Multi-omic 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, DNA methylation, and RNA expression across eight rat tissues following endurance exercise training (EET) to map epigenomic changes to transcriptional changes and determine key TFs involved. We uncovered tissue-specific changes and TF motif enrichment across all omic layers, differentially accessible regions (DARs), differentially methylated regions (DMRs), and differentially expressed genes (DEGs). We discovered distinct routes of EET-induced regulation through either epigenomic alterations providing better access for TFs to affect target genes, or via changes in TF expression or activity enabling target gene response. We identified TF motifs enriched among correlated epigenomic and transcriptomic alterations, DEGs correlated with exercise-related phenotypic changes, and EET-induced activity changes of TFs enriched for DEGs among their gene targets. This analysis elucidates the unique transcriptional regulatory mechanisms mediating diverse organ effects of 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 peak
TSS	Transcription start site
TES	Transcription end site
TF	Transcription factor

Introduction

Regular exercise impacts health and modulates disease processes through the body. Exercise maintains muscle function, improves cardiovascular wellness and cognitive performance, and lowers the risk of cardiovascular disease and many other disorders, ranging from dementia to several cancer types[1]. The molecular processes mediating the adaptations induced by exercise training across tissues are poorly understood.

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, including DNA methylation, 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].

Our companion multi-tissue analysis of the molecular response dynamics during endurance exercise training found that the majority of differentially regulated genes are tissue specific whereas a small proportion are shared across multiple tissues[5,6]. 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[7], regulation of adipogenesis[8], and hippocampal BDNF activity and its cognitive effects[9]. As complex regulatory patterns drive tissue-specific gene regulation[10], 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, chromatin accessibility, and DNA methylation[11], and have concentrated on 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, chromatin accessibility, and DNA methylation were assayed in 8 tissues from age-matched male and female rats. By integrative analysis of all three omes within the same tissues, we establish a map of the regulatory transcriptional responses to training across tissues.

Results

Characterization of epigenetic and transcriptional responses to endurance training

To understand the epigenetic and transcriptional response mechanisms elicited during eight weeks of EET, we analyzed ATAC-seq, RNA-seq and RRBS profiles 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. We identified differentially accessible regions (DARs; F test adjusted p value < 0.1), differentially methylated regions

(DMRs: 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, DMRs, and DEGs. Although most expressed genes, open chromatin sites, and methylation sites were detectable in multiple tissues, the majority of DARs (90%), DMRs (91%), 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[6].

We then examined the distribution patterns of log2 fold change (L2FC) in gene expression, L2FC in chromatin accessibility, and L2FC in methylation across time points and sexes (Fig. S2, S3, S4 respectively). While the ratios of up- to down-regulated analytes (i.e. DEGs, DARs, DMRs) were similar across the majority of tissues, L2FC patterns differed between tissues. Notably, the spread of L2FC was notably higher among DMRs than DEGs or DARs in each tissue. Additionally, we observed sex differences among DEGs in HIPPOC, BAT, and WAT-SC. Sex differences were more prevalent among DARs and DMRs in a majority of tissues. Overall, the proportion of DEGs showing concordant changes (L2FC) across timepoints and sex groups was higher than that of DARs and DMRs (Fig. 1d). Notably, heart, SKM-GN, and kidney exhibited the most consistency in expression changes across all 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 S5). 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. S6a) and to sex in WAT (Fig. 1f and Fig. S6b). 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.

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[12,13], 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[14], the enrichment of DARs in this region suggested that EET results in the transcriptional activation of target genes. DMRs were significantly enriched at downstream regions across all tissues except WAT, relative to all methylation sites in each tissue (Fig 1i,j,l). Unlike DARs and DMRs, 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 S7); however, ATAC-seq peaks that mapped to either expressed genes or DEGs contained a higher proportion of intronic peaks and a lower proportion of distal intergenic peaks compared to all open chromatin peaks in the dataset.

Tissue specificity is not only limited to the training response itself. Within each ome, we identified a consistent pattern of tissue-specific enrichment of the analytes that exhibit a significant training response in a given tissue (Figure 2). In the case of both RNAseq and ATACseq data, DEGs and DARs in a given tissue were, on average, more highly expressed (Figure 2a,d) and more highly accessible (Figure 2b,e), respectively, at baseline in the tissue with the significant training response. Conversely, DMRs in a given tissue were more hypomethylated in the tissue with the significant training response (Figure 2c,f). While there is some overlap in significant training responses across tissues, the majority of training response analytes represent tissue-enriched analytes, even if they are not tissue-specific.

Identification of distal correlated epigenetic regulation despite few DARs and DMRs mapping to adjacent 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 higher proportion of overlap between DARs and DEGs (Fig 3a). 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 3a). Similarly, we investigated DMR-DEG associations by assigning each DMR to the nearest gene and determining the fraction that were annotated to DEGs. Only BAT and WAT-SC have significant overlaps between DMRs and DEGs (Fig 3b). BAT and WAT-SC are the two tissues with the largest number of DMRs in the dataset by a considerable margin.

The binding of TFs to distal open chromatin regions can regulate gene transcription[15,16]. 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 3c). BAT, SKM-GN, and liver contributed most of the DARs adjacent to a DEG, confirming our earlier observations (Fig 3a). A similar pattern is seen when measuring the distance between a DMR and the nearest DEG (Fig 3d) and the distance between a DMR and the nearest DAR (Fig 3e). BAT and WAT-SC contain the majority of closest DMR-DEG pairs likely because of their greater DMR-DEG overlap (Fig 3b) and BAT contains the majority of closest DMR-DAR relationships.

DARs that were adjacent to DEGs tended to be more highly correlated with gene expression changes (Fig 3f, Fig S8), a pattern predominantly seen in SKM-GN and liver. In contrast, the majority of DMRs adjacent to DEGs exhibit strong positive or negative correlations driven primarily by BAT and WAT-SC (Fig 3g, Fig S9). Interestingly, while instances of adjacent DMRs and DARs are more limited in this study, in tissues more populated by DAR-DMR pairs, positive correlation tended to occur when the DMR was upstream of the DAR, and correlation was decreased or negative when the DMR was downstream of the DAR (Fig 3h, Fig S10).

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

We focused on DARs and DEGs located within 500 kb from 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; 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. Pathway enrichment analysis among DAR-DEG pairs identified distinct patterns of enrichment for each tissue (Fig S11). In agreement with the training-associated increase in immune cell types inferred from cell type deconvolution analysis (see Fig. 1e), DAR-DEG pairs in BAT showed enrichment for several immune pathways. Lung also showed considerable enrichment for immune-associated pathways, suggesting activating roles in antigen defense by exercise. 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 S11). These results suggested that the correlated epigenetic and transcriptional changes induced by training affected tissue-specific functions.

To identify key regulators of the training response in each tissue, we analyzed TF motif enrichment at the DARs of DAR-DEG pairs. Notably, we identified MAZ and SMAD3 as regulatory TFs in specific tissues (Fig. 4a-c). 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 4a,b). MAZ can act both as a transcriptional activator and a repressor[14,17]. In SKM-GN, MAZ target genes included: Igf2 (Fig 4d), which plays pivotal roles in exercise response[18,19], SKM growth, and differentiation[20]; Ppp1r15a, which is associated with innate immunity[21]; and Sall2 (Fig 4e), a TF typically associated with development and neuronal differentiation[22]. In the lung, MAZ targets included immune response genes such as Mpeg1, Oas2, Nfkb2 (Fig 4f), as well as stress response gene Hspb6.

