulatory TFs**

We sought to determine which biological processes were associated with the DEGs that correlated, either positively or negatively, with DARs (located within 500 kb) across time points and sexes. Pathway enrichment analysis identified distinct patterns of enrichment for each tissue (Fig 3a). In agreement with the training-associated increase in immune cell types inferred from cell type deconvolution analysis (see Fig. 1e), BAT DAR-DEG pairs showed enrichment for several immune pathways. Lung also showed considerable enrichment for immune-associated pathways. By contrast, liver DAR-DEG pairs were primarily enriched for lipid biosynthesis and metabolic processes, while heart DAR-DEG pairs were enriched for muscle movement and filament sliding, and SKM-GN DAR-DEG pairs were enriched for myofiber synthesis and muscle contraction (Fig 3a). These results suggested that the correlated epigenetic and transcriptional changes induced by training affected tissue-specific functions.

We analyzed TF motif enrichment at the DARs of DAR-DEG pairs to identify potential regulators of the training response in each tissue. Notably, we identified MAZ and SMAD3 as regulatory TFs in specific tissues (Fig. 3b-d). In SKM-GN, an 8.6% enrichment for MAZ binding sites was found among DAR-DEG pairs vs. a 2.7% enrichment among active peaks in that tissue (p-value = 0.01912). In lung, there was a 26% enrichment for MAZ binding sites among DAR-DEG pairs vs. a 2.7% general enrichment (p-value = 1.168e-04). In SKM-GN and in the lung, MAZ motifs were predominantly found in DARs that were negatively correlated with differential gene expression (Fig 3b,c). MAZ can act both as a transcriptional activator and a repressor[13,16]. In SKM-GN, MAZ target genes included: Igf2 (Fig 3e), which plays pivotal roles in exercise response[17,18], SKM growth, and differentiation[19]; Ppp1r15a, which is associated with innate immunity[20]; and Sall2 (Fig 3f), a TF typically associated with development and neuronal differentiation[21]. In the lung, MAZ targets included immune response genes such as Mpeg1, Oas2, Nfkb2 (Fig 3g), as well as stress response gene Hspb6.

SMAD3 binding sites were enriched in the liver (Fig 3d), with a 22% enrichment among DAR-DEG pairs vs. a 5.4% general enrichment (p-value = 1.255e-05). Some of the SMAD3 target genes in liver had paired DARs corresponding to SMAD3 motifs, suggesting combinatorial transcriptional regulation (Fig

3d). Paired DAR sets were positively correlated with *Glul* expression (Fig S7a,b), negatively correlated with *Lpar3* (Fig S7c,d) expression, and negatively correlated with multiple members of the *Serpina* gene family (Fig 3d, S7e,f). *Fkbp4* expression was also negatively correlated with a single SMAD3 motif-containing DAR (Fig 3h). The remaining SMAD3 targets *Abhd2* (Fig S7g), *Onecut1* (Fig S7h), *Ccnd1* (Fig S7i), and *Xbp1* (Fig S7j) shared a similar training response pattern with lower L2FC in male subjects than female subjects in most time points. Notably, SMAD3 has been identified as a major regulator of exercise response in human SKM[22]. Other potential transcriptional regulators along with target genes are illustrated in Fig. S7.

**Characterization of TF expression responses to EET in each tissue**

As putative transcriptional regulators were inferred from DAR-DEG correlations in a restricted number of tissues, we sought to independently characterize TF expression responses to EET per tissue over the 8-week training period. We measured the RNA levels of all TF-encoding genes and assessed their protein levels, when available, based on mass spectrometry data obtained from a subset of 6 tissues. Various subsets of TFs exhibited significant changes at the transcriptome (Fig 4a) and proteome (Fig 4b) levels in each tissue. BAT had the largest and most significant changes in TF gene expression (Fig 4a), which included immune-related TF genes such as *Irf1*, *Irf8*, *Pou2f2*, and *Spi1*. While TF gene responses were largely tissue-specific, some gene expression changes were shared across tissues. Most notably, both *Fos* and *Egr1* transcript levels decreased in SKM-GN, kidney, heart, lung, and to a lesser extent HIPPOC. *Klf9* increased slightly in all tissues, and dramatically in SKM-GN, and *Myb* expression increased in WAT-SC, HIPPOC, kidney, and lung. To evaluate inter-tissue differences in TF responses, we performed pairwise analyses of the fold-change in TF gene expression. Lung and kidney showed the highest correlation, and SKM-GN and heart, both striated muscle tissues, showed a relatively high correlation as well (Fig 4c).

Similar to TF gene expression changes, TF protein level responses to EET were largely tissue-specific (Fig 4b). WAT-SC and lung exhibited the most significant changes. While cross-tissue TF protein level changes showed little overlap, a few shared TFs decreased in protein abundance in both tissues, namely MAX, STAT4, ELF1, RUNX1, and HOXB4. CLOCK protein levels increased in WAT-SC and decreased in heart. NR4A1 (NUR77) and MEF2A levels decreased in SKM-GN, while MEF2C levels increased. ATF3 decreased at the proteomic level in heart and liver and at the transcriptomic level in SKM-GN. RORC levels decreased in lung, kidney, and liver, but increased in heart. Pairwise comparisons of the fold change in TF protein abundance across the subset of 6 tissues showed the strongest positive correlation between kidney and lung (Fig. 4d), which was consonant with TF gene expression fold changes. On the other hand, heart and kidney were the most negatively correlated, whereas they were positively correlated at the RNA level. Post-transcriptional or translational regulatory mechanisms could be responsible for this discordance.

We next examined the target genes of TFs regulated in SKM-GN at both transcript and protein levels. Genes with MEF2C or NR4A1 motifs in open proximal promoter regions were more likely to be DEGs (Fig 4e, Fig S8). There was a substantial increase in MEF2C protein levels in SKM-GN after 2 weeks of EET, followed by a return towards baseline. DEGs that harbored MEF2C motifs at the promoter showed diverse EET responses, yet the majority showed expression changes that positively correlated with MEF2C protein changes (Fig 4f). Among the MEF2C target genes, clock gene *Per1* demonstrated the most differential expression throughout EET, while *Sema6c* and *Phkg1* showed the strongest positive correlation with MEF2C protein changes. *Dystrophin* (*Dmd*), a critical protein for muscle fiber integrity[23], was the most negatively correlated with MEF2C. Decreased expression of the four DEG targets of NR4A1

coincided with reduced NR4A1 protein levels in SKM-GN (Fig 4g). Heat shock protein *Hspa1l* and dual specificity phosphatase *Dupd1* were the most differentially regulated. Altogether, these findings support the functional relevance of EET-regulated TFs.

**DARs vs. DEGaPs show distinct TF motif enrichment patterns that differentially correlate with TF gene expression**

The lack of nearby DARs for the majority of DEGs within each tissue (see Fig. 2a,b) led us to hypothesize that DARs and DEGaPs may mediate different paths of transcriptional regulation: i) DARs coordinating a combination of direct and long-range regulatory mechanisms, ii) a combination of statically open cis-regulatory elements (DEGaPs) and changes in TF behavior influencing differential gene expression. To address this, we analyzed TF binding site enrichment at either DARs or DEGaPs relative to all open chromatin peaks in each tissue (Fig. 5a,b; see Methods).

Due to their very low number of DARs, WAT and HIPPOC were removed from the analysis (Fig. 5a). Motif enrichment patterns varied greatly across the six remaining tissues, reflecting a high degree of tissue-specificity. Motifs for both FOX and KLF families of TFs were over-represented among SKM-GN DARs. KLFs are zinc-finger TFs that have been associated with myogenesis and muscle fusion via their recruitment to Muscle Creatine Kinase (MCK) promoters[24]. SIX2 and MEF2C motifs were enriched in heart, COUP-TFII in kidney, and FOXO3 and SP2 in liver. HOXA10 and HOXD12 motifs were enriched in lung, while both IRF8 and IRF3 motifs were enriched in BAT. Tissue-specific patterns of motif enrichment were maintained when measuring the frequency of motif presence in DARs across tissues (Fig S9a).

Motif enrichment patterns among DEGaPs differed considerably from those in DARs (Fig. 5b). Indeed, motif enrichment at DARs vs. that at DEGaPs per tissue were weakly correlated (<0.39; Fig S11). Motif enrichment significance in DEGaPs was greater in lung, BAT, and WAT-SC, presumably due to their higher proportions of DEGs (see Fig. 1b). MEF2 TF motifs were enriched across lung, BAT, and WAT-SC, as well as in SKM and heart, forming one cluster; on the other hand, ETS and ELF TF motifs were more exclusively enriched in lung, BAT, and WAT-SC and formed another cluster. MEF2 TFs are typically involved in muscle tissue regeneration[25]. ETS and ELF TFs are associated with the regulation of immunity[26,27], suggesting that they may be related to the immune cell type composition changes occurring in adipose tissues. With respect to tissue-specific enrichment, motifs of immediate early genes including JUN and FOS were enriched in HIPPOC (Fig S10a), HNF1 and ZIC families of TFs in kidney, and FOX and SOX TF families in liver.

To establish correlations between the motifs enriched at either DARs or DEGaPs and the relative expression levels of the corresponding TFs, we examined TF gene expression patterns in control tissues (Fig S9b, Fig S10b respectively) and L2FC in TF gene expression following EET (Fig S9c, Fig S10c, respectively). We found a stronger correlation between motif enrichment and TF control gene expression levels among DEGaPs (Fig S10d-f) than among DARs (Fig S9d-f). Conversely, there was no correlation between motif enrichment and L2FC in TF gene expression, be it among DARs (Fig. S9a,c) or DEGaPs (Fig. S10a,c). These findings suggest that the influence of DEGaPs on DEG expression is more dependent upon the expression level of the associated TFs, whereas DARs can influence DEGs directly, supporting the concept that DEGaPs vs. DARs mediate two distinct paths of transcriptional regulation.

**DARs and DEGaPs show cross-tissue motif enrichment conservation at specific genomic regions**

The different genomic distributions of DARs vs. DEGaPs (see Fig. 1h,j) also suggested that some regions may contribute more to the regulation of EET gene responses than others. Thus, following the mapping of tissue DARs and DEGaPs to various genomic regions (i.e. promoter, upstream, downstream, 3'UTR, 5'UTR, intron, exon, distal intergenic), we examined motif enrichment among each subcategory of DARs and DEGaPs. We found that tissues shared conserved patterns of TF motif enrichment at the DARs mapped to proximal promoter, downstream, and 3'UTR regions (Fig 5c). Proximal promoter DARs showed strong enrichment for SP and KLF TFs, ZFX, NF1, and the circadian clock TF BHLHE41. Downstream DARs were enriched for NPAS and BMAL1, two core circadian clock TFs, as well as NKX and HOX TFs. The NF1 subcluster also showed enrichment among 3' UTR DARs, while THRB was most highly enriched among 3' UTR DARs.

Correlation analysis at DEGaPs revealed that TF motif enrichment was conserved across tissues predominantly at distal intergenic/intronic regions, followed by proximal promoter and exonic regions (Fig. 5d). When we selected the motifs showing the most differential enrichment between the different DEGaP subcategories, we found that proximal promoters were enriched for SP and KLF TFs, as in DARs, in conjunction with LRF, E2F3, and E2F4 (Fig. 5e). In contrast with DARs, AP-2γ was enriched at distal promoter DEGaPs. Members of the FOX TF family were enriched among intronic and distal intergenic DEGaPs. THRB and NF-E2 motifs were the most enriched among upstream DEGaPs. Overall, genomic features appeared to be a driving force in clustering the enriched binding motifs among DEGaPs and DARs, and in both instances motif enrichment was conserved across tissues at proximal promoters.

**Differential TF motif enrichment at DEGaPs of upregulated vs. downregulated DEGs**

In a concurrent MoTrPAC-study (Nair et al. manuscript in preparation), the biological pathways that were over-represented among upregulated vs. downregulated DEGs in each tissue after 8 weeks of training were divergent, suggesting that different sets of TFs were involved in their regulation. We sought to predict the sets of TFs that may preferably bind to up-regulated genes, and those that may bind to down-regulated genes. After identifying the DEGaPs related to up-regulated DEGs and those linked to downregulated DEGs in each tissue, we refined each DEGaP category into distal intergenic, intron, and promoter peak subsets (the most prevalent genomic features among DEGaPs, as shown in Fig. 5d), and determined their respective patterns of TF motif enrichment. We then generated z scores for each TF enrichment across the peak sets, and applied hierarchical clustering to the TFs. We found that, in most tissues, clustering was mainly driven by differences in TF motif enrichment between the promoter peaks associated with up-regulated DEGs vs. those associated with down-regulated DEGs (Fig S12). Adipose tissues had the greatest motif enrichment overlap between promoter peaks associated with upregulation and those associated with down-regulation.

Restricting the TF motifs enriched within each subset of promoter peaks (i.e. promoter peaks associated with either up- or down-regulation) to those conserved across all four time points (Fig. S13), revealed some overlap in motif enrichment between the tissues (Fig. 6a,b). For instance, CLOCK, BHLHE41, and MYC (c-Myc) were among the TF motifs enriched across WAT-SC, heart, lung, SKM-GN, and kidney at the promoter peaks associated with upregulation; another set of TFs enriched primarily in heart and lung included ZFX and EBF2, while ATF2 was enriched in SKM-GN and kidney. Among the promoter peaks associated with downregulation, there was less consistent enrichment across tissues, however NKX2 was enriched across SKM-GN, kidney, liver, and lung, and SOX9 was enriched among 6 of the 8 tissues. Twenty seven TFs were shared between the 50 TFs associated with upregulation (Fig. 6a) and the 39 associated with downregulation (Fig. 6b), suggesting their involvement in both positive and negative

regulation depending on the tissue. Heart, SKM-GN, and kidney shared the most enriched motifs associated with upregulation and lung, liver, and HIPPOC shared the most enriched motifs associated with downregulation (Fig 6c,d). Overall, while gene responses and TF responses were largely tissue-specific, tissues showed an overlap in the sets of TFs associated with either upregulated or downregulated DEGs, suggesting that specific patterns of EET-induced regulation could be conserved across tissues.

**Identification of DEGs associated with EET-induced phenotypic changes**

Exercise elicits various phenotypic changes such as increased aerobic capacity (VO2max) and reduced body fat mass, all of which reflect physiological adaptations. As several parameters were clinically measured throughout the EET period, we sought to: i) assess possible correlations between phenotypic alterations and gene expression responses to training, ii) infer key transcriptional regulators based on their motif enrichment at the accessible promoters of DEGs that correlated with phenotypic changes.

Parameters measured included weight, body fat mass, body lean mass, body water, lactate levels, and VO2 max (Fig. S14). Body weight and fat mass were significantly lower in 8-week trained rats than in controls (Fig S14a,b), with the greatest difference observed in males. VO2 max significantly increased in both females and males in response to EET (Fig S14c). Body lean mass also increased in both sexes, though significance was only reached in females (Fig S14d). Lactate (Fig S14e) and body water (Fig S14f) showed no significant changes relative to control, though a significant decrease in lactate levels was observed between week 4 and week 8 in males. Positively correlated parameters included VO2 max vs. body lean mass, and body weight and body fat mass (Fig 7a).