SMAD3 binding sites were enriched in the liver (Fig 4c), with a 22% enrichment among DAR-DEG pairs vs. a 5.4% general enrichment (p-value = 1.255e-05), suggesting combinatorial transcriptional regulation. Paired DAR sets were positively correlated with *Glul* expression (Fig S12a,b), negatively correlated with *Lpar3* (Fig S12c,d) expression, and negatively correlated with multiple members of the *Serpina* gene family (Fig 4c, S12e,f). *Fkbp4* expression was also negatively correlated with a single SMAD3 motif-containing DAR (Fig 4g). The remaining SMAD3 targets *Abhd2* (Fig S12g), *Onecut1* (Fig S12h), *Ccnd1* (Fig S12i), and *Xbp1* (Fig S12j) 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[23]. Other potential transcriptional regulators along with target genes are illustrated in Fig. S12.

Correlated DMR-DEG pairs enriched for key TFs in adipose and lung tissue

We identified DMR-DEG pairs located within 500kb of each other and whose L2FC training responses were either positively or negatively correlated across time points and sexes (> 0.5 or < -0.5) (Supplementary Table 2). The majority of DMR-DEG pairs were found in WAT-SC, 329 of 393 in total, while 33 were found in lung, 13 in heart, 12 in SKM-GN, and 6 in liver. The heavy skew towards WAT-SC is accounted for by the increased numbers of both DEGs and DMRs found in the tissue. We identified 24 TFs whose motifs were significantly enriched among the DMRs in correlated neighboring DMR-DEG pairs in each tissue (Fig 4H). The majority of enriched TFs were found in WAT-SC, while 4 were found in lung, and LRF was the only TF enriched in SKM-GN. NF1-halfsite was the most significantly enriched TF in WAT-SC, with

motifs found in 27 DMRs correlated to 55 distinct DEGs (Fig S13a), including chr2-375476_cluster1, whose training response is highly correlated ($r > 0.75$ or $r < -0.75$) for five genes, four negatively correlated: *Rab25* (Fig S13b), *Crabp2* (Fig S13c), *Paqr6* (Fig S13d), *Tmem79* (Fig S13e), and one positively correlated: *Syt11* (Fig S13f).

AP-2gamma is enriched in multiple tissues, with motifs found in 19 DMRs correlated to 51 DEGs in WAT-SC (Fig S13g), and found in 5 DMRs correlated to 7 DEGs in lung (Fig S13h). *B4galnt1* is a member of positively correlated target DMR-DEG pairs in both WAT-SC (Fig S13i) and lung (Fig S13j). Members of the Rho GTPase family are represented by target DMR-DEG pairs in both tissues as well, including *Arhgap9* in lung (Fig S13k) and *Arhgap9*, *Arhgef4*, *Arhgef25*, and *Arhgef2* in WAT-SC. The TMEM family is also represented in both tissues: *Tmem176b* in lung, and *Tmem79* in WAT-SC.

Characterization of TF expression responses to EET

As putative transcriptional regulators were inferred from DAR-DEG and DMR-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 abundances and phosphorylation levels, based on mass spectrometry data from a subset of 6 tissues. Various subsets of TFs exhibited significant changes at the transcriptome (Fig 5a), proteome (Fig 5b), and phosphoproteome (Fig 5c) levels in each tissue. BAT and WAT-SC had the largest and most significant changes in TF gene expression (Fig 5a), which included immune-related TF genes such as *Irf8* and *Pou2f2* responding in both tissues and *Irf1* and *Spi1* specifically in BAT. Transcript levels of *Egr1* decreased significantly across multiple tissues including SKM-GN, heart, kidney, and lung. Fellow early response gene *Fos* also decreased across the same tissues, reaching significance in kidney. *Myb* expression was significantly altered in both HIPPOC, where transcript levels increased in response to training in both sexes, and kidney where transcript levels increased in females, but decreased in males.

Similar to TF gene expression changes, TF protein level responses to EET were largely tissue-specific (Fig 5b). WAT-SC and lung exhibited the most significant changes. Select TFs showed significant protein level changes in multiple tissues, including RORC which decreased in female subjects in both lung and kidney. PBX1 decreased in WAT-SC while PBX2 decreased in kidney, and ATF1 and ATF3 increased in male subject protein levels in WAT-SC and heart, respectively. NR4A1 (NUR77) and MEF2A levels decreased in SKM-GN at week 8 of training, while MEF2C levels increased at week 2 of training.

MEF2C showed significant protein phosphorylation changes in SKM-GN, lung, and WAT-SC (Fig 5c). Other MEF family members MEF2A and MEF2D also exhibited changes in phosphorylation in lung and SKM-GN, respectively. A number of TFs had multiple significant phosphosite changes in a single tissue, including two NF1 phosphosites in heart, two STAT3 sites and two ATF2 sites in liver, two RFX5 and two MEF2C sites in lung, and two USF2 sites and four TRPS1 sites in WAT-SC.

Now that we determined which TFs were found to have significant training responses at either a transcriptome, proteome or phosphoproteome level, we wanted to find which of their motifs were enriched among the promoter regions of DEGs, identifying potential training-induced regulatory relationships. We found a number of TFs whose motifs were enriched among DEGs (Fig 5d, Fig S14-S15), including eight statistically significant motif enrichments. The majority of significantly enriched TFs were found in SKM-GN: SF1, SIX1 (Fig 5e), SIX2 who exhibited a

significant training response at the transcriptome level, MEF2C with significant training responses at both the proteome and phosphoproteome levels (Fig 5f), and MEF2D with a significant training response at the proteome level. MEF2A had a significant training response at the phosphoproteome level in heart, while PU.1 had a significant training response at the transcriptome level in lung, and IRF:BATF had a significant training response at the proteome level in lung (Fig 5g).

SIX1 DEG targets in SKM-GN exhibit a range of functions and training responses, including collagen gene *Col3a1*, and muscle contraction associated gene *Lmod1*, both of which decreased in response to training. Protein modification-associated genes *Golga4* and *Art1* increased over training while ubiquitin gene *Usp2* decreased. Malic enzyme *Me3*, which is involved in the oxidative decarboxylation of malate to pyruvate, is consistently higher expressed during training. MEF2C protein levels significantly increased at week 2 of training then returned to baseline, while phosphorylation increased significantly throughout the eight weeks of training. Among MEF2C target genes, clock gene *Per1* demonstrated the highest increase in expression at the onset of training, while *Sema6c*, *Ankh*, *Ptpn1*, and *Phkg1* exhibit decreased expression following training. *Dystrophin* (*Dmd*), a critical protein for muscle fiber integrity[24], was the most negatively correlated with MEF2C protein level changes, but positively correlated with MEF2C phosphorylation changes. IRF:BATF targets in lung predominantly exhibited decreased expression following training, including tubulin gene *Tuba1c*, mitochondrial biogenesis-associated gene *Perm1* and actin cytoskeleton organizational gene *Cfl1*. We examine the DEG targets of other TFs with significant training responses in Supplemental Figures S16 and S17, including JUND, which is enriched for DEGs in SKM-GN, heart and WAT-SC, and NR4A1 (NUR77) which significantly decreased in protein level in SKM-GN as did the majority of its DEG targets including heat shock protein *Hspa1l* and dual specificity phosphatase *Dupd1*. Altogether, these findings support the functional relevance of EET-regulated TFs.

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

The lack of nearby DARs or DMRs for the majority of DEGs within each tissue (see Fig. 3a,b) led us to hypothesize that DARs, DMRs, and DEGaPs may mediate different paths of transcriptional regulation: i) DARs and DMRs 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, DMRs, or DEGaPs relative to all open chromatin peaks in each tissue (Fig. 6a,b,c; see Methods).

Due to their very low number of DARs, WAT and HIPPOC were removed from the analysis (Fig. 6a). 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[25]. SIX2 and MEF2C motifs were enriched in heart, COUP-TFII in kidney, 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 S18a).