In most tissues, DEGs formed 2 separate clusters based on their either positive or negative correlation with a measured parameter change (Fig S15). Correlation with body fat involved the highest proportion of DEGs in most tissues (Fig. 7b), except for BAT and HIPPOC.

**Identification of key TFs associated with phenotypic changes**

To identify key regulators of the DEGs that were associated with phenotypic changes, we analyzed motif enrichment at the corresponding promoter DEGaPs. In SKM-GN, the promoter DEGaPs of genes that were positively correlated with changes in VO2 max were significantly enriched for NR5A2 and ERRG motifs (Fig. 7c). The VO2 max-correlated DEG targets of those two TFs and of AP-2GAMMA and AP-2ALPHA, whose motif enrichment did not reach significance, showed considerable overlap (Fig 7d). The positive correlation between three of the DEGs targeted by either NR5A2, ERRG, or/and AP-2GAMMA (*Me3*, *Rora*, and *Lgi3*) and VO2 max change is depicted in Fig. 7e-g.

In SKM-GN, we also found a positive correlation between DEGs and body weight changes, and identified the MyoD motif as significantly enriched at the corresponding promoter DEGaPs (Fig. 7h). While an enrichment for the COUP-TFII and ERRA motifs was also detected, it did not reach significance (Fig 7h). The *Chd7, Rnf13,* and *RGD1562029* DEGs showed concurrent enrichment for the MyoD, COUP-TFII, and ERRA motifs (Fig 7i). The positive correlation between *Chd7* expression and body weight change is illustrated in Fig. 7j.

As MAZ and SMAD3 have been identified as major regulators of the DARs-associated DEGs (DAR-DEG pairs; see Fig. 3), we asked whether those DEGs correlated with phenotypic changes. In SKM-GN, we found that MAZ targets *Igf2* and *Sall2* were positively and negatively correlated with body fat change, respectively (Fig. 7k,l); in lung, *Oas2* and *Nfkb2* were negatively and positively correlated with body fat

change, respectively (Fig. 7m,n); in the liver, SMAD3-target *Fkbp4* was also positively correlated with body fat change (Fig 7o). In each of those correlations, as well as in the positive correlation between *Chd7* and body weight, we noted sex differences both in EET-dependent RNA levels and in phenotypic changes, as week 8-trained males deviated more from controls than week 8-trained females (Fig. 7j-o).

In view of the characterization of TF expression responses to EET, we next evaluated correlations between TF expression changes, at either the RNA or protein level (see Fig. 4) and phenotypic changes. We observed the strongest correlations between TF RNAs and body fat changes (Fig S16a), with *Pknox1* in SKM-GN being the most positively correlated, and *Srebf1* in the kidney being the most negatively correlated. *Srebf1* in the kidney was also the most positively correlated with body lean mass changes. *Rora* was the most positively correlated with VO2 max changes. *Zeb1* and *Zeb2* shared similar correlation patterns, namely a positive correlation with body fat in WAT-SC. *Gata2* had diverging correlation patterns depending on the tissue: positively correlated with body weight and body fat in WAT-SC, and negatively correlated with both measures in the lung. Similar to the correlations with TF RNA level-changes, the strongest correlations at the protein level were with body fat changes. Remarkably, liver showed the highest frequency of strong correlations (Fig S16b). In WAT-SC, JUN was the most positively correlated with body fat, while TCFL2 and EHF were the most negatively correlated. ZEB1 and ZEB2 maintained their correlation structure seen at the RNA level, with ZEB2 being also positively correlated with body fat in the lung. IRF3 in SKM-GN was the most positively correlated TF with body lean changes and CLOCK in WAT-SC was negatively correlated with body fat. Connecting EET-modulated genes and proteins, and specifically TFs, with correlated phenotypic changes suggest potential mechanistic roles. In particular, MAZ and SMAD3 target gene correlation with body fat changes reinforces their functional role in the training response.

# Discussion

Regular exercise has a variety of physiological benefits affecting many organ systems. We integrated chromatin accessibility and transcriptomic data from 8 rat tissues to infer the TFs underlying EET responses in each tissue. We found multiple layers of regulation characterizing EET adaptation, including utilization of the innate tissue-specific and genomic region-specific TF machinery, changes in TF expression at transcriptome and proteome levels, direct relationships between proximal promoter DARs and their associated DEGs, interactions between more distant highly correlated DAR-DEG pairs enriched for known TFs and specific training responses with distinct TF enrichment patterns, some conserved across tissues (Fig 8). We isolated MAZ and SMAD3 as key regulatory TFs in SKM-GN, lung, and liver, found the down-regulation of immediate-early response genes over EET and enrichment of KLF and SP TF motifs among proximal promoter DEGaPs. Consistent with the findings of a MoTrPAC companion study, we found that the DEGaPs of upregulated vs. downregulated DEGs show differential motif enrichment, which predominantly occurs at promoter regions. Finally, we singled out TFs that correlated with EET-induced phenotypic changes including VO2 max and body weight.

The EET of rats allowed studying responses in systems not feasible in human subjects. Limitations of our study include the use of inbred rats, which eliminates effects of genetic diversity, and high variance in some tissues, thus reducing statistical power. However, differences in cell type composition within testing groups of same strain, same age, same sex rats suggest greater than initially anticipated variation and stronger applicability to the diversity expected in human responses to EET. Exercise effects on cell type composition could only be addressed computationally in assays of bulk tissue. Many integrated

multiomic EET responses that we identified, such as muscle development in SKM-GN, innate immunity in lung, increased immune cell type proportions in adipose tissues, and metabolic processes in liver may contribute to known health effects of exercise. However, as the goal of the MoTrPAC study is to provide a systematic compendium of exercise effects, follow-up studies of the multi-organ roadmap of genomic regulatory responses that we have identified are warranted.

In liver, the SMAD3 TF motif was the most enriched among correlated DAR-DEG pairs following EET. While SMAD3 was recently established as a key regulator of the acute exercise response in SKM[22], it is also an intracellular mediator of the TGF-β signaling pathway, which has been associated with hepatic stellate cell (HSC) activation and liver fibrosis[28,29]. SMAD3-target genes include collagens and fibrogenic markers such as *smooth muscle actin* and *cadherin*. In this work, SMAD3 motif-containing DARs anti-correlated with *Serpina* family gene responses. Low serum levels of *Serpina4*, also known as *kallistatin*, have been associated with liver diseases, including liver fibrosis[30]. SMAD3-target gene *Lpar3* has been linked to liver regeneration following damage[31] and *Lpar* family members were previously associated with liver fibrosis[32]. RNA expression of *Glul*, another potential SMAD3-target gene, was reported to decrease in activated HSCs, which are responsible for extracellular matrix deposition in liver fibrosis[33]. We speculate that TGF-β/SMAD signaling, which regulates ECM production and cytoskeletal organization in HSCs, may be modulated by EET.

In SKM-GN and lung, MAZ was the most enriched TF motif among correlated DAR-DEG pairs. Although it was not previously linked to EET, MAZ can act both as a transcriptional activator and a repressor[13,16]. Multiple MAZ target genes found among the DAR-DEG pairs were associated with immune response: *Ppp1r15a* (*Gadd34*) is necessary for interferon production[34]; *Igf2*, which is primarily a growth hormone during development, is also involved in immune response[35–37]; *Hspb6* acts as a mediator of platelet aggregation along with smooth muscle relaxation[38]; *Mpeg1* expresses a transmembrane protein in macrophages in an antimicrobial capacity[39]; *Oas2* is a type I-interferon response gene[40]; mutations in *Nfkb2* greatly damage the immune system[41]. Thus, MAZ could be involved in the inflammatory response occurring in SKM following exercise[42].

Several TFs that were downregulated by EET in SKM-GN either at the RNA level (*Fos, Jun, Egr1,* and *Atf3*) or at the protein level (NR4A1) represent immediate-early response genes (IEGs), which were previously reported to be induced by acute exercise[43–46]. Stress-inducible ATF3 was shown to reduce the RNA expression of inflammatory chemokines and cytokines in mouse SKM following acute exercise, and ATF3 knockout resulted in impairment of some of the molecular adaptations to exercise training[43,47]. In SKM, NR4A1 regulates genes associated with glucose uptake, glycogen synthesis, and promoting muscle growth[48] as well as mediating inflammatory response[49,50]. As exercise represents a stressful stimulus, modulation of IEGs may help tissues like SKM-GN to recover their homeostasis and thus facilitate their adaptation to exercise training. The two gene targets of NR4A1 with the largest correlated decreases in EET-response gene expression were the heat shock protein *Hspa1l*[51], a critical element of the cellular stress response, and *Dupd1* (*Dusp29*), which has been linked to regulation of muscle cell differentiation, development, and atrophy[52].

In the present work, we determined that DARs vs. DEGaPs have different genomic distributions and distinct TF motif enrichment profiles, and found a much stronger correlation between motif enrichment and control TF gene expression levels at DEGaPs compared to DARs. These findings suggest that DARs and DEGaPs represent two distinct patterns of transcriptional regulation in response to EET. The DAR-mediated mechanism involves EET-induced changes in chromatin accessibility at either promoter or distal gene regions, resulting in the activation or repression of gene transcription. The DEGaP-mediated

mechanism involves chromatin regions that are already accessible in the control tissue, and thus is most likely driven by TF activity. TF activity is influenced by post-translational modifications (PTMs), heterodimerization, and interactions with co-factors. PTMs include phosphorylation, sumoylation, ubiquitination, acetylation, glycosylation, and methylation[53]. PTMs can: i) alter TF conformation and subsequently their affinity for DNA or their stability, ii) influence TF nuclear translocation and hence their access to DNA. Although our results support a model wherein one path is predominantly controlled by the impact of EET on chromatin structure, and the other is predominantly driven by the effect of EET on the transcriptional activity of TFs, further analysis is needed to investigate this concept.

In both DARs and DEGaPs, the proximal promoter region best illustrated the clustering of motif enrichment. Notably, KLF and SP TFs, whose motifs were over-represented at promoter DEGaPs, bind to GC-rich regions and GT boxes[54], which are typically enriched at proximal promoters. These members of the KLF and SP TF families are known to act on a variety of biological pathways depending upon tissue and context. Tissue-specific expression of KLF and SP family members is likely the critical step for determining their functionality. For example, *Klf5* gene expression was highest in lung, while *Klf14* was primarily expressed in WAT-SC.

We found that binding sites for circadian clock TFs were over-represented at the promoter DEGaPs of DEGs that were associated with VO2 max changes in SKM-GN. Notably, there were enriched motifs for the nuclear receptors NR5A2 and ERRG[55] which bind to the core clock proteins CLOCK[56] and BMAL1[57], respectively. Moreover, we found an over-representation of binding sites for circadian clock-related TFs at promoter DARs (BHLHE41 motif), at downstream DARs (NPAS and BMAL1 motifs), and at the DEGaPs of upregulated DEGs (CLOCK motif) in multiple tissues, suggesting a cross-tissue, exercise training effect on circadian clock factors. It is noteworthy that the EET program in rats was conducted during an active, dark phase, thus excluding that the enrichment for circadian clock TF motifs would be related to exercise training in the inactive phase. Recent studies revealed interactions between exercise and circadian rhythms. Some demonstrated that time of exercise can modify the transcriptional response to acute exercise, while others showed that exercise can modify the muscle clock phase and expand the circadian transcriptome in SKM[58–63].

In summary, our multi-omic analysis across rat tissues allowed us to map the epigenomic changes to the transcriptional changes occurring during EET and infer the TFs driving training responses. By providing a view of the complex interplay between chromatin structure modifications, gene transcription, and TF abundance throughout EET, this work sets the stage for a molecular understanding of the multi-organ effects of regular exercise. Ultimately, this knowledge will be utilized to maximize the benefits of exercise to healthy individuals and develop targeted therapies for patients with a disease or disability.

# Methods

## Animal study design

### Animal care

Male and female Fischer 344 (F344) inbred rats from the National Institute on Aging were transported to the University of Iowa a minimum of 4 weeks prior to starting exercise training. Rats were housed with

the same sex, 2 per cage (146.4 square inches of floor space) in ventilated racks (Thoren Maxi-Miser IVC Caging System) on Tekland 7093 Shredded Aspen bedding and fed the Lab Diet 5L79 pelleted diet.

Rats were acclimated to a reverse dark-light cycle with lights OFF at 9:00am and lights ON at 9:00pm, with temperature controlled at 68-77°F and humidity of 25-55%. For these studies we will use Zeitgeber Time (ZT) to refer to time of day relative to the time of lights off/lights on with lights off at ZT12. All experimental interventions and husbandry were performed during the active, dark phase of the rats under red light. All animal procedures were approved by the Institutional Animal Care and Use Committee at the University of Iowa.

**Treadmill familiarization**

Prior to exercise training, rats were acclimated to treadmill exercise on a Panlab 5-lane rat treadmill (Harvard Instruments, Model LE8710RTS). Day 1-2 consisted of static treadmill exploration for 10 minutes. Days 3-5 consisted of running at 6m/min at 0° incline for 10 minutes, speed was increased to 10m/min between Days 6-12. On Day 12, rat running behavior was scored for compliance with running at 12m/min for 5 minutes at a 10° incline. Based on running behavior, rats received a score from 1-4 with 4 being the highest score. Rats that were assigned a score of 1 were removed from the study. 25 male and 25 female compliant rats were randomized to a control or training group and the body weights of the rats within each group were not different..

**Progressive exercise training protocol**

At 6-months of age, 1, 2, 4 or 8 weeks of exercise training began. Control rats were placed on a static treadmill for 15 min per day, 5 times per week. Exercise training consisted of a progressive training protocol 5 times per week at ZT13-20, to target 70% VO2max (see below). Week 1 sessions started at 13m/min for males and 16m/min for females at 5° for 20 minutes, with duration increased by one minute each day until reaching 50 min on day 31 of training. The treadmill grade was increased from 5° to 10° at the start of week 3. The treadmill speed increased at the start of week 2 (15m/min males, 18m/min females), 4 (18m/min males, 21m/min females), 5 (20m/min males, 23m/min females), 6 (23m/min males, 26m/min females), and 7 (25m/min males, 28m/min females) and was fixed for the final 10 days of training. Rats performing less than 4 days of training per week were removed from the study and euthanized as described below.

**Body composition measurements**

Body composition (lean tissue%, fat tissue% and body fluid) was measured for all rats 13 days prior to the start of training and 5 days prior to euthanasia in the 4 and 8-week training groups using the minispec LF90II Body Composition Rat and Mice Analyzer (Bruker, 6.2 MHz Time Domain Nuclear Magnetic Resonance (TD-NMR) system). VO2max was determined prior to commencing training in all rats, and during the last week of training for rats in the 4- and 8-week exercise groups, in a single-lane enclosed treadmill (Columbus Instruments Metabolic Modular Treadmill), with rats acclimated two days prior to testing. For testing, the rat was placed in the treadmill and testing began once oxygen consumption stabilized. The testing protocol consisted of a 15 min warm up at 9 m/min and 0° incline. The incline was increased to 10° and treadmill speed was increased by 1.8 m/min every 2 minutes[64]. During the test, electric shocks were used if the rat stopped running and sat on the shock area. Testing stopped when the rat sat on the shock area 3 consecutive times and did not respond to increased shock. Blood was then taken from the tail to measure lactate. VO2max was determined as a leveling off of

oxygen uptake, despite increased workload, a respiratory exchange ratio above 1.05, and an unhaemolyzed blood lactate concentration ≥6 mM.