A similar pattern of tissue-specific motif enrichment was observed among DMRs for each tissue (Fig 6b), although with different TFs enriched among DMRs vs DARs. MYOG was the most

significantly enriched TF among DMRs in SKM-GN, along with two neuronal TFs NEUROD1, NEUROG2. Another nervous system-associated TF NKX2.2 was the most significantly enriched in heart. PRDM9 was the most enriched in HIPPOC, as was PAX5 in kidney, and multiple HOX TFs in liver including HOXA10 and HOXD12 which were enriched among DARs in lung. TF enrichment significance among DMRs overlapped more between lung, BAT and WAT-SC tissue with ELK4, ELK1, ETV1 and ETV4 all enriched across the three tissues. GLIS3 enrichment was isolated to WAT-SC. As with DARs, the frequency of motif presence in DMRs across tissues maintained tissue specificity and even more sharply distinguished lung, BAT and WAT-SC tissues (Fig S19a).

Motif enrichment patterns among DEGaPs differed considerably from those in DARs and DMRs (Fig. 6c). Indeed, pairwise motif enrichment comparisons between DARs, DMRs, and DEGaPs per tissue were weakly correlated (<0.39; Fig S21, Fig S22, Fig S23). 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[26]. ETS and ELF TFs are associated with the regulation of immunity[27,28], 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 S20a), 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, DMRs, or DEGaPs and the relative expression levels of the corresponding TFs, we examined TF gene expression patterns in control tissues (Fig S18b, Fig S19b, Fig S20b respectively) and L2FC in TF gene expression following EET (Fig S18c, Fig S19c, Fig S20c, respectively). We found a stronger correlation between motif enrichment and TF control gene expression levels among DEGaPs (Fig S20d-f) than among DARs (Fig S18d-f), and even found a negative correlation consistently among DMRs (Fig S19d-f). Conversely, there was no correlation between motif enrichment and L2FC in TF gene expression, be it among DARs (Fig. S18a,c), DMRs (Fig. S19a,c), or DEGaPs (Fig. S20a,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 and DMRs can influence DEGs directly, supporting the concept that DEGaPs vs. DARs and DMRs mediate distinct paths of transcriptional regulation.

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

The different genomic distributions of DARs and DMRs (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, DMRs, 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, DMRs, 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 6d). Proximal promoter DARs showed strong enrichment for SP and KLF TFs, MAZ and PITX1. Downstream DARs were enriched for NPAS and BMAL1, two core circadian clock TFs, as well as NKX TFs. The NF1 subcluster also showed enrichment among 3' UTR DARs, while THRB was most highly enriched among 3' UTR DARs.

In contrast, TFs were most highly enriched among downstream DMRs versus other genomic regions (Fig 6e), including multiple KLF TFs, MYOG, MAZ and EGR1. ETV2 and ETS1 were enriched among proximal promoter DMRs, while BM2 and PIT1 were enriched among upstream DMRs. Another cluster of TFs were enriched among exon DMRs including c-MYC, HIF2A, and MYOD. TF enrichment among DEGaPs was most significant in intronic regions, one of the regions most populated with DEGaPs (Fig 6f). Notably, however, there was no enrichment among distal intergenic DEGaPs despite their large proportion of total DEGaPs. TFs enriched among intronic DEGaPs include SCL, SMAD3, THRB and ERRA. Significant enrichment was also found among exonic DEGaPs, highly correlated to intronic DEGaP enrichments. While also highly enriched among intronic DEGaPs, the most enriched TFs among promoter DEGaPs include MAZ and multiple KLF family members. Overall, genomic features appeared to be a driving force in clustering the enriched binding motifs among DEGaPs, DARs, and DMRs.

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

A companion EET MoTrPAC manuscript (Nair et al. manuscript under review), found that the biological pathways that were over-represented among upregulated vs. downregulated DEGs in each tissue after 8 weeks of training were distinct, 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 peaks, as shown in Fig. 1g), 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 S24). Adipose tissues had the greatest motif enrichment overlap between promoter peaks associated with up-regulation 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. S25), revealed some overlap in motif enrichment between the tissues (Fig. 7a,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. 7a) and the 39 associated with downregulation (Fig. 7b), 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 SKM-GN shared the most enriched motifs associated with downregulation (Fig 7c,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. S26). Body weight and fat mass were significantly lower in 8-week trained rats than in controls (Fig S26a,b), with the greatest difference observed in males. VO2 max significantly increased in both females and males in response to EET (Fig S26c). Body lean mass also increased in both sexes, though significance was only reached in females (Fig S26d). Lactate (Fig S26e) and body water (Fig S26f) 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 8a).

In most tissues, DEGs formed 2 separate clusters based on their either positive or negative correlation with a measured parameter change (Fig S27). Correlation with body fat involved the highest proportion of DEGs in most tissues (Fig. 8b), except for BAT and HIPPOC.

Connection of key TFs to 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. 8c). 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 8d). 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. 8e-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. 8h). While an enrichment for the COUP-TFII and ERRA motifs was also detected, it did not reach significance (Fig 8h). The *Chd7, Rnf13,* and *RGD1562029* DEGs showed concurrent enrichment for the MyoD, COUP-TFII, and ERRA motifs (Fig 8i). The positive correlation between *Chd7* expression and body weight change is illustrated in Fig. 8j.

As MAZ and SMAD3 have been identified as major regulators of the DARs-associated DEGs (DAR-DEG pairs; see Fig. 4), 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. 8k,l); in lung, *Oas2* and *Nfkb2* were negatively and positively correlated with body fat change, respectively (Fig. 8m,n); in the liver, SMAD3-target *Fkbp4* was also positively correlated with body fat change (Fig 8o). 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. 8j-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. 5) and phenotypic changes. We observed the strongest correlations between TF RNAs and body fat changes (Fig S28a), 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 S28b). 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. The comprehensive study of rats allowed extending the understanding of the molecular mechanisms of EET to important tissues not feasible in human subjects. We integrated chromatin accessibility, DNA methylation, 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, post-translational phosphorylation of TF proteins, direct relationships between proximal promoter DARs and DMRs and their associated DEGs, interactions between more distant highly correlated DAR-DEG and DMR-DEG pairs enriched for known TFs and specific training responses with distinct TF enrichment patterns, some conserved across tissues (Fig 9). 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. We identified MEF2A, MEF2C and MEF2D TFs with significant changes in protein level or phosphorylation whose motifs were enriched in DEGs in SKM-GN and heart. We found that the DEGaPs of upregulated vs. downregulated DEGs show differential motif enrichment, which predominantly occurs at promoter regions. Finally, we identified TFs associated with EET-induced phenotypic changes including VO2 max and body weight.

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[23], 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[29,30]. 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[31]. SMAD3-target gene *Lpar3* has been linked to liver regeneration following damage[32] and *Lpar* family members were previously associated with liver fibrosis[33].

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[34]. 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. Not previously linked to EET, MAZ can act both as a transcriptional activator and a repressor[14,17]. Multiple MAZ target genes found among the DAR-DEG pairs were associated with immune response: *Ppp1r15a* (*Gadd34*) is necessary for interferon production[35]; *Igf2*, which is primarily a growth hormone during development, is also involved in immune response[36-38]; *Hspb6* acts as a mediator of platelet aggregation along with smooth muscle relaxation[39]; *Mpeg1* expresses a transmembrane protein in macrophages in an antimicrobial capacity[40]; *Oas2* is a type I-interferon response gene[41]; mutations in *Nfkb2* greatly damage the immune system[42]. Thus, MAZ could be involved in the inflammatory response occurring in SKM following exercise[43].

The motif of AP2-gamma, also known as TFAP2C, was enriched among correlated DMR-DEG pairs in WAT-SC and lung. The AP2-gamma motif was also significantly enriched among DEGs in SKM-GN that are correlated with a change in VO2max over EET. AP2-gamma's primarily studied function is in early development morphogenesis[44]; however, it has also been shown to have an inflammatory role by activating Th17 and Th1 cells[45]. The NF1-halfsite motif was also enriched among correlated DMR-DEG pairs in WAT-SC. NF1 negatively regulates the RAS/MAPK pathway[46], and mutations in the protein lead to neurofibromatosis type 1 which can have numerous negative effects including musculoskeletal defects and impaired exercise capabilities[47].

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[48-51]. 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[48,52]. In SKM, NR4A1 regulates genes associated with glucose uptake, glycogen synthesis, and promoting muscle growth[53] as well as mediating inflammatory response[54,55]. 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 in our study were the heat shock protein *Hspa1l*, a critical element of the cellular stress response[56], and *Dupd1* (*Dusp29*), which has been linked to regulation of muscle cell differentiation, development, and atrophy[57].

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[58] which bind to the core clock proteins CLOCK[59] and BMAL1[60], respectively. Moreover, we found an over-representation of binding sites for circadian clock-related TFs at promoter and downstream DARs (NPAS and BMAL1 motif), 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[61-66].

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 interindividual 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. The goal of the MoTrPAC project is to provide the first systematic compendium of exercise effects. Follow-up studies of the multi-organ roadmap of genomic regulatory responses that we have identified are warranted.

Our multi-omic analysis across rat tissues allowed us to map the epigenomic changes to the transcriptional changes occurring during EET and infer TFs driving training responses. By providing a view of the complex interplay between chromatin structure modifications, DNA methylation, gene transcription, and TF abundance and activity throughout EET, this work provides a comprehensive survey of the multi-organ gene control mechanisms underlying the effects of regular exercise. This insight helps further the goal of maximizing the benefits of exercise in healthy individuals and developing targeted exercise therapies for patients with 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.

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[67]. 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 deeply anesthetized with approximately 1-2% isoflurane in oxygen, and gastrocnemius, white adipose, liver, lung, and heart were removed under anesthesia. Following removal of the heart, the rat was decapitated using a guillotine. The brain was removed from the skull and hippocampus dissected. The remaining tissues (kidney, brown adipose, and hippocampus) were dissected post death. All tissues 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[68]. 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)[69]. Samples from the same sex and training group were analyzed as biological replicates. Reads were aligned to genome rn6.0.96[70] with Bowtie 2 (v2.3.4.3)[71]. Duplicate and mitochondrial chromosome mapped reads were removed. Peaks were called using MACS2 (v2.2.4)[72], 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[73] 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)[74] 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[75]. 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[76] (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.

DNA methylation data generation

Rat tissues were disrupted in GenFind v2 lysis buffer (Beckman Coulter, Indianapolis, IN) with a tissue ruptor (Omni International, Kennesaw, GA). Genomic DNA was extracted in a BiomekFx automation workstation (Beckman Coulter, Chaska, MN). DNA sample quantification was done by Qubit assay (dsDNA HR assay, Thermo Fisher Scientific) and quality was determined by Nanodrop A260/280 and A260/230 ratios. Reduced representation bisulfite sequencing (RRBS)[77] libraries were generated with the Ovation® RRBS Methyl-Seq kit from Tecan Genomics (Baldwin Park, CA). Quantity of the libraries was determined by Qubit High Sensitivity assays (Thermo Fisher Scientific) and quality evaluation was done by Bioanalyzer High Sensitivity DNA Chip (Agilent Technologies, Santa Clara, CA). Libraries were sequenced on a NovaSeq 6000 platform (Illumina, San Diego, CA) using paired-end 100 base-pair run configuration.

RRBS data processing and normalization

bcl2fastq (version 2.20) was used to demultiplex reads with options --use-bases-mask Y*,I8Y*,I*,Y* --mask-short-adapter-reads 0 --minimum-trimmed-read-length 0 (Illumina, San Diego, CA). FastQC (v0.11.8) was used to calculate pre-alignment QC metrics[78] and reads were indexed and aligned to the Ensembl *Rattus norvegicus* (rn6) genome using Bismark (v0.20.0)[79]. Bowtie 2 (v2.3.4.3) was used to quantify the percent of reads that mapped to globin, rRNA and phix sequence contaminants and spike-ins[71]. Chromosome mapping percentages were calculated with SAMtools (v1.3.1)[80] and NuGEN's "nodup.py" script quantified PCR duplicates.

CpG sites were selected for downstream analysis if they exhibited methylation coverage >= 10 in all samples. Individual CpG sites were divided into 500 base-pair windows and clustered with the Markov clustering algorithm R package *MCL*[81]. Quantile normalization of sites/clusters was conducted separately on each tissue using R package *preprocessCore*[82].

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[70]. 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[71]. 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)[83]. 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[84]. 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[85]. 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[86]. Peptides were labeled using tandem mass tag (TMT)[87] 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).

Phosphoproteomics data generation

Phosphopeptide enrichment was performed through immobilized metal affinity chromatography (IMAC)[5]. Phosphopeptides were eluted off IMAC beads in 3x70 µl agarose bead elution buffer, desalted with C18 stage tips, eluted with 50% ACN, and then lyophilized. Samples were reconstituted in 3% ACN / 0.1% FA for LC-MS/MS analysis. Heart and liver samples were separated by a nanoflow Proxeon EASY-nLC 1200 UHPLC system (Thermo Fisher Scientific) and analyzed with a Q-Exactive HFX mass spectrometer (Thermo Fisher Scientific). SKM-GN, WAT-SC, kidney and lung samples were separated by a Dionex Ultimate 3000 UHPLC direct-inject system (Thermo Fisher Scientific) then analyzed with a Q-Exactive HFX mass spectrometer.

Proteomics and phosphoproteomics 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 and phosphosites. 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). Phosphoproteome data was not normalized to the total proteome due to the lack of complete overlap of phosphosites and total proteome features (80.5% - 89.7%).

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, proteomics, and phosphoproteomics data[88]; the edgeR pipeline for methylation analysis was used for RRBS data[89]; DESeq2 was used for RNA-Seq[78]. 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, *edgeR::glmQLFTest*) 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[90]. Training-differential features were selected at 10% 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, for edgeR models we used *edgeR::glmQLFTest*. 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)[91]. 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[92] using the getallSPVs function. Marker sets were generated using the IRIS (Immune Response *In Silico*[93] and DMAP (Differentiation Map) reference datasets[94]. 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[95] 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. MoTrPAC Study Group *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. Nair, V. D., *et al*. Molecular adaptations in response to exercise training are associated with tissue-specific transcriptomic and epigenomic signatures. *Cell Genomics*. In press.

7. Oliveira, A. N. & Hood, D. A. Exercise is mitochondrial medicine for muscle. *Sports Med. Health Sci.* **1**, 11–18 (2019).

8. Gregoire, F. M., Smas, C. M. & Sul, H. S. Understanding Adipocyte Differentiation. *Physiol. Rev.* **78**, 783–809 (1998).

9. 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).

10. Sonawane, A. R. *et al.* Understanding Tissue-Specific Gene Regulation. *Cell Rep.* **21**, 1077–1088 (2017).

11. Ramachandran, K. *et al.* Dynamic enhancers control skeletal muscle identity and reprogramming. *PLOS Biol.* **17**, e3000467 (2019).

12. Miao, Z. *et al.* Single cell regulatory landscape of the mouse kidney highlights cellular differentiation programs and disease targets. *Nat. Commun.* **12**, 2277 (2021).

13. 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).

14. 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).

15. 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).

16. Panigrahi, A. & O'Malley, B. W. Mechanisms of enhancer action: the known and the unknown. *Genome Biol.* **22**, 108 (2021).

17. 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).

18. 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).

19. 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).

20. 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).

21. 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).

22. Farkas, C. *et al.* Characterization of SALL2 Gene Isoforms and Targets Across Cell Types Reveals Highly Conserved Networks. *Front. Genet.* **12**, 613808 (2021).