### Euthanasia and tissue collection

On the day of euthanasia, food was removed at ZT11.5, 3 hours before tissue collections which took place between ZT14.5-17.5, a minimum of 48 hours post their last exercise bout. Rats were sedated with approximately 1-2% isoflurane in oxygen, then hearts removed followed by immediate decapitation using a guillotine. The brain was removed from the skull and hippocampus dissected. All tissues (heart, lung, liver, kidney, white adipose, brown adipose, skeletal muscle, and hippocampus) were cleaned of excess connective/fat tissue and immediately flash-frozen in liquid nitrogen, placed in cryovials and stored at -80°C. Rat tissues were archived and cyropulverized at the MoTrPAC Biospecimens Repository, until distributed to Chemical Analysis Sites for respective assays[5].

## Data production and quantification

### ATAC-seq data generation

Nuclei from 30 mg white adipose, 15 mg brown adipose, and 10 mg of other tissue samples were extracted using the Omni-ATAC protocol with modifications[65]. The white adipose, brown adipose, and hippocampus tissues were processed using no-douncing nuclei extraction. Cryopulverized tissue powder was incubated in the homogenization buffer for 10 min at 4°C, tubes inverted every 2-3 minutes. Other tissue powder was incubated in the homogenization buffer for 5 minutes on ice and dounced 10 times using pestle A and 20 times with pestle B. Nuclei were stained with DAPI and counted using an automated cell counter. 50,000 nuclei (or max. 500 µl nuclei) were added to 1 ml ATAC-RSB buffer and spun at 1000 g for 10 minutes, and the supernatant was removed.

The nuclei pellet was resuspended in 50 µl of transposition mixture and incubated at 37°C for 30 minutes with 1000 rpm shaking. The transposed DNA was purified using Qiagen MinElute Purification kits (Qiagen # 28006), and amplified using NEBnext High-Fidelity 2x PCR Master Mix (NEB, M0541L) and custom indexed primers[65]. 1.8x SPRIselect beads were used to clean the PCR reaction and remove primer dimers. The ATAC-seq libraries were sequenced on a NovaSeq 6000 using 2x50bp with 35 million pairs of reads per sample.

### ATAC-seq data processing and normalization

Reads were demultiplexed with bcl2fastq2 (v2.20.0) and processed with the ENCODE ATAC-seq pipeline (v1.7.0) (https://github.com/ENCODE-DCC/atac-seq-pipeline)[66]. Samples from the same sex and training group were analyzed as biological replicates. Reads were aligned to genome rn6.0.96[67] with Bowtie 2 (v2.3.4.3)[68]. Duplicate and mitochondrial chromosome mapped reads were removed. Peaks were called using MACS2 (v2.2.4)[69], both from reads from each sample and pooled reads from all biological replicates. Pooled peaks were compared with the peaks called for each replicate individually using irreproducible discovery rate[70] and thresholded to generate an optimal set of peaks. Optimal peaks from all workflows were concatenated, trimmed to 200 base pairs around the summit, sorted and merged with bedtools (v2.29.0)[71] to generate a master peak list. This peak list was intersected with the filtered alignments from each sample using bedtools coverage to generate a peak by sample matrix of raw counts. Peaks from non-autosomal chromosomes were removed. Peaks that did not have at least 10

read counts in four samples in a tissue were removed. Filtered raw counts were then quantile-normalized with limma-voom[72]. This version of the normalized data was used for downstream analyses.

**ATAC-seq peak assignment to genomic features**

Accessible regions identified using ATAC-seq were assigned to one of the nine terms of genomic features using Ensembl Rn6 GTF (gene annotation file) and function annotatePeak of package ChIPseeker[73] (v1.8.6).

Nine genomic features are defined as:
Promoter (<=1 kb) (proximal promoter): within +/- 1 kb from the transcription start site (TSS);
Promoter (1-2 kb): +/- 1 to 2 kb from the TSS;
Upstream (<5kb) 2-5 kb upstream from the TSS;
Downstream (< 5 kb): within 5kb downstream of the transcription end site (TES);
5' UTR (5' untranslated region); Exon; Intron; 3' UTR (3' untranslated region);
Distal Intergenic: regions > 5kb downstream of TES or > 5kb upstream from next TSS;
Overlaps Gene: overlaps with gene annotation, but not in any terms above.

All ATAC-seq identified accessible regions were assigned to the closest genomic feature of a genome. Differentially expressed gene associated peaks (DEGaPs) are defined as all accessible regions assigned to the differentially expressed gene.

**RNA-seq data generation**

Rat tissue powders were further disrupted using Agencourt RNAdvance tissue lysis buffer (Beckman Coulter, Brea, CA) with a tissue ruptor (Omni International, Kennesaw, GA, # 19-040E). The total RNA was quantified using NanoDrop (ThermoFisher Scientific, # ND-ONE-W) and Qubit assay (ThermoFisher Scientific). Total RNA quality was determined by either Bioanalyzer or Fragment Analyzer analysis.

500 ng total RNA was used for library generation. Universal Plus mRNA-Seq kit from NuGEN/Tecan (# 9133) was used to select polyadenylated RNA. The generated sequencing libraries contain dual barcodes (i7 and i5) and UMIs (unique molecular identifiers) to accurately quantify the transcript levels. The RNA-seq libraries were sequenced on a NovaSeq 6000 using 2x100 bp with 35 million pairs of reads per sample.

**RNA-seq data processing and normalization**

Reads were demultiplexed with bcl2fastq2 (v2.20.0). Adapters were trimmed with cutadapt (v1.18). STAR (v2.7.0d) was used to index and align reads to genome rn6.0.96 and gene annotations[67]. Bowtie 2 (v2.3.4.3) was used to index and align reads to globin, rRNA, and phix sequences in order to quantify the percent of reads that mapped to these contaminants and spike-ins[68]. UMIs were used to accurately quantify PCR duplicates with NuGEN's "nudup.py" script (https://github.com/tecangenomics/nudup). QC metrics from every stage of the quantification pipeline were compiled, in part with multiQC (v1.6)[74]. Lowly expressed genes (having 0.5 or fewer counts per million in all but one sample) were removed and normalization was performed separately in each tissue. These filtered raw counts were used as input for differential analysis with DESeq2[75]. To generate normalized sample-level data, filtered gene counts were TMM-normalized using edgeR::calcNormFactors, followed by conversion to log counts per million with edgeR::cpm[76]. The same normalization technique was used on the 8 week control samples of each tissue for cross-tissue comparisons.

### Proteomics data generation

Liquid chromatography tandem mass spectrometry (LC-MS/MS) was conducted on six tissues: heart and liver at the Broad Institute and skeletal muscle, kidney, lung, and white adipose at Pacific Northwest National Laboratory (PNNL). Sample processing followed a modified version of a previous protocol[77]. Peptides were labeled using tandem mass tag (TMT)[78] and samples were grouped into sex- and training time point-based TMT11 multiplexes. Multiplex samples were fractionated by high pH reversed phase separation. Heart and liver samples underwent online separation with a nanoflow Proxeon EASY-nLC 1200 UHPLC system (Thermo Fisher Scientific), and then analyzed with a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific). The remaining tissues' samples underwent online separation with a nanoAcquity M-Class UHPLC system (Waters), and analyzed with a Q Exactive HF mass spectrometer (Thermo Fisher Scientific).

### Proteomics data processing and normalization

For heart and liver, raw MS/MS data samples were processed by a Spectrum Mill (v.7.09.215) (Agilent Technologies). For the remaining tissues, sample processing was implemented by an in-house cloud-based proteomics pipeline executed in the Google Cloud Platform[5]. In all tissues, MS2 spectra were processed and searched against the rat RefSeq protein database (downloaded November 2018). $Log_2$ TMT ratios to the common reference were used as quantitative values for all proteins. Principal component analysis and median protein abundance across samples were used to find sample outliers. Proteomics features that were not fully quantified in at least two plexes within a tissue and non-rat contaminants were removed. Median-centering and mean absolute deviation scaling of $Log_2$ TMT ratios were done for sample normalization. Plex batch effects were removed using *limma::removeBatchEffect* function in R (v 3.48.0).

## Statistical analysis

### Differential analysis

Differential analyses were performed in each tissue of each ome. Males and females in one dataset were analyzed separately. Limma with empirical Bayes variance shrinkage was used for ATAC-seq and proteomics data[79]; DESeq2 was used for RNA-Seq[75]. For all proteomics and ATAC-seq data, the input for differential analysis was normalized as described above. For RNA-Seq, the input was filtered raw counts, in accordance with the DESeq2 workflow.

To select analytes that changed over the training time course, we performed F-tests (limma) or likelihood ratio tests (DESeq2::nbinomLRT, lrtest) to compare a full model with ome-specific technical covariates and training group as a factor variable (i.e. sedentary control, 1 week, 2 weeks, 4 weeks, 8 weeks) against a reduced model with only technical covariates. For each analyte, male- and female-specific p-values were combined using Fisher's sum of logs meta-analysis to provide a single p-value, referred to as the training p-value. To account for false discovery rate across all statistical tests, the training p-values were adjusted across all datasets within each ome using Independent Hypothesis Weighting (IHW) with tissue as a covariate[80]. Training-differential features were selected at 5% IHW FDR.

We used the contrasts of each training time point versus the sex-matched sedentary controls to calculate time- and sex-specific effect sizes, their variance, and their p-values (e.g., using linear F-tests), referred to as the timewise summary statistics. Specifically, for limma models we used

limma::contrasts.fit and limma::eBayes, for DESeq2 models we used DESeq2::DESeq. Covariates were selected from assay-specific technical metrics that explained variance in the data and were not correlated with exercise training: RNA integrity number (RIN), median 5'-3' bias, percent of reads mapping to globin, and percent of PCR duplicates as quantified with Unique Molecular Identifiers (UMIs) for RNA-Seq; fraction of reads in peaks and library preparation batch for ATAC-seq.

**Identification of transcription factor motifs using HOMER**

Transcription factor motif enrichment analysis was performed on sets of DARs and DEGaPs for each tissue. DARs for motif enrichment analysis were selected for each tissue by satisfying an adjusted p-value threshold of 0.1. Similarly, DEGaPs for each tissue were selected by isolating the DEGs that satisfied an adjusted p-value threshold of 0.1, and selecting peaks annotated to the DEGs that contained a median normalized accessibility of -1. For genomic feature-specific analysis, DEGaPs were divided based upon their gene region annotation. The analysis was carried out by findMotifsGenome.pl (HOMER v4.11.1)[81]. It was performed on the ±50 bp flanking regions of the peak summits. The search lengths of the motifs were 8, 10, and 12 bp. We applied the -find flag to generate a list of all known rat motifs contained within the ±50 bp flanking regions of the summits for each peak in the ATACseq dataset, using the same settings as above.

**DAR genomic feature TF motif enrichment analysis**

Applying the output from HOMER, the top ten significantly enriched TF motifs among DARs and DEGaPs in each tissue were selected for further downstream analysis and cross-tissue comparisons. TFs were removed from further analysis if their gene was not expressed in the tissue in which their motifs were enriched. TF motif enrichments for differentially accessible regions (DARs) divided into gene features were calculated using the Fisher test. The test compared the ratio of DARs containing the motif for a specific TF/non-DARs containing this motif in one genomic feature, and the ratio of DARs containing this motif / non-DARs containing this motif in other genomic features. p values were adjusted and FDR cutoff = 0.1 to select significant motifs in specific genomic features.

**Correlations between DARs and DEGs**

We selected DARs whose centers were within 500kb of a DEG TSS in each tissue. We then calculated the Pearson correlation of the L2FC of the DAR and the DEG for each sex at each week. We considered a DAR-DEG pair for further analysis if their training response Pearson correlation was greater than 0.5 or less than -0.5 and the DAR contained a known motif for a TF expressed within the tissue.

**Cell-type deconvolution**

Cell type deconvolution was conducted by the R package CellCODE[82] using the getallSPVs function. Marker sets were generated using the IRIS (Immune Response *In Silico*[83] and DMAP (Differentiation Map) reference datasets[84]. The Kruskal-wallis test was implemented with the R function *kruskal.test* to determine if the variability in cell type proportion across samples in a given tissue would suggest a significant training response or sex difference.

**Pathway enrichment**

Pathway enrichment analysis of up-regulated DEGs at each time point was performed using the R package gprofiler2:gost[85] against Gene Ontology Biological Process, Reactome, WikiPathways and KEGG databases. Top 10 pathway enrichments for each tissue are displayed as bubble plots with sizes

indicating the number of significant genes enriched in different pathways relative to the pathway size (number of genes in that pathway) and colors indicating the significance (BH-corrected p-value). At least 10 genes were required to be enriched in a pathway with a maximum of 200 genes.

**Correlations between Phenotypic Measures and DEGs**

Phenotypic measures were calculated at weeks 4 and 8 of EET and in week 8 controls. Measures were presented as changes between time point and original baseline measurements in each rat. For each phenotypic measure and DEG combination, we calculated the Pearson correlation between the change in phenotypic measure between baseline and a given time point, and the gene expression of the DEG at the time point for each animal subject. We isolated the DEGs that exhibited > 0.5 or < -0.5 correlations with each phenotypic measure in each tissue and selected the DEGaPs annotated to the promoter region of the DEGs. TF motif enrichment significance among a set of positively or negatively correlated DEG's promoter DEGaPs in a tissue were determined by an exact binomial test comparing the frequency of enrichment among phenotype-correlated DEGs versus general enrichment among the promoter DEGaPs in the tissue.

# Data Availability

MoTrPAC data will be publicly available at time of publication via motrpac-data.org/data-access. Data access inquiries should be sent to motrpac-helpdesk@lists.stanford.edu. Additional resources can be found at motrpac.org and motrpac-data.org.

# Code Availability

MoTrPAC data processing pipelines for RNA-Seq, ATAC-seq, RRBS, and proteomics will be made public at the time of publication: https://github.com/MoTrPAC/motrpac-rna-seq-pipeline, https://github.com/MoTrPAC/motrpac-atac-seq-pipeline, https://github.com/MoTrPAC/motrpacrrbs-pipeline, https://github.com/MoTrPAC/motrpac-proteomics-pipeline. Normalization and QC scripts will be available at https://github.com/MoTrPAC/motrpac-bic-norm-qc. Code for the underlying differential analysis for the manuscript will be provided in the MotrpacRatTraining6mo R package (motrpac.github.io/MotrpacRatTraining6mo). Code for conducting the analysis and generating the figures contained within this paper will be available at https://github.com/gsmith990306/MoTrPAC_PASS1B_Transcription_Factor_Paper.

# References


1. Neufer, P. D. *et al.* Understanding the Cellular and Molecular Mechanisms of Physical Activity-Induced Health Benefits. *Cell Metab.* **22**, 4–11 (2015).