23. Amar, D. *et al.* Time trajectories in the transcriptomic response to exercise - a meta-analysis. *Nat. Commun.* **12**, 3471 (2021).

24. 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).

25. 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).

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

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

28. 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).

29. 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.

30. 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).

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

32. 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).

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

34. 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).

35. 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.

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

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

38. 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).

39. 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).

40. 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).

41. 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).

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

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

44. Auman, H. J., *et al*. Transcription factor AP-2γ is essential in the extra-embryonic lineages for early postimplantation development. *Development*. **129**, 2733-2747 (2002).

45. Zhang, H., *et al.* TFAP2C exacerbates psoriasis-like inflammation by promoting Th17 and Th1 cells activation through regulating TEAD4 transcription. *Allergol. Immunopathol.* **51**, 124-134 (2023).

46. Trovó-Marqui, A. B., *et al.* Neurofibromin: a general outlook. *Clin. Genet.* **70**, 1-13 (2006).

47. Ferreira de Souza, J., *et al.* Exercise capacity impairment in individuals with neurofibromatosis type 1. *Am. J. Med. Genet. Part A.* **161A**, 393-395 (2013).

48. 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).

49. 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).

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

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

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

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

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

55. 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).

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

57. 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).

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

59. 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).

60. 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).

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

62. 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).

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

64. 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).

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

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

67. 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).

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

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

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

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

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

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

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

75. 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).

76. 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).

77. Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res.* **33**, 5868–5877 (2005).

78. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. https://qubeshub.org/resources/fastqc (2015).

79. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics*. **27**, 1571-2 (2011).

80. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

81. Jäger, M. L. MCL: Markov Cluster Algorithm. https://CRAN.R-project.org/package=MCL (2015).

82. Bolstad, B. preprocessCore: A collection of pre-processing functions. *GitHub* https://github.com/bmbolstad/preprocessCore (2021).

83. 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).

84. 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).

85. 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).

86. 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).

87. 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).

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

89. Chen, Y., Pal, B., Visvader, J. E. & Smyth, G. K. Differential methylation analysis of reduced representation bisulfite sequencing experiments using edgeR. *F1000Research* **6**, 2055 (2017).

90. 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).

91. 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).

92. 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).

93. 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).

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

95. 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-S28 and corresponding legends.

Supplementary Table S1. Correlated DAR-DEG pairs 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.

Supplementary Table S2. Correlated DMR-DEG pairs annotated with an expressed TF motif. For each pair, we listed the tissue, DEG ensembl id and symbol, DMR id, distance between DMR midpoint and DEG TSS, Pearson correlation, known TF motifs within the DMR, and the L2FC of the DEG and the DMR for each time point relative to eight week control in both female and male samples. Brown Adipose DMR-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

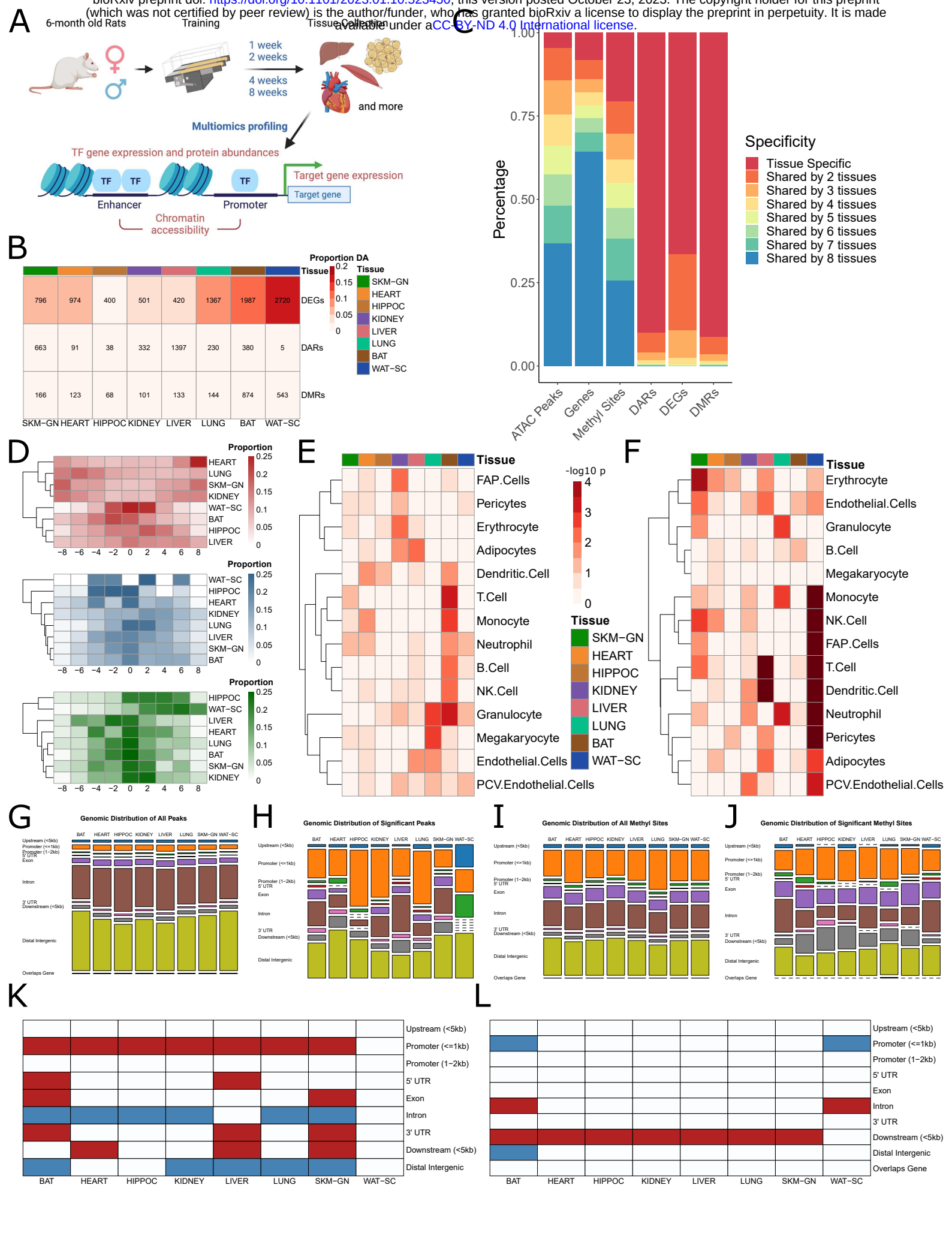

Figure 1: Epigenetic and transcriptional responses to training programs. (a) 6-month old rats of both sexes underwent training programs. Tissues were collected and subjected to multiomics profiling, including ATAC-seq, RNA-seq, proteomics. (b) Higher number and percentage of differentially expressed genes (DEGs) were identified than differentially accessible regions (DARs) and differentially methylated regions (DMRs) after training in most tissues (F test FDR<0.1). (c) Many accessible regions, methylation sites, and genes were identified in all tissues and training-induced features were highly tissue-specific.(d) Distribution of L2FC positive/negative consistency in RNAseq, ATACseq and RRBS differential analytes across tissues. The sum of the sign(L2FC) at each time point in males and females for each analyte is calculated with values ranging from -8 (negative L2FC at all time points and sexes) to 0 (half positive and half negative L2FC) to 8 (positive L2FC at all time points and sexes). Heart and SKM-GN RNAseq DEGs are more consistently up or down-regulated while WAT-SC RNAseq DEGs are less consistent, for example. (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), all methylation sites (i) and DMRs (j). (k) DARs enriched for the proximal promoter compared to all accessible regions. (l) DMRs are enriched for the downstream region compared to all methylation sites.

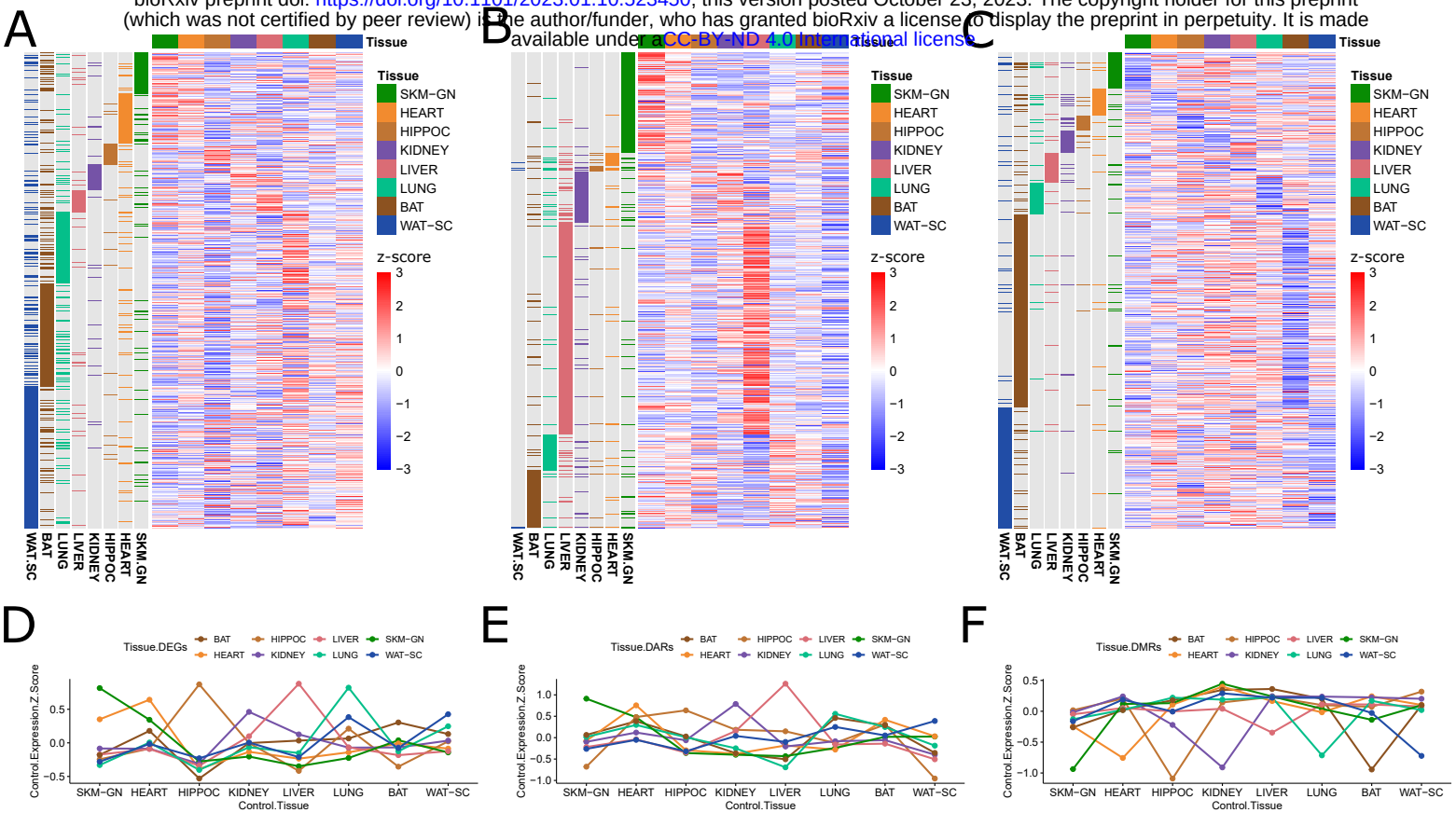

Figure 2: In general, genes are more highly expressed, chromatin peaks are more highly accessible and methylation sites are more hypomethylated in the tissues in which they exhibit a differential training response. (a-c) heatmaps of all differential analytes across eight tissues in RNAseq (a), ATACseq (b), and RRBS (c). Columns reflect z scores of baseline expression (a), accessibility (b), or methylation (c), across the eight tissues. Rows are annotated by the tissues in which each analyte exhibits a differential training response. (d-f) Mean z-score of control gene expression (d), chromatin peak accessibility (e), and site methylation (f), for the differential training response analytes for each tissue.