2. Mansueto, G. *et al.* Transcription Factor EB Controls Metabolic Flexibility during Exercise. *Cell Metab.* **25**, 182–196 (2017).

3. Bléher, M. *et al.* Egr1 loss-of-function promotes beige adipocyte differentiation and activation specifically in inguinal subcutaneous white adipose tissue. *Sci. Rep.* **10**, 15842 (2020).
4. Popov, D. V. *et al.* Contractile activity-specific transcriptome response to acute endurance exercise and training in human skeletal muscle. *Am. J. Physiol.-Endocrinol. Metab.* **316**, E605–E614 (2019).
5. Group, M. S. *et al.* Temporal dynamics of the multi-omic response to endurance exercise training across tissues. 2022.09.21.508770 Preprint at https://doi.org/10.1101/2022.09.21.508770 (2022).
6. Oliveira, A. N. & Hood, D. A. Exercise is mitochondrial medicine for muscle. *Sports Med. Health Sci.* **1**, 11–18 (2019).
7. Gregoire, F. M., Smas, C. M. & Sul, H. S. Understanding Adipocyte Differentiation. *Physiol. Rev.* **78**, 783–809 (1998).
8. Moore, D. & Loprinzi, P. D. Exercise influences episodic memory via changes in hippocampal neurocircuitry and long-term potentiation. *Eur. J. Neurosci.* **54**, 6960–6971 (2021).
9. Sonawane, A. R. *et al.* Understanding Tissue-Specific Gene Regulation. *Cell Rep.* **21**, 1077–1088 (2017).
10. Ramachandran, K. *et al.* Dynamic enhancers control skeletal muscle identity and reprogramming. *PLOS Biol.* **17**, e3000467 (2019).
11. Miao, Z. *et al.* Single cell regulatory landscape of the mouse kidney highlights cellular differentiation programs and disease targets. *Nat. Commun.* **12**, 2277 (2021).
12. Yan, F., Powell, D. R., Curtis, D. J. & Wong, N. C. From reads to insight: a hitchhiker's guide to ATAC-seq data analysis. *Genome Biol.* **21**, 22 (2020).
13. Izzo, M. W., Strachan, G. D., Stubbs, M. C. & Hall, D. J. Transcriptional Repression from the c-myc P2 Promoter by the Zinc Finger Protein ZF87/MAZ *. *J. Biol. Chem.* **274**, 19498–19506 (1999).
14. Nair, V. D. *et al.* Differential analysis of chromatin accessibility and gene expression profiles identifies cis-regulatory elements in rat adipose and muscle. *Genomics* **113**, 3827–3841 (2021).

15. Panigrahi, A. & O'Malley, B. W. Mechanisms of enhancer action: the known and the unknown. *Genome Biol.* **22**, 108 (2021).
16. Maity, G. *et al.* The MAZ transcription factor is a downstream target of the oncoprotein Cyr61/CCN1 and promotes pancreatic cancer cell invasion via CRAF–ERK signaling. *J. Biol. Chem.* **293**, 4334–4349 (2018).
17. Bang, P. *et al.* Exercise-induced changes in insulin-like growth factors and their low molecular weight binding protein in healthy subjects and patients with growth hormone deficiency. *Eur. J. Clin. Invest.* **20**, 285–292 (1990).
18. Stein, A. M. *et al.* Physical exercise, IGF-1 and cognition A systematic review of experimental studies in the elderly. *Dement. Neuropsychol.* **12**, 114–122 (2018).
19. Jiao, S. *et al.* Differential regulation of IGF-I and IGF-II gene expression in skeletal muscle cells. *Mol. Cell. Biochem.* **373**, 107–113 (2013).
20. Clavarino, G. *et al.* Protein phosphatase 1 subunit Ppp1r15a/GADD34 regulates cytokine production in polyinosinic:polycytidylic acid-stimulated dendritic cells. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 3006–3011 (2012).
21. Farkas, C. *et al.* Characterization of SALL2 Gene Isoforms and Targets Across Cell Types Reveals Highly Conserved Networks. *Front. Genet.* **12**, 613808 (2021).
22. Amar, D. *et al.* Time trajectories in the transcriptomic response to exercise - a meta-analysis. *Nat. Commun.* **12**, 3471 (2021).
23. García-Pelagio, K. P., Bloch, R. J., Ortega, A. & González-Serratos, H. Biomechanics of the sarcolemma and costameres in single skeletal muscle fibers from normal and dystrophin-null mice. *J. Muscle Res. Cell Motil.* **31**, 323–336 (2011).
24. Himeda, C. L., Ranish, J. A., Pearson, R. C. M., Crossley, M. & Hauschka, S. D. KLF3 regulates muscle-specific gene expression and synergizes with serum response factor on KLF binding sites. *Mol. Cell. Biol.* **30**, 3430–3443 (2010).

25. Taylor, M. V. & Hughes, S. M. Mef2 and the skeletal muscle differentiation program. *Semin. Cell Dev. Biol.* **72**, 33–44 (2017).

26. Gallant, S. & Gilkeson, G. ETS transcription factors and regulation of immunity. *Arch. Immunol. Ther. Exp. (Warsz.)* **54**, 149–163 (2006).

27. Seifert, L. L. *et al.* The ETS transcription factor ELF1 regulates a broadly antiviral program distinct from the type I interferon response. *PLoS Pathog.* **15**, e1007634 (2019).

28. TGF-β/SMAD Pathway and Its Regulation in Hepatic Fibrosis - Fengyun Xu, Changwei Liu, Dandan Zhou, Lei Zhang, 2016. https://journals.sagepub.com/doi/10.1369/0022155415627681.

29. Latella, G. *et al.* Targeted disruption of Smad3 confers resistance to the development of dimethylnitrosamine-induced hepatic fibrosis in mice. *Liver Int.* **29**, 997–1009 (2009).

30. Cheng, Z. *et al.* Kallistatin, a new and reliable biomarker for the diagnosis of liver cirrhosis. *Acta Pharm. Sin. B* **5**, 194–200 (2015).

31. Simo, K. A. *et al.* Altered lysophosphatidic acid (LPA) receptor expression during hepatic regeneration in a mouse model of partial hepatectomy. *HPB* **16**, 534–542 (2014).

32. Choi, J. W. *et al.* LPA receptors: subtypes and biological actions. *Annu. Rev. Pharmacol. Toxicol.* **50**, 157–186 (2010).

33. Choi, W.-M. *et al.* Metabotropic Glutamate Receptor 5 in Natural Killer Cells Attenuates Liver Fibrosis by Exerting Cytotoxicity to Activated Stellate Cells. *Hepatology* **74**, 2170–2185 (2021).

34. Induction of GADD34 Is Necessary for dsRNA-Dependent Interferon-β Production and Participates in the Control of Chikungunya Virus Infection | PLOS Pathogens. https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1002708.

35. Du, L. *et al.* IGF-2 Preprograms Maturing Macrophages to Acquire Oxidative Phosphorylation-Dependent Anti-inflammatory Properties. *Cell Metab.* **29**, 1363-1375.e8 (2019).

36. De Giovanni, C. *et al.* Immune targeting of autocrine IGF2 hampers rhabdomyosarcoma growth and metastasis. *BMC Cancer* **19**, 126 (2019).

37. Frasca, F. *et al.* Insulin Receptor Isoform A, a Newly Recognized, High-Affinity Insulin-Like Growth Factor II Receptor in Fetal and Cancer Cells. *Mol. Cell. Biol.* **19**, 3278–3288 (1999).

38. Li, F., Xiao, H., Zhou, F., Hu, Z. & Yang, B. Study of HSPB6: Insights into the Properties of the Multifunctional Protective Agent. *Cell. Physiol. Biochem.* **44**, 314–332 (2017).

39. Bayly-Jones, C., Pang, S. S., Spicer, B. A., Whisstock, J. C. & Dunstone, M. A. Ancient but Not Forgotten: New Insights Into MPEG1, a Macrophage Perforin-Like Immune Effector. *Front. Immunol.* **11**, (2020).

40. Liao, X. *et al.* 2′, 5′-Oligoadenylate Synthetase 2 (OAS2) Inhibits Zika Virus Replication through Activation of Type I IFN Signaling Pathway. *Viruses* **12**, 418 (2020).

41. Klemann, C. *et al.* Clinical and Immunological Phenotype of Patients With Primary Immunodeficiency Due to Damaging Mutations in NFKB2. *Front. Immunol.* **10**, (2019).

42. Rogeri, P. S. *et al.* Crosstalk Between Skeletal Muscle and Immune System: Which Roles Do IL-6 and Glutamine Play? *Front. Physiol.* **11**, (2020).

43. Fernández-Verdejo, R. *et al.* Activating transcription factor 3 attenuates chemokine and cytokine expression in mouse skeletal muscle after exercise and facilitates molecular adaptation to endurance training. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **31**, 840–851 (2017).

44. Kawasaki, E. *et al.* Role of local muscle contractile activity in the exercise-induced increase in NR4A receptor mRNA expression. *J. Appl. Physiol. Bethesda Md 1985* **106**, 1826–1831 (2009).

45. Rundqvist, H. C. *et al.* Acute sprint exercise transcriptome in human skeletal muscle. *PLOS ONE* **14**, e0223024 (2019).

46. Puntschart, A. *et al.* Expression of fos andjun genes in human skeletal muscle after exercise. *Am. J. Physiol.-Cell Physiol.* **274**, C129–C137 (1998).

47. Ku, H.-C. & Cheng, C.-F. Master Regulator Activating Transcription Factor 3 (ATF3) in Metabolic Homeostasis and Cancer. *Front. Endocrinol.* **11**, (2020).

48. Zhang, L. *et al.* The Orphan Nuclear Receptor 4A1: A Potential New Therapeutic Target for Metabolic Diseases. *J. Diabetes Res.* **2018**, 9363461 (2018).

49. Pei, L., Castrillo, A. & Tontonoz, P. Regulation of Macrophage Inflammatory Gene Expression by the Orphan Nuclear Receptor Nur77. *Mol. Endocrinol.* **20**, 786–794 (2006).

50. Liebmann, M. *et al.* Nur77 serves as a molecular brake of the metabolic switch during T cell activation to restrict autoimmunity. *Proc. Natl. Acad. Sci.* **115**, E8017–E8026 (2018).

51. Mayer, M. P. & Bukau, B. Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell. Mol. Life Sci.* **62**, 670–684 (2005).

52. Cooper, L. M., West, R. C., Hayes, C. S. & Waddell, D. S. Dual-specificity phosphatase 29 is induced during neurogenic skeletal muscle atrophy and attenuates glucocorticoid receptor activity in muscle cell culture. *Am. J. Physiol.-Cell Physiol.* **319**, C441–C454 (2020).

53. Filtz, T. M., Vogel, W. K. & Leid, M. Regulation of transcription factor activity by interconnected post-translational modifications. *Trends Pharmacol. Sci.* **35**, 76–85 (2014).

54. Presnell, J. S., Schnitzler, C. E. & Browne, W. E. KLF/SP Transcription Factor Family Evolution: Expansion, Diversification, and Innovation in Eukaryotes. *Genome Biol. Evol.* **7**, 2289–2309 (2015).

55. Zhao, X. *et al.* Nuclear receptors rock around the clock. *EMBO Rep.* **15**, 518–528 (2014).

56. Oiwa, A. *et al.* Synergistic regulation of the mouse orphan nuclear receptor SHP gene promoter by CLOCK–BMAL1 and LRH-1. *Biochem. Biophys. Res. Commun.* **353**, 895–901 (2007).

57. Rey, G. *et al.* Genome-Wide and Phase-Specific DNA-Binding Rhythms of BMAL1 Control Circadian Output Functions in Mouse Liver. *PLOS Biol.* **9**, e1000595 (2011).

58. Sato, S. *et al.* Atlas of exercise metabolism reveals time-dependent signatures of metabolic homeostasis. *Cell Metab.* **34**, 329-345.e8 (2022).

59. Kemler, D., Wolff, C. A. & Esser, K. A. Time of Day Dependent Effects of Contractile Activity on the Phase of the Skeletal Muscle Clock. 2020.03.05.978759 Preprint at https://doi.org/10.1101/2020.03.05.978759 (2020).

60. Casanova-Vallve, N. *et al.* Daily running enhances molecular and physiological circadian rhythms in skeletal muscle. *Mol. Metab.* **61**, 101504 (2022).

61. Silva, B. S. de A. *et al.* Exercise as a Peripheral Circadian Clock Resynchronizer in Vascular and Skeletal Muscle Aging. *Int. J. Environ. Res. Public. Health* **18**, 12949 (2021).

62. Mansingh, S. & Handschin, C. Time to Train: The Involvement of the Molecular Clock in Exercise Adaptation of Skeletal Muscle. *Front. Physiol.* **13**, (2022).

63. Duglan, D. & Lamia, K. A. Clocking In, Working Out: Circadian Regulation of Exercise Physiology. *Trends Endocrinol. Metab.* **30**, 347–356 (2019).

64. Wisløff, U., Helgerud, J., Kemi, O. J. & Ellingsen, O. Intensity-controlled treadmill running in rats: VO(2 max) and cardiac hypertrophy. *Am. J. Physiol. Heart Circ. Physiol.* **280**, H1301-1310 (2001).

65. Corces, M. R. *et al.* An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).

66. Moore, J. E. *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* **583**, 699–710 (2020).

67. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

68. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

69. Gaspar, J. M. Improved peak-calling with MACS2. 496521 Preprint at https://doi.org/10.1101/496521 (2018).

70. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).

71. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.* **26**, 841–842 (2010).

72. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).

73. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).

74. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinforma. Oxf. Engl.* **32**, 3047–3048 (2016).

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

76. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinforma. Oxf. Engl.* **26**, 139–140 (2010).

77. Mertins, P. *et al.* Reproducible workflow for multiplexed deep-scale proteome and phosphoproteome analysis of tumor tissues by liquid chromatography–mass spectrometry. *Nat. Protoc.* **13**, 1632–1661 (2018).

78. Zecha, J. *et al.* TMT Labeling for the Masses: A Robust and Cost-efficient, In-solution Labeling Approach *[S]. *Mol. Cell. Proteomics* **18**, 1468–1478 (2019).

79. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).

80. Ignatiadis, N., Klaus, B., Zaugg, J. B. & Huber, W. Data-driven hypothesis weighting increases detection power in genome-scale multiple testing. *Nat. Methods* **13**, 577–580 (2016).

81. Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).

82. Chikina, M., Zaslavsky, E. & Sealfon, S. C. CellCODE: a robust latent variable approach to differential expression analysis for heterogeneous cell populations. *Bioinformatics* **31**, 1584–1591 (2015).

83. Abbas, A. R., Wolslegel, K., Seshasayee, D., Modrusan, Z. & Clark, H. F. Deconvolution of Blood Microarray Data Identifies Cellular Activation Patterns in Systemic Lupus Erythematosus. *PLOS ONE* **4**, e6098 (2009).