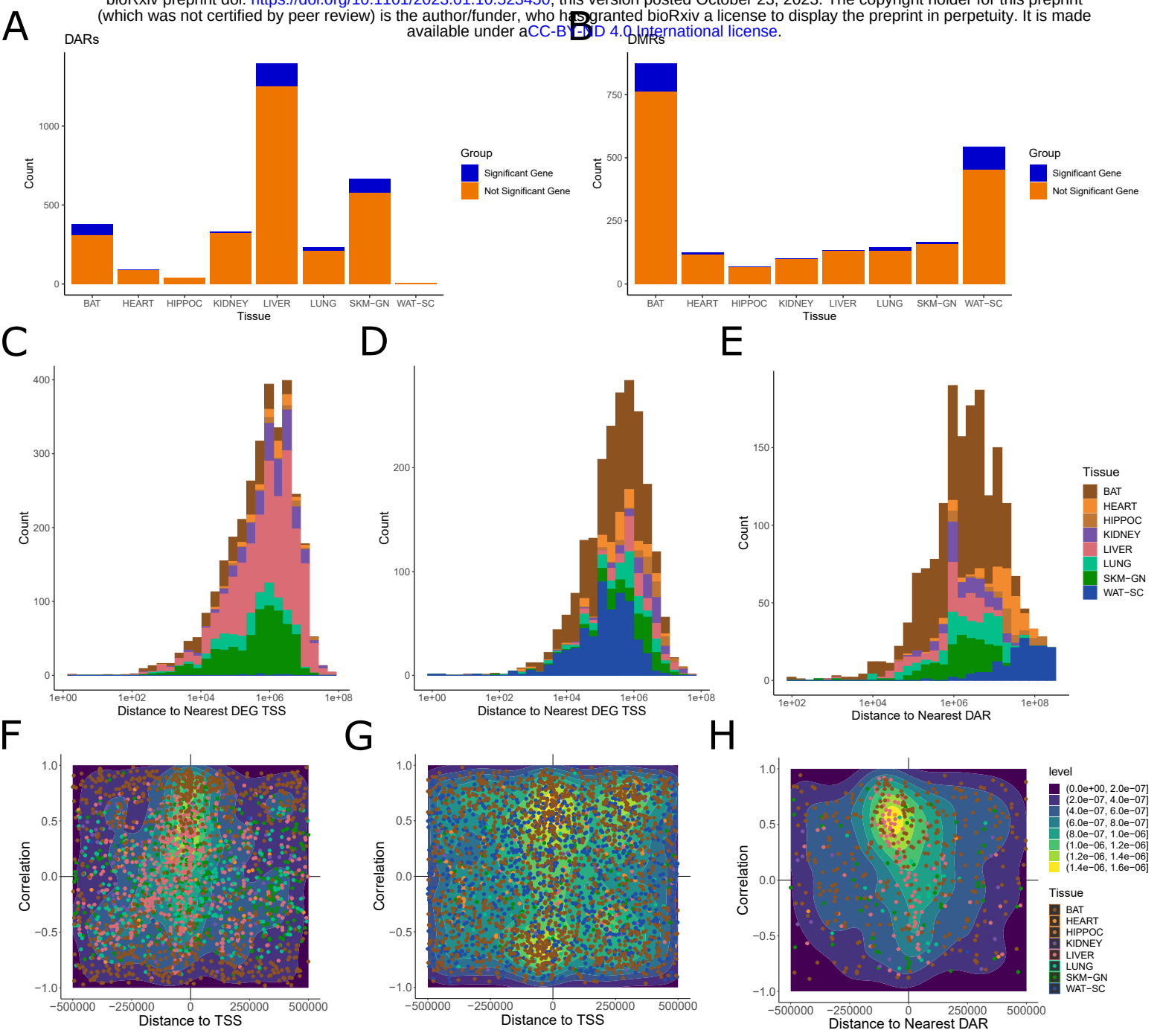

Figure 3: Responses in chromatin accessibility and DNA methylation may not directly link with the expression of the closest genes. (a) Count of DARs annotated to a DEG or a non-DEG. The closest gene to most DARs is not a DEG. * reflects $p < 0.05$ for hypergeometric test measuring the significance of the DAR-DEG overlap. (b) Count of DMRs annotated to a DEG or a non-DEG. Similar to DARs, the closest gene to most DMRs is not a DEG. * reflects $p < 0.05$ for hypergeometric test measuring the significance of the DMR-DEG overlap (c-e) Distributions of distance between DARs and nearest DEG TSS (c), DMRs and nearest DEG TSS (d), and DMRs and nearest DAR (e). DARs (c) and DMRs (d,e) are colored by tissue. (f-h) Density scatter plots of DAR-DEG training response correlation vs. distance (f), DMR-DEG training response correlation vs distance (g), and DMR-DAR training response correlation vs distance (h). DARs with high positive correlation to gene expression enriched for TSS-proximal regions in most tissues while DMRs with high positive and negative correlation to gene expression enriched for TSS-proximal regions. DAR-DMR correlations tended more positive when DMRs were upstream of DARs, and tended more negative when DMRs were downstream of DARs.

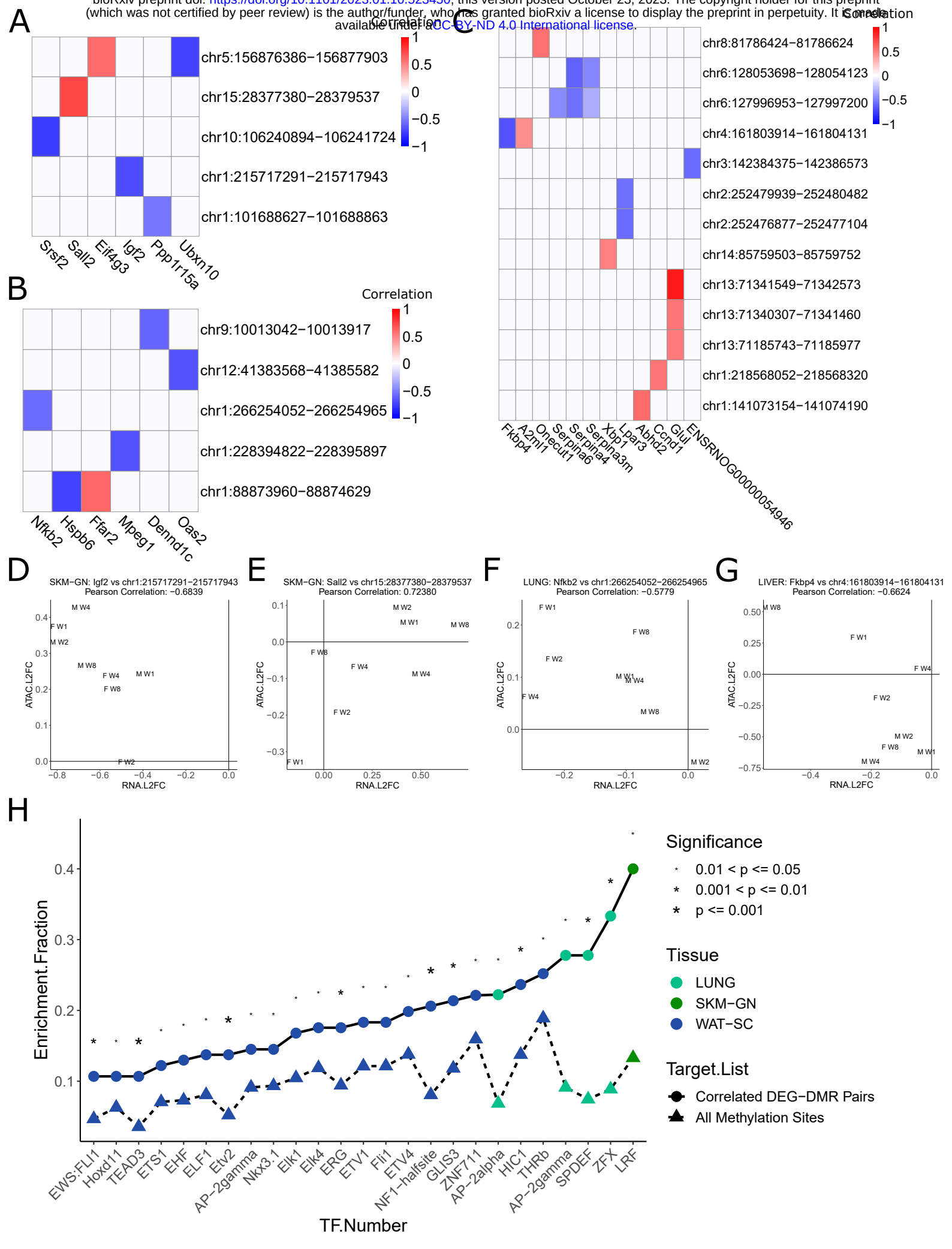

Figure 4: Responsive transcription factors in correlated DAR-DEG pairs. (a-c) Correlation heatmaps of correlated DAR-DEG pairs with binding sites for enriched TFs Maz in SKM-GN (a), and lung (b), and Smad3 in liver (c). (d-g) Training response L2FC scatter plots of correlated DAR-DEG Maz and Smad3 targets. Maz-target DARs in SKM-GN correlated with Igf2 (d), and Sall2 (e), and in lung correlated with Nfkb2 (f). SMAD3-target DAR in liver correlated with Fkbp4 (g).