84. Novershtern, N. *et al.* Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis. *Cell* **144**, 296–309 (2011).

85. Raudvere, U. *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* **47**, W191–W198 (2019).

# Acknowledgements

**Funding:** The MoTrPAC Study is supported by NIH grants U24OD026629 (Bioinformatics Center), U24DK112349, U24DK112342, U24DK112340, U24DK112341, U24DK112326, U24DK112331, U24DK112348 (Chemical Analysis Sites), U01AR071133, U01AR071130, U01AR071124–01, U01AR071128, U01AR071150, U01AR071160, U01AR071158 (Clinical Centers), U24AR071113 (Consortium Coordinating Center), U01AG055133, U01AG055137 and U01AG055135 (PASS/Animal Sites).

Parts of this work were performed in the Environmental Molecular Science Laboratory, a U.S. Department of Energy national scientific user facility at Pacific Northwest National Laboratory in Richland, Washington.

# Author Contributions

Gregory R. Smith, supervised the investigation, performed data and bioinformatics analysis, prepared figures, drafted and revised the original manuscript
Bingqing Zhao, supervised the investigation, performed data and bioinformatics analysis, prepared figures, drafted and revised the original manuscript
Malene E. Lindholm, performed data and bioinformatics analysis, prepared figures, drafted and revised the original manuscript
Archana Raja, performed data and bioinformatics analysis, prepared figures, revised the original manuscript
Mark Viggars, contributed to data interpretation, drafted and revised the original manuscript
Hanna Pincas, contributed to data interpretation, revised the original manuscript
Nicole R. Gay, performed data and bioinformatics analysis, revised the original manuscript
Yifei Sun, performed and analyzed experiments
Yongchao Ge, performed data and bioinformatics analysis

Venugopalan D. Nair, contributed to data interpretation
James A. Sanford, contributed to data interpretation, revised the original manuscript
Mary Anne S. Amper, performed and analyzed experiments
Mital Vasoya, performed and analyzed experiments
Kevin S. Smith, contributed to data interpretation
Stephen Montgomery, contributed to data interpretation, revised the original manuscript
Elena Zaslavsky, contributed to data interpretation
Sue C. Bodine, contributed to data interpretation
Karyn A. Esser, contributed to data interpretation, drafted and revised the original manuscript
Martin J. Walsh, contributed to data interpretation, revised the original manuscript
Michael P. Snyder, provided overall leadership and guidance to the investigation
Stuart C. Sealfon, provided overall leadership and guidance to the investigation

# Competing Interests

S.C.B. has equity in Emmyon, Inc and receives grant funding from Calico Life Sciences. M.P.S. is a cofounder and scientific advisor of Personalis, Qbio, January AI, Filtricine, SensOmics, Protos, Fodsel, Rthm, Marble and scientific advisor of Genapsys, Swaz, Jupiter. S.B.M. is a consultant for BioMarin, MyOme and Tenaya Therapeutics.

# Supplementary Information

Supplementary Figures. Supplementary Figures S1-S16 and corresponding legends.

Supplementary Table S1. DAR-DEG pairs with Pearson correlation > 0.5 or < -0.5 and annotated with an expressed TF motif. For each pair, we listed the tissue, DEG ensembl id and symbol, DAR id, distance between DAR midpoint and DEG TSS, Pearson correlation, known TF motifs within the DAR, and the L2FC of the DEG and the DAR for each time point relative to eight week control in both female and male samples. Brown Adipose DAR-DEG pairs are excluded because of confounding cell type composition effects.

MoTrPAC Study Group Author List. Extended author list of MoTrPAC consortium members.

# Figures and Legends

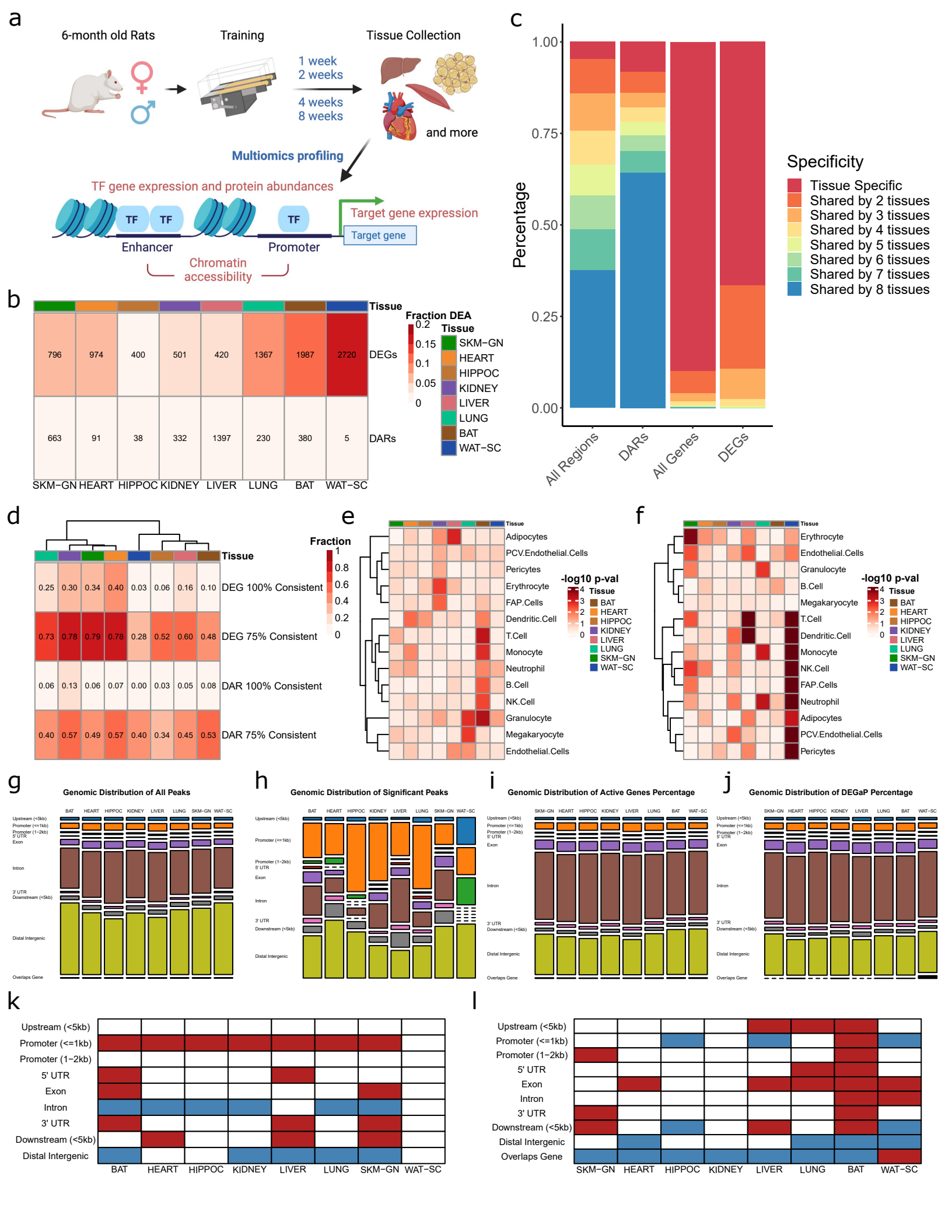
a
6-month old Rats
Training
Tissue Collection
1 week
2 weeks
4 weeks
8 weeks
and more
Multiomics profiling
TF gene expression and protein abundances
TF
TF
TF
Target gene expression
Target gene
Enhancer
Promoter
Chromatin accessibility
b
Tissue
796 974 400 501 420 1367 1987 2720 DEGs
663 91 38 332 1397 230 380 5 DARs
SKM−GN HEART HIPPOC KIDNEY LIVER LUNG BAT WAT−SC
Fraction DEA
0.2
0.15
0.1
0.05
0
Tissue
SKM−GN
HEART
HIPPOC
KIDNEY
LIVER
LUNG
BAT
WAT−SC
c
Percentage
1.00
0.75
0.50
0.25
0.00
All Regions
DARs
All Genes
DEGs
Specificity
Tissue Specific
Shared by 2 tissues
Shared by 3 tissues
Shared by 4 tissues
Shared by 5 tissues
Shared by 6 tissues
Shared by 7 tissues
Shared by 8 tissues
d
Tissue
Fraction
1
0.8
0.6
0.4
0.2
0
0.25 0.30 0.34 0.40 0.03 0.06 0.16 0.10 DEG 100% Consistent
0.73 0.78 0.79 0.78 0.28 0.52 0.60 0.48 DEG 75% Consistent
0.06 0.13 0.06 0.07 0.00 0.03 0.05 0.08 DAR 100% Consistent
0.40 0.57 0.49 0.57 0.40 0.34 0.45 0.53 DAR 75% Consistent
e
Tissue
Adipocytes
PCV.Endothelial.Cells
Pericytes
Erythrocyte
FAP.Cells
Dendritic.Cell
T.Cell
Monocyte
Neutrophil
B.Cell
NK.Cell
Granulocyte
Megakaryocyte
Endothelial.Cells
-log10 p-val
4
3
2
1
0
Tissue
BAT
HEART
HIPPOC
KIDNEY
LIVER
LUNG
SKM−GN
WAT−SC
f
Tissue
Erythrocyte
Endothelial.Cells
Granulocyte
B.Cell
Megakaryocyte
T.Cell
Dendritic.Cell
Monocyte
NK.Cell
FAP.Cells
Neutrophil
Adipocytes
PCV.Endothelial.Cells
Pericytes
-log10 p-val
4
3
2
1
0
Tissue
BAT
HEART
HIPPOC
KIDNEY
LIVER
LUNG
SKM−GN
WAT−SC
g
Genomic Distribution of All Peaks
BAT HEART HIPPOC KIDNEY LIVER LUNG SKM−GN WAT−SC
Upstream (<5kb)
Promoter (<=1kb)
Promoter (1−2kb)
5' UTR
Exon
Intron
3' UTR
Downstream (<5kb)
Distal Intergenic
Overlaps Gene
h
Genomic Distribution of Significant Peaks
BAT HEART HIPPOC KIDNEY LIVER LUNG SKM−GN WAT−SC
Upstream (<5kb)
Promoter (<=1kb)
Promoter (1−2kb)
5' UTR
Exon
Intron
3' UTR
Downstream (<5kb)
Distal Intergenic
i
Genomic Distribution of Active Genes Percentage
SKM−GN HEART HIPPOC KIDNEY LIVER LUNG BAT WAT−SC
Upstream (<5kb)
Promoter (<=1kb)
Promoter (1−2kb)
5' UTR
Exon
Intron
3' UTR
Downstream (<5kb)
Distal Intergenic
Overlaps Gene
j
Genomic Distribution of DEGaP Percentage
SKM−GN HEART HIPPOC KIDNEY LIVER LUNG BAT WAT−SC
Upstream (<5kb)
Promoter (<=1kb)
Promoter (1−2kb)
5' UTR
Exon
Intron
3' UTR
Downstream (<5kb)
Distal Intergenic
Overlaps Gene
k
Upstream (<5kb)
Promoter (<=1kb)
Promoter (1−2kb)
5' UTR
Exon
Intron
3' UTR
Downstream (<5kb)
Distal Intergenic
BAT HEART HIPPOC KIDNEY LIVER LUNG SKM−GN WAT−SC
l
Upstream (<5kb)
Promoter (<=1kb)
Promoter (1−2kb)
5' UTR
Exon
Intron
3' UTR
Downstream (<5kb)
Distal Intergenic
Overlaps Gene
SKM−GN HEART HIPPOC KIDNEY LIVER LUNG BAT WAT−SC

Figure 1: Epigenetic and transcriptional responses to training programs. (a) 6-month old rats of both sexes underwent training programs. Tissues were collected and subjected to multiomics profiling, including ATAC-seq, RNA-seq, proteomics. (b) Higher number and percentage of differentially expressed genes (DEGs) were identified than differentially accessible regions (DARs) after training in most tissues (F test FDR<0.1). (c) Many accessible regions and genes were identified in all tissues and training-induced features were highly tissue-specific. (d) Frequency of differential genes and analytes L2FC direction consistent in 100% (8) or 75% (6) sex and week-specific data points for each tissue. Training-induced responses were more consistent between genders and time points in DEGs than in DARs, highest in heart, SKM-GN and kidney. (e,f) Cell type deconvolution analysis-generated -log10 p-values of Kruskal-Wallis test measuring significant predicted changes in tissue cell type composition based on training (e) or sex (f). Brown adipose exhibited increased proportions of immune cell types after training. White adipose exhibited sex-specific changes in proportions of immune cell types and pericytes. (g-j) Distribution of genomic locations of all accessible regions (g), DARs (h), accessible regions in all active genes (i) and accessible regions in DEGs (DEGaPs) (j). (k) DARs enriched for the proximal promoter compared to all accessible regions. (l) DEGaPs enriched for exon and contained fewer distal intergenic features in some tissues.

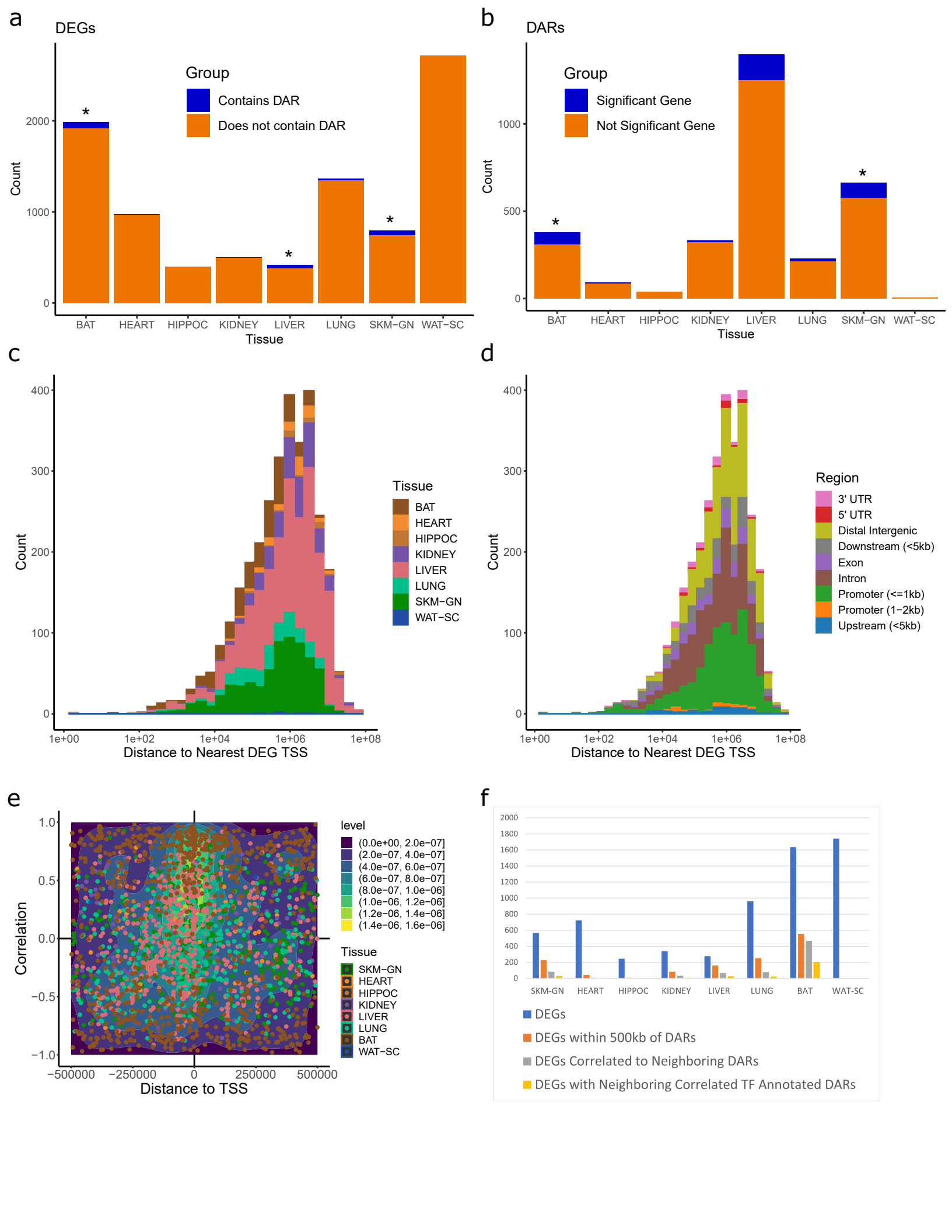
a
DEGs
Group
Contains DAR
Does not contain DAR
Count
Tissue
BAT
HEART
HIPPOC
KIDNEY
LIVER
LUNG
SKM−GN
WAT−SC
b
DARs
Group
Significant Gene
Not Significant Gene
c
Tissue
Distance to Nearest DEG TSS
d
Region
3' UTR
5' UTR
Distal Intergenic
Downstream (<5kb)
Exon
Intron
Promoter (<=1kb)
Promoter (1−2kb)
Upstream (<5kb)
e
level
Correlation
Distance to TSS
f
DEGs
DEGs within 500kb of DARs
DEGs Correlated to Neighboring DARs
DEGs with Neighboring Correlated TF Annotated DARs

Figure 2: Responses in chromatin accessibility may not directly link with the expression of the closest genes. (a) Count of DARs annotated to a DEG or a non-DEG. The closest gene to most DARs is not a DEG. * reflects $p < 0.05$ for hypergeometric test measuring the significance of the DAR-DEG overlap. (b) Count of DEGs with an annotated DAR. Most DEGs did not contain DARs in their vicinity. * reflects $p < 0.05$ for hypergeometric test measuring the significance of the DAR-DEG overlap (c-d) Distribution of distance between DARs and nearest DEG TSS. DARs colored by tissue (c) and genomic feature (d). (e) Density scatter plot of DAR-DEG training response correlation vs. distance. DARs with high correlation to gene expression enriched for TSS-proximal regions in most tissues. (f) Subsets of tissue DEGs with DARs located within 500 kb, DEGs correlated with within-500 kb DARs, and DEGs correlated with within-500 kb DARs containing annotated TF motifs.

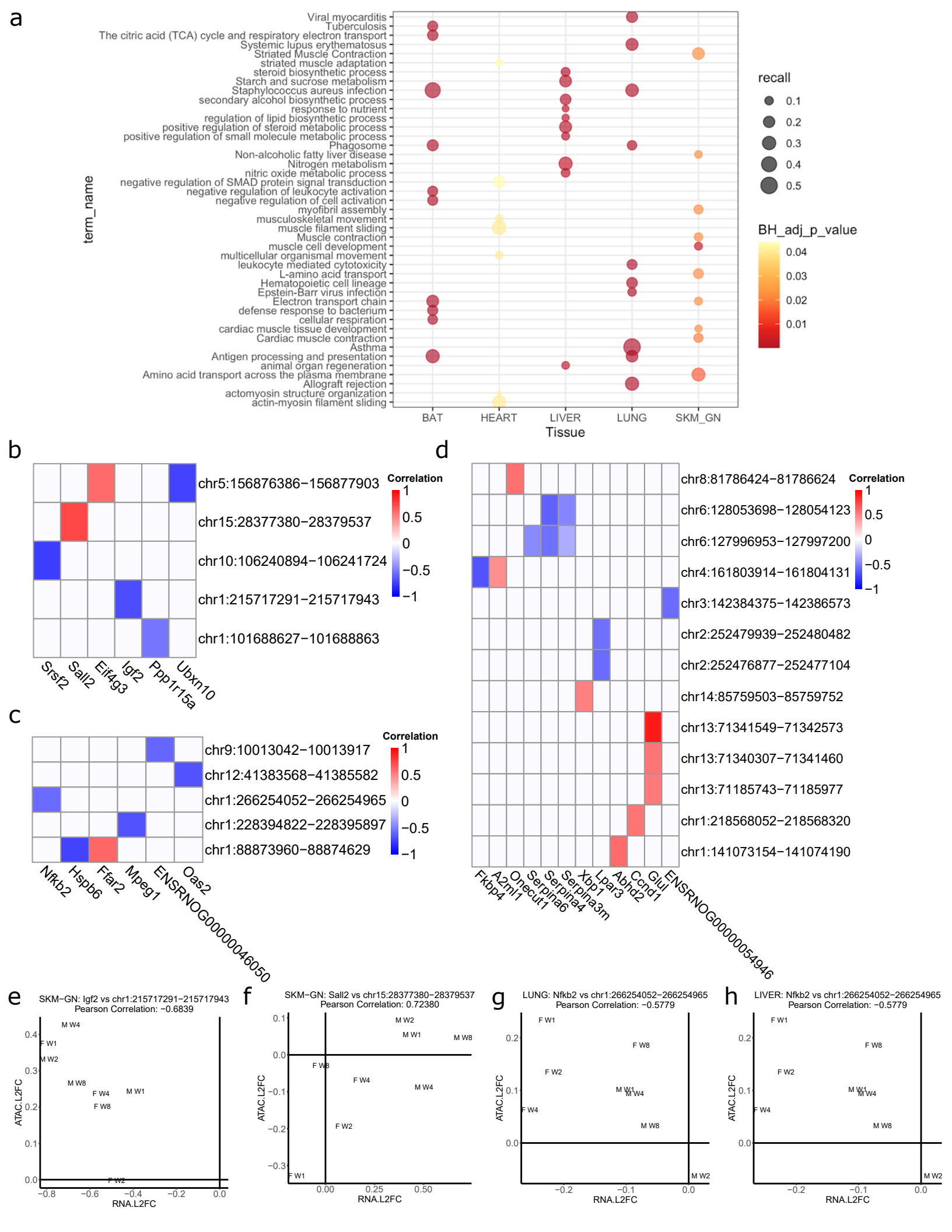
a
Viral myocarditis
Tuberculosis
The citric acid (TCA) cycle and respiratory electron transport
Systemic lupus erythematosus
Striated Muscle Contraction
striated muscle adaptation
steroid biosynthetic process
Starch and sucrose metabolism
Staphylococcus aureus infection
secondary alcohol biosynthetic process
response to nutrient
regulation of lipid biosynthetic process
positive regulation of steroid metabolic process
positive regulation of small molecule metabolic process
Phagosome
Non-alcoholic fatty liver disease
Nitrogen metabolism
nitric oxide metabolic process
negative regulation of SMAD protein signal transduction
negative regulation of leukocyte activation
negative regulation of cell activation
myofibril assembly
musculoskeletal movement
muscle filament sliding
Muscle contraction
muscle cell development
multicellular organismal movement
leukocyte mediated cytotoxicity
L-amino acid transport
Hematopoietic cell lineage
Epstein-Barr virus infection
Electron transport chain
defense response to bacterium
cellular respiration
cardiac muscle tissue development
Cardiac muscle contraction
Asthma
Antigen processing and presentation
animal organ regeneration
Amino acid transport across the plasma membrane
Allograft rejection
actomyosin structure organization
actin-myosin filament sliding
term_name
BAT
HEART
LIVER
LUNG
SKM_GN
Tissue
recall
0.1
0.2
0.3
0.4
0.5
BH_adj_p_value
0.04
0.03
0.02
0.01
b
chr5:156876386−156877903
chr15:28377380−28379537
chr10:106240894−106241724
chr1:215717291−215717943
chr1:101688627−101688863
Srsf2
Sall2
Eif4g3
Igf2
Ppp1r15a
Ubxn10
Correlation
1
0.5
0
−0.5
−1
c
chr9:10013042−10013917
chr12:41383568−41385582
chr1:266254052−266254965
chr1:228394822−228395897
chr1:88873960−88874629
Nfkb2
Hspb6
Ffar2
Mpeg1
ENSRNOG00000046050
Oas2
Correlation
1
0.5
0
−0.5
−1
d
chr8:81786424−81786624
chr6:128053698−128054123
chr6:127996953−127997200
chr4:161803914−161804131
chr3:142384375−142386573
chr2:252479939−252480482
chr2:252476877−252477104
chr14:85759503−85759752
chr13:71341549−71342573
chr13:71340307−71341460
chr13:71185743−71185977
chr1:218568052−218568320
chr1:141073154−141074190
Fkbp4
A2ml1
Onecut1
Serpina6
Serpina4
Serpina3m
Xbp1
Lpar3
Abhd2
Ccnd1
Glul
ENSRNOG00000054946
Correlation
1
0.5
0
−0.5
−1
e
SKM−GN: Igf2 vs chr1:215717291−215717943
Pearson Correlation: −0.6839
ATAC.L2FC
RNA.L2FC
f
SKM−GN: Sall2 vs chr15:28377380−28379537
Pearson Correlation: 0.72380
ATAC.L2FC
RNA.L2FC
g
LUNG: Nfkb2 vs chr1:266254052−266254965
Pearson Correlation: −0.5779
ATAC.L2FC
RNA.L2FC
h
LIVER: Nfkb2 vs chr1:266254052−266254965
Pearson Correlation: −0.5779
ATAC.L2FC
RNA.L2FC

Figure 3: Enriched pathway and responsive transcription factors in correlated DAR-DEG pairs. (a) Per tissue pathway enrichment analysis of DEGs with correlated DARs located within 500 kb . (b-d) Correlation heatmaps of correlated DAR-DEG pairs with binding sites for enriched TFs Maz in SKM-GN (b), and lung (c), and Smad3 in liver (d). (e-h) Training response L2FC scatter plots of correlated DAR-DEG Maz and Smad3 targets. Maz-target DARs in SKM-GN correlated with *Igf2* (e), and *Sall2* (f), and in lung correlated with *Nfkb2* (g). SMAD3-target DAR in liver correlated with Fkbp4 (h).

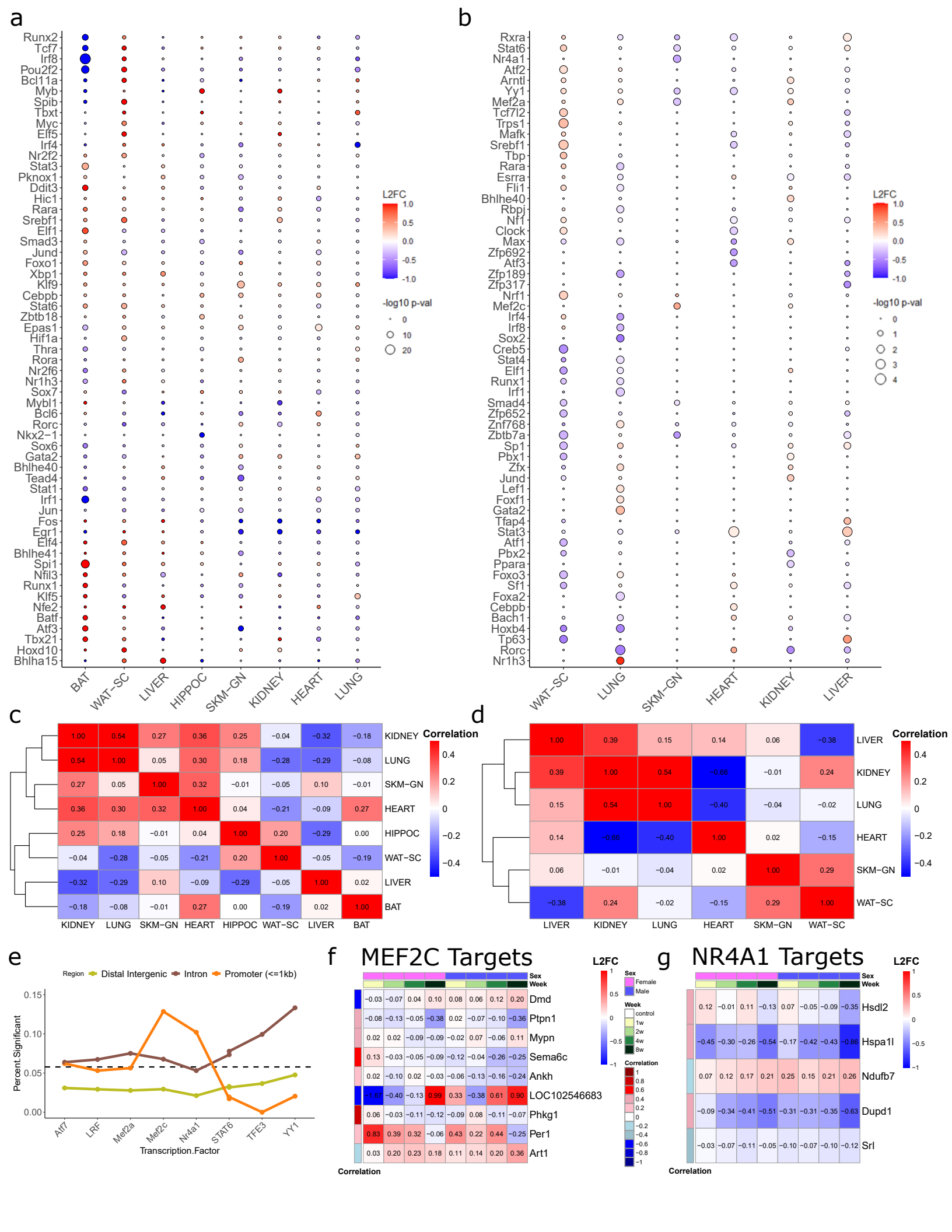
a
Runx2
Tcf7
Irf8
Pou2f2
Bcl11a
Myb
Spib
Tbxt
Myc
Elf5
Irf4
Nr2f2
Stat3
Pknox1
Ddit3
Hic1
Rara
Srebf1
Elf1
Smad3
Jund
Foxo1
Xbp1
Klf9
Cebpb
Stat6
Zbtb18
Epas1
Hif1a
Thra
Rora
Nr2f6
Nr1h3
Sox7
Mybl1
Bcl6
Rorc
Nkx2−1
Sox6
Gata2
Bhlhe40
Tead4
Stat1
Irf1
Jun
Fos
Egr1
Elf4
Bhlhe41
Spi1
Nfil3
Runx1
Klf5
Nfe2
Batf
Atf3
Tbx21
Hoxd10
Bhlha15
BAT
WAT−SC
LIVER
HIPPOC
SKM−GN
KIDNEY
HEART
LUNG
L2FC
1.0
0.5
0.0
−0.5
−1.0
-log10 p-val
0
10
20
b
Rxra
Stat6
Nr4a1
Atf2
Arntl
Yy1
Mef2a
Tcf7l2
Trps1
Mafk
Srebf1
Tbp
Rara
Esrra
Fli1
Bhlhe40
Rbpj
Nf1
Clock
Max
Zfp692
Atf3
Zfp189
Zfp317
Nrf1
Mef2c
Irf4
Irf8
Sox2
Creb5
Stat4
Elf1
Runx1
Irf1
Smad4
Zfp652
Znf768
Zbtb7a
Sp1
Pbx1
Zfx
Jund
Lef1
Foxf1
Gata2
Tfap4
Stat3
Atf1
Pbx2
Ppara
Foxo3
Sf1
Foxa2
Cebpb
Bach1
Hoxb4
Tp63
Rorc
Nr1h3
WAT−SC
LUNG
SKM−GN
HEART
KIDNEY
LIVER
L2FC
1.0
0.5
0.0
−0.5
−1.0
-log10 p-val
0
1
2
3
4
c
Correlation
0.4
0.2
0
−0.2
−0.4
KIDNEY 1.00 0.54 0.27 0.36 0.25 −0.04 −0.32 −0.18
LUNG 0.54 1.00 0.05 0.30 0.18 −0.28 −0.29 −0.08
SKM−GN 0.27 0.05 1.00 0.32 −0.01 −0.05 0.10 −0.01
HEART 0.36 0.30 0.32 1.00 0.04 −0.21 −0.09 0.27
HIPPOC 0.25 0.18 −0.01 0.04 1.00 0.20 −0.29 0.00
WAT−SC −0.04 −0.28 −0.05 −0.21 0.20 1.00 −0.05 −0.19
LIVER −0.32 −0.29 0.10 −0.09 −0.29 −0.05 1.00 0.02
BAT −0.18 −0.08 −0.01 0.27 0.00 −0.19 0.02 1.00
KIDNEY LUNG SKM−GN HEART HIPPOC WAT−SC LIVER BAT
d
Correlation
0.4
0.2
0
−0.2
−0.4
LIVER 1.00 0.39 0.15 0.14 0.06 −0.38
KIDNEY 0.39 1.00 0.54 −0.66 −0.01 0.24
LUNG 0.15 0.54 1.00 −0.40 −0.04 −0.02
HEART 0.14 −0.66 −0.40 1.00 0.02 −0.15
SKM−GN 0.06 −0.01 −0.04 0.02 1.00 0.29
WAT−SC −0.38 0.24 −0.02 −0.15 0.29 1.00
LIVER KIDNEY LUNG HEART SKM−GN WAT−SC
e
Region
Distal Intergenic
Intron
Promoter (<=1kb)
Percent.Significant
0.15
0.10
0.05
0.00
Atf7
LRF
Mef2a
Mef2c
Nr4a1
STAT6
TFE3
YY1
Transcription.Factor
f
MEF2C Targets
Sex
Week
L2FC
1
0.5
0
−0.5
−1
Sex
Female
Male
Week
control
1w
2w
4w
8w
Correlation
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
Dmd −0.03 −0.07 0.04 0.10 0.08 0.06 0.12 0.20
Ptpn1 −0.08 −0.13 −0.05 −0.38 0.02 −0.07 −0.10 −0.36
Mypn 0.02 0.02 −0.09 −0.09 −0.02 0.07 −0.06 0.11
Sema6c 0.13 −0.03 −0.05 −0.09 −0.12 −0.04 −0.26 −0.25
Ankh 0.02 −0.10 0.02 −0.03 −0.06 −0.13 −0.16 −0.24
LOC102546683 −1.67 −0.40 −0.13 0.99 0.33 −0.38 0.61 0.90
Phkg1 0.06 −0.03 −0.11 −0.12 −0.09 0.08 −0.11 −0.07
Per1 0.83 0.39 0.32 −0.06 0.43 0.22 0.44 −0.25
Art1 0.03 0.20 0.23 0.18 0.11 0.14 0.20 0.36
Correlation
g
NR4A1 Targets
Sex
Week
L2FC
1
0.5
0
−0.5
−1
Hsdl2 0.12 −0.01 0.11 −0.13 0.07 −0.05 −0.09 −0.35
Hspa1l −0.45 −0.30 −0.26 −0.54 −0.17 −0.42 −0.43 −0.86
Ndufb7 0.07 0.12 0.17 0.21 0.25 0.15 0.21 0.26
Dupd1 −0.09 −0.34 −0.41 −0.51 −0.31 −0.31 −0.35 −0.63
Srl −0.03 −0.07 −0.11 −0.05 −0.10 −0.07 −0.10 −0.12
Correlation

Figure 4: TFs showed significant EET responses at multiple omic levels. (a-b) Significant TF training responses at the transcriptomic (a) and proteomic (b) levels. Adjusted p-values generated by the F test of differential analysis and highest magnitude L2FC across sexes and time points are displayed for each TF. (c-d) Pearson correlation of cross-tissue highest magnitude L2FC for each TF at the transcriptomic (c) and proteomic (d) levels. Kidney and lung shared the most similar TF training responses. (e) Frequency of DEGs among gene targets of TFs with significant proteomic training response in skeletal muscle. The dashed line represents the frequency of DEGs among active genes in skeletal muscle. Gene targets are divided by presence of TF motif within active peaks annotated to the proximal promoter (orange), intron (brown) or distal intergenic (gold) features. (f,g) L2FC training response of DEG proximal promoter targets of MEF2C (f) and NR4A1 (NUR77) (g). Pearson correlation between target DEG L2FC and TF protein L2FC is annotated for each row.

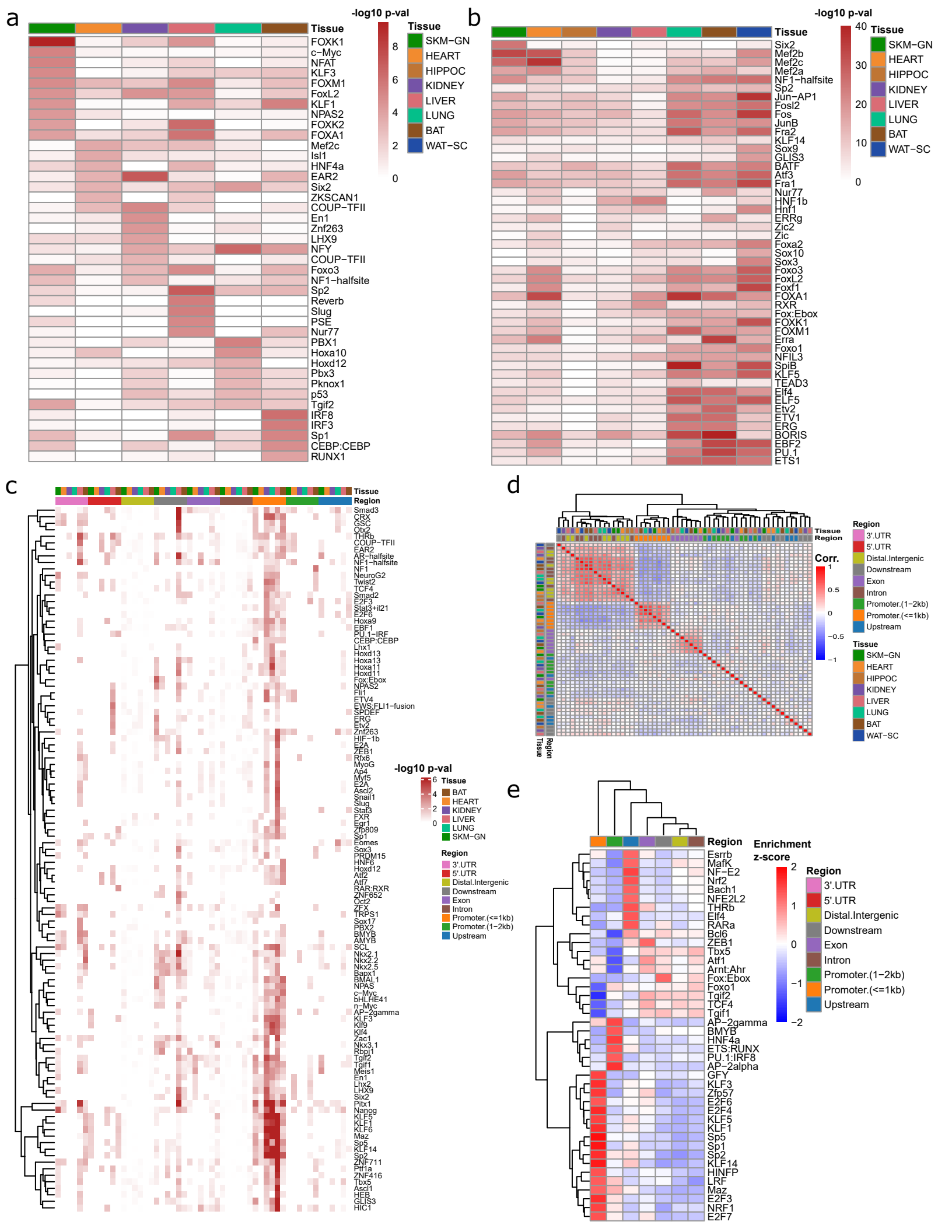
a
-log10 p-val
Tissue
SKM−GN
HEART
HIPPOC
KIDNEY
LIVER
LUNG
BAT
WAT−SC
FOXK1
c−Myc
NFAT
KLF3
FOXM1
FoxL2
KLF1
NPAS2
FOXK2
FOXA1
Mef2c
Isl1
HNF4a
EAR2
Six2
ZKSCAN1
COUP−TFII
En1
Znf263
LHX9
NFY
COUP−TFII
Foxo3
NF1−halfsite
Sp2
Reverb
Slug
PSE
Nur77
PBX1
Hoxa10
Hoxd12
Pbx3
Pknox1
p53
Tgif2
IRF8
IRF3
Sp1
CEBP:CEBP
RUNX1
b
-log10 p-val
Tissue
SKM−GN
HEART
HIPPOC
KIDNEY
LIVER
LUNG
BAT
WAT−SC
Six2
Mef2b
Mef2c
Mef2a
NF1−halfsite
Sp2
Jun−AP1
Fosl2
Fos
JunB
Fra2
KLF14
Sox9
GLIS3
BATF
Atf3
Fra1
Nur77
HNF1b
Hnf1
ERRg
Zic2
Zic
Foxa2
Sox10
Sox3
Foxo3
FoxL2
Foxf1
FOXA1
RXR
Fox:Ebox
FOXK1
FOXM1
Erra
Foxo1
NFIL3
SpiB
KLF5
TEAD3
Elf4
ELF5
Etv2
ETV1
ERG
BORIS
EBF2
PU.1
ETS1
c
Tissue
Region
Smad3
CRX
GSC
Otx2
THRb
COUP−TFII
EAR2
AR−halfsite
NF1−halfsite
NF1
NeuroG2
Twist2
TCF4
Smad2
E2F3
Stat3+il21
E2F6
Hoxa9
EBF1
PU.1−IRF
CEBP:CEBP
Lhx1
Hoxd13
Hoxa13
Hoxa11
Hoxd11
Fox:Ebox
NPAS2
Fli1
ETV4
EWS:FLI1−fusion
SPDEF
ERG
Etv2
Znf263
HIF−1b
E2A
ZEB1
Rfx6
MyoG
Ap4
Myf5
E2A
Ascl2
Snail1
Slug
Stat3
FXR
Egr1
Zfp809
Sp1
Eomes
Sox3
PRDM15
HNF6
Hoxd12
Atf2
Atf7
RAR:RXR
ZNF652
Oct2
ZFX
TRPS1
Sox17
PBX2
BMYB
AMYB
SCL
Nkx2.1
Nkx2.2
Nkx2.5
Bapx1
BMAL1
NPAS
c−Myc
bHLHE41
n−Myc
AP−2gamma
KLF3
Klf9
Klf4
Zac1
Nkx3.1
Rbpj1
Tgif2
Tgif1
Meis1
En1
Lhx2
LHX9
Six2
Pitx1
Nanog
KLF5
KLF1
KLF6
Maz
Sp5
KLF14
Sp2
ZNF711
Ptf1a
ZNF416
Tbx5
Ascl1
HEB
GLIS3
HIC1
-log10 p-val
Tissue
BAT
HEART
KIDNEY
LIVER
LUNG
SKM−GN
Region
3'.UTR
5'.UTR
Distal.Intergenic
Downstream
Exon
Intron
Promoter.(<=1kb)
Promoter.(1−2kb)
Upstream
d
Tissue
Region
Corr.
Region
3'.UTR
5'.UTR
Distal.Intergenic
Downstream
Exon
Intron
Promoter.(1−2kb)
Promoter.(<=1kb)
Upstream
Tissue
SKM−GN
HEART
HIPPOC
KIDNEY
LIVER
LUNG
BAT
WAT−SC
e
Region
Enrichment
z-score
Esrrb
MafK
NF−E2
Nrf2
Bach1
NFE2L2
THRb
Elf4
RARa
Bcl6
ZEB1
Tbx5
Atf1
Arnt:Ahr
Fox:Ebox
Foxo1
Tgif2
TCF4
Tgif1
AP−2gamma
BMYB
HNF4a
ETS:RUNX
PU.1:IRF8
AP−2alpha
GFY
KLF3
Zfp57
E2F6
E2F4
KLF5
KLF1
Sp5
Sp1
Sp2
KLF14
HINFP
LRF
Maz
E2F3
NRF1
E2F7
Region
3'.UTR
5'.UTR
Distal.Intergenic
Downstream
Exon
Intron
Promoter.(1−2kb)
Promoter.(<=1kb)
Upstream

Figure 5: TF motif enrichment at the tissue level and at the genomic feature level in each tissue. (a-b) -log10 p-value of TF motif enrichment in tissue DARs (a) and DEGaPs (b). Top enriched TFs were selected for each tissue in (a) and (b). (c) -log10 p-value of TF motif enrichment among sets of DARs split by genomic feature. TF motifs significantly enriched in proximal promoter regions (<=1kb from TSS) were shared by multiple tissues. (d) Correlation between TF motif enrichment of tissue DEGaPs split by genomic feature. TF motif enrichment in proximal promoter DEGaPs among all tissues were highly correlated. (e) Most differentially enriched TFs between genomic feature-classified DEGaPs.

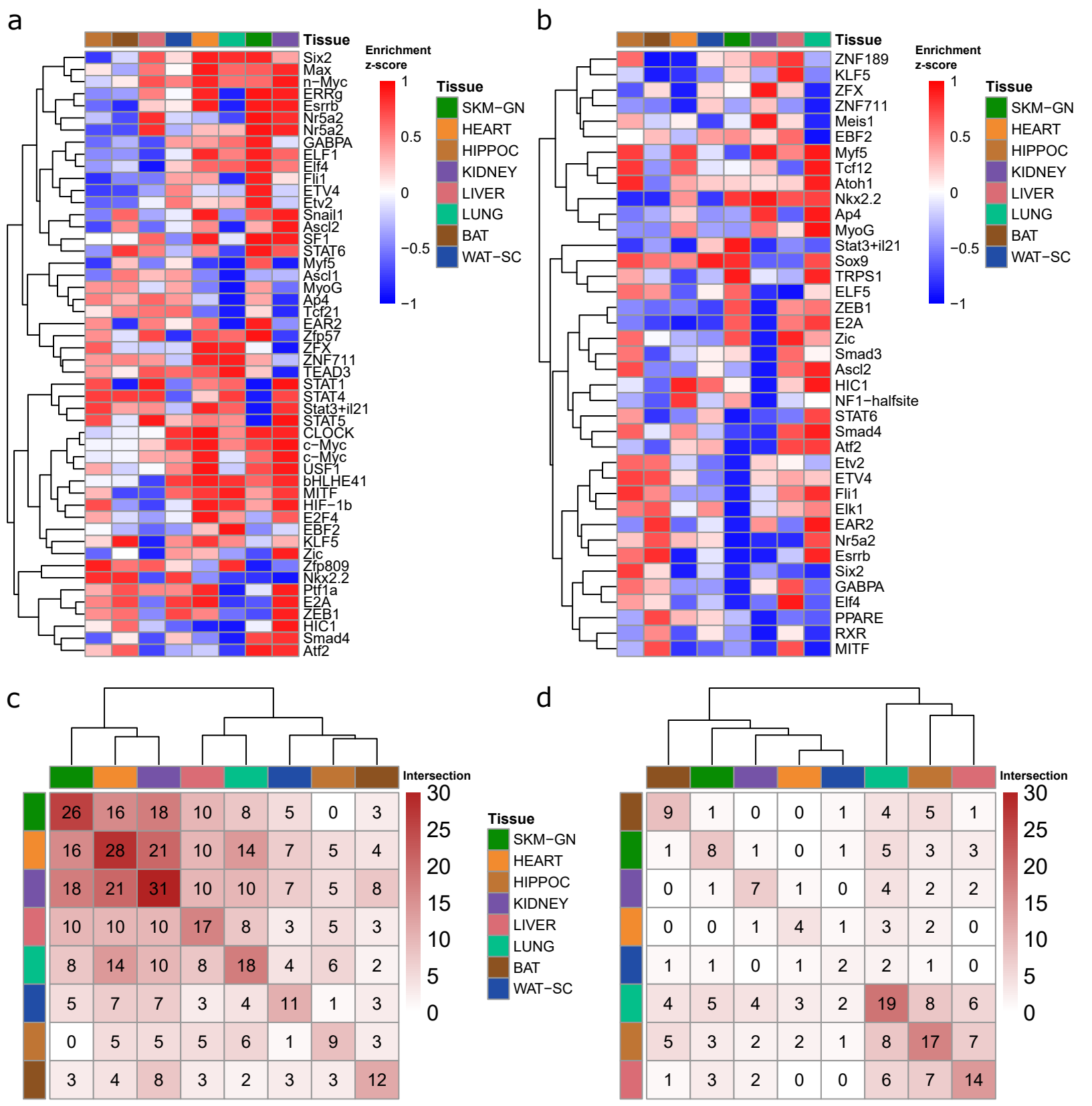
a
Tissue
Six2
Max
n−Myc
ERRg
Esrrb
Nr5a2
Nr5a2
GABPA
ELF1
Elf4
Fli1
ETV4
Etv2
Snail1
Ascl2
SF1
STAT6
Myf5
Ascl1
MyoG
Ap4
Tcf21
EAR2
Zfp57
ZFX
ZNF711
TEAD3
STAT1
STAT4
Stat3+il21
STAT5
CLOCK
c−Myc
c−Myc
USF1
bHLHE41
MITF
HIF−1b
E2F4
EBF2
KLF5
Zic
Zfp809
Nkx2.2
Ptf1a
E2A
ZEB1
HIC1
Smad4
Atf2
Enrichment z-score
1
0.5
0
−0.5
−1
Tissue
SKM−GN
HEART
HIPPOC
KIDNEY
LIVER
LUNG
BAT
WAT−SC
b
Tissue
ZNF189
KLF5
ZFX
ZNF711
Meis1
EBF2
Myf5
Tcf12
Atoh1
Nkx2.2
Ap4
MyoG
Stat3+il21
Sox9
TRPS1
ELF5
ZEB1
E2A
Zic
Smad3
Ascl2
HIC1
NF1−halfsite
STAT6
Smad4
Atf2
Etv2
ETV4
Fli1
Elk1
EAR2
Nr5a2
Esrrb
Six2
GABPA
Elf4
PPARE
RXR
MITF
Enrichment z-score
1
0.5
0
−0.5
−1
Tissue
SKM−GN
HEART
HIPPOC
KIDNEY
LIVER
LUNG
BAT
WAT−SC
c
Intersection
30
25
20
15
10
5
0
Tissue
SKM−GN
HEART
HIPPOC
KIDNEY
LIVER
LUNG
BAT
WAT−SC
d
Intersection
30
25
20
15
10
5
0

Figure 6: Tissues shared sets of TFs enriched in the promoter regions of genes that were up- or down-regulated after 8 weeks of training. (a-b) Heatmaps of relative enrichment of TFs among the promoter peaks of up- (a) or down-regulated genes (b) across tissues. (c-d) Number of enriched TFs shared between pairs of tissues in the promoter peaks of up- (c) or down-regulated genes (d).

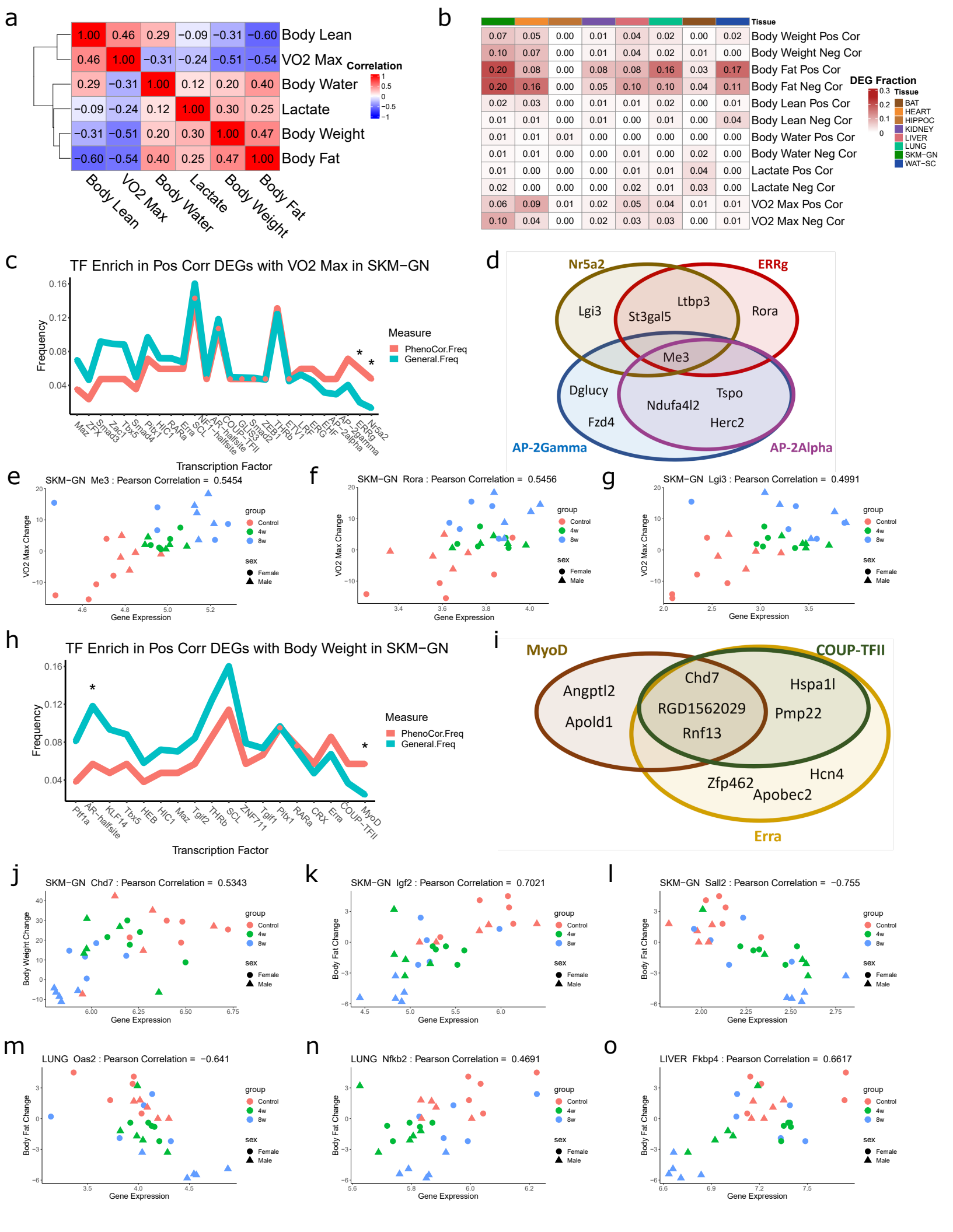
a
Correlation
Body Lean
VO2 Max
Body Water
Lactate
Body Weight
Body Fat
b
Tissue
DEG Fraction
BAT
HEART
HIPPOC
KIDNEY
LIVER
LUNG
SKM−GN
WAT−SC
Body Weight Pos Cor
Body Weight Neg Cor
Body Fat Pos Cor
Body Fat Neg Cor
Body Lean Pos Cor
Body Lean Neg Cor
Body Water Pos Cor
Body Water Neg Cor
Lactate Pos Cor
Lactate Neg Cor
VO2 Max Pos Cor
VO2 Max Neg Cor
c
TF Enrich in Pos Corr DEGs with VO2 Max in SKM−GN
Frequency
Measure
PhenoCor.Freq
General.Freq
Transcription Factor
d
Nr5a2
ERRg
Lgi3
Ltbp3
St3gal5
Rora
Me3
Dglucy
Tspo
Ndufa4l2
Fzd4
Herc2
AP-2Gamma
AP-2Alpha
e
SKM−GN Me3 : Pearson Correlation = 0.5454
VO2 Max Change
Gene Expression
group
Control
4w
8w
sex
Female
Male
f
SKM−GN Rora : Pearson Correlation = 0.5456
g
SKM−GN Lgi3 : Pearson Correlation = 0.4991
h
TF Enrich in Pos Corr DEGs with Body Weight in SKM−GN
i
MyoD
COUP-TFII
Angptl2
Apold1
Chd7
RGD1562029
Rnf13
Hspa1l
Pmp22
Zfp462
Apobec2
Hcn4
Erra
j
SKM−GN Chd7 : Pearson Correlation = 0.5343
Body Weight Change
k
SKM−GN Igf2 : Pearson Correlation = 0.7021
Body Fat Change
l
SKM−GN Sall2 : Pearson Correlation = −0.755
m
LUNG Oas2 : Pearson Correlation = −0.641
n
LUNG Nfkb2 : Pearson Correlation = 0.4691
o
LIVER Fkbp4 : Pearson Correlation = 0.6617

Figure 7: TF enrichment among phenotype-correlated DEGs. (a) Heatmap of Pearson correlation between phenotypic measures. Strong relationships between body lean and VO2 max and body fat and weight gain. (b) Frequency of tissue DEGs positively (> 0.5) or negatively (< -0.5) correlated with each phenotypic measure. (c) Comparison of TF enrichment among the active promoter peaks of VO2 max positively correlated DEGs vs TF enrichment among the total active promoter peak set in SKM-GN. * represents binomial test significance ($p < 0.05$) for difference in phenotype correlated DEG frequency and general frequency. (d) Overlap of target DEGs for most enriched TFs among VO2 max positively correlated DEG promoter peaks. (e-g) Scatter plots of gene expressed vs correlated phenotypic measure. In SKM-GN, VO2-max change is positively correlated with *Me3* (e), *Rora* (f), and *Lgi3* (g). (h) Comparison of TF enrichment among the active promoter peaks of body weight positively correlated DEGs vs TF enrichment among the total active promoter peak set in SKM-GN. * represents binomial test significance ($p < 0.05$) for difference in phenotype correlated DEG frequency and general frequency. (i) Overlap of target DEGs for most enriched TFs among body weight positively correlated DEG promoter peaks. (j-o) Scatter plots of gene expression vs correlated phenotypic measure. In SKM-GN, weight gain is positively correlated with *Chd7* (j), and body fat change is positively correlated with *Igf2* (k), and negatively correlated with *Sall2* (l). In lung, body fat change is negatively correlated with *Oas2* (m), and positively correlated with *Nfkb2* (n), and in liver, body fat change is positively correlated with *Fkbp4* (o).

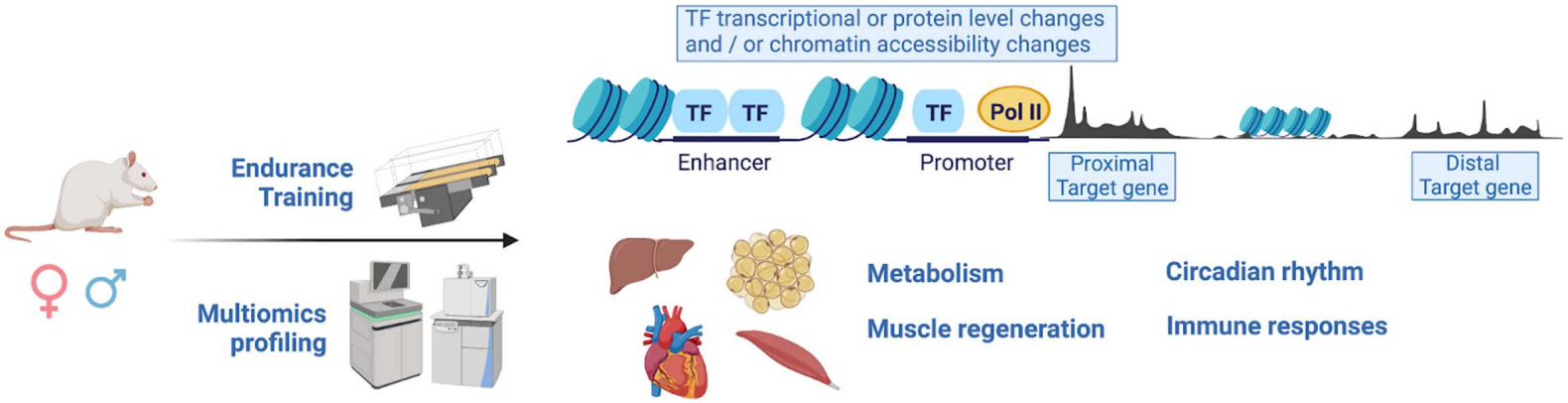
TF transcriptional or protein level changes
and / or chromatin accessibility changes
TF
TF
TF
Pol II
Enhancer
Promoter
Proximal
Target gene
Distal
Target gene
Endurance
Training
Multiomics
profiling
Metabolism
Muscle regeneration
Circadian rhythm
Immune responses

Figure 8: Transcription factors regulate exercise training-induced gene expression via multiple methods. Following eight weeks of endurance training, multiomic analysis across eight tissues have noted gene regulation through direct proximal promoter DAR to DEG relationships, DAR to distant correlated DEG relationships, tissue-specific or gene-region-specific TF machinery, and through changes in TF expression within a specific tissue. Enriched TF gene targets are associated with metabolism, muscle regeneration, immune responses and circadian rhythm pathways.