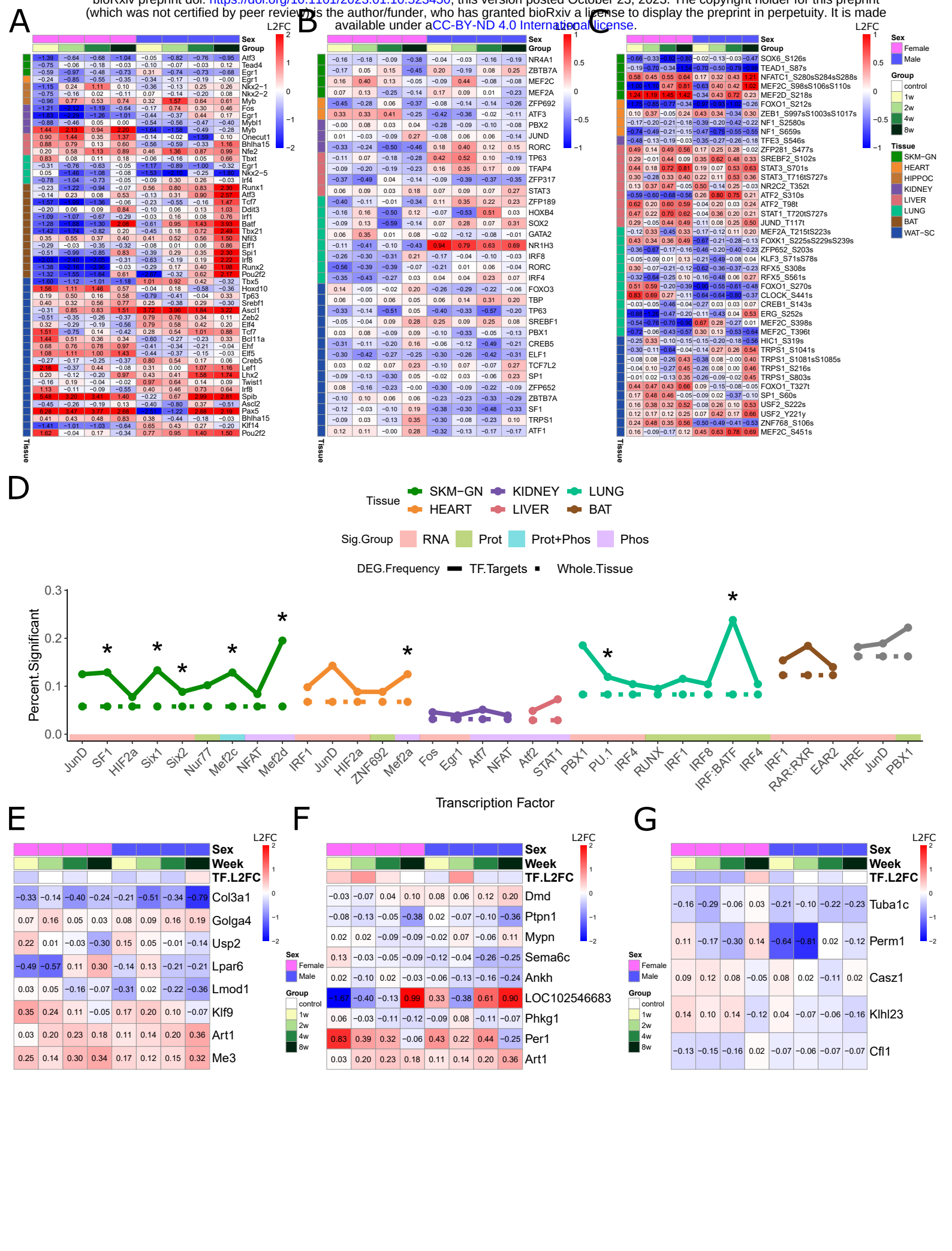

Figure 5: TFs showed significant EET responses at multiple omic levels. (a-c) Significant TF training responses at the transcriptomic (a), proteomic (b) and phosphoproteomic (c) levels. All TFs satisfy p-value < 0.1 in the specified ome and L2FC values are displayed at each time point and sex. Tissue of significance is color-annotated on the left side of each row. (d) TFs with significant changes at the transcriptomic, proteomic and phosphoproteomic levels and whose proximal promoter motif targets are enriched for DEGs are shown. Significant enrichments ($p < 0.05$) are represented with a *. Solid lines reflect the frequency of TF proximal promoter motif targets to be DEGs while the dashed lines show the frequency of DEGs among expressed genes in each tissue. Lines are colored by the tissue of TF training response and target DEG enrichment and TFs are individually colored by the source of their TF training response (RNA, Protein, Protein and Phosphoprotein, and Phosphoprotein). (e-g) Examples of TFs with significant enrichments for DEGs among proximal promoter motif targets. Heatmaps display the targets for the TFs Six1 in SKM-GN (e), MEF2C in SKM-GN (f), and IRF:BATF in Lung (g). For each target, the L2FC across time points and sexes is displayed as well as the L2FC for the TF itself.

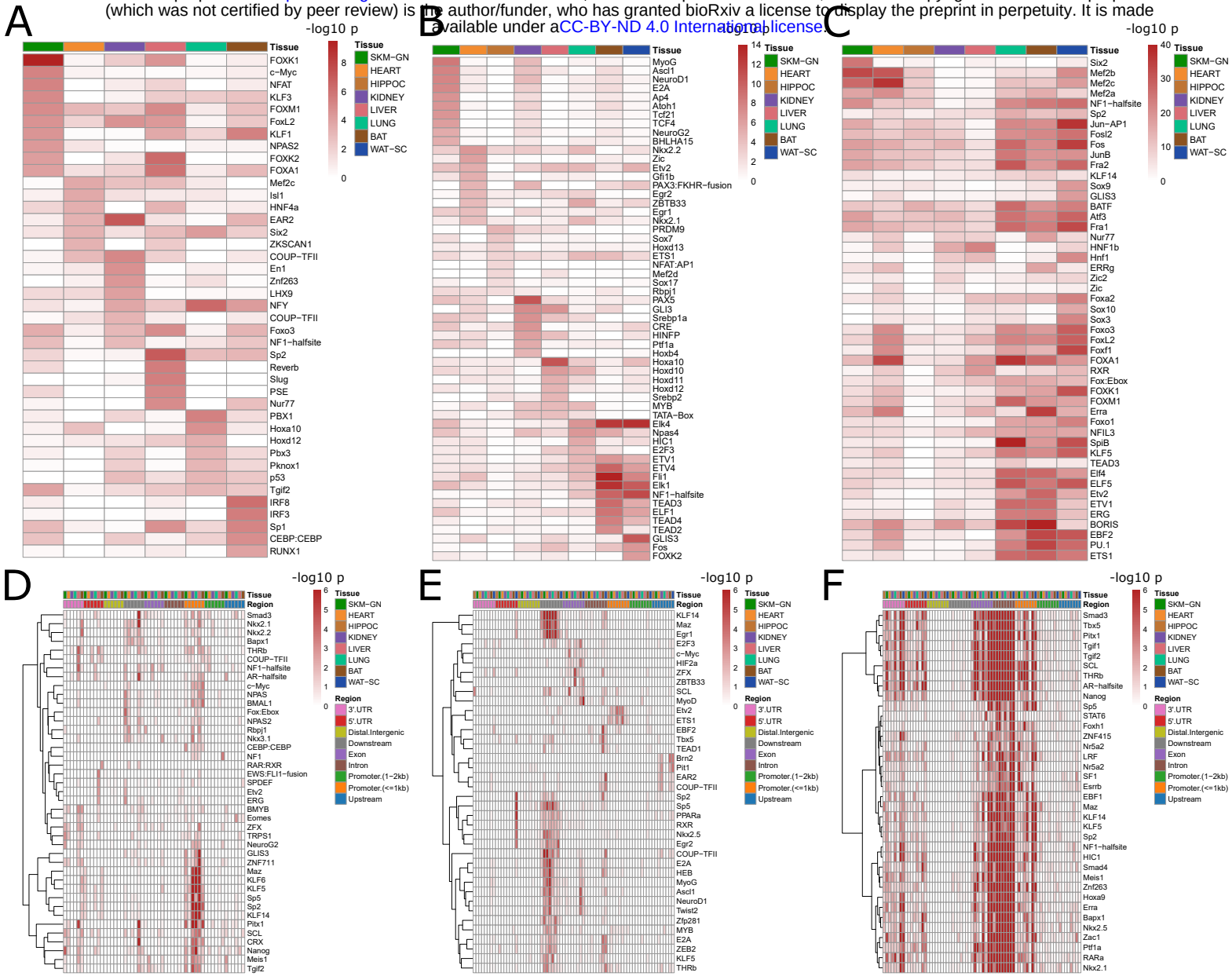

Figure 6: TF motif enrichment at the tissue level and at the genomic feature level in each tissue. (a-c) -log10 p-value of TF motif enrichment in tissue DARs (a), DMRs (b), and DEGaPs (c). Top enriched TFs were selected for each tissue in (a-c). (d-f) -log10 p-value of TF motif enrichment among sets of DARs (d), DMRs (e), and DEGaPs (f) split by genomic feature. Distinct sets of TF motifs are significantly enriched in proximal promoter regions (<=1kb from TSS) and downstream regions, shared by multiple tissues for DARs and DMRs, while TFs are most heavily enriched among intronic DEGaPs in most tissues.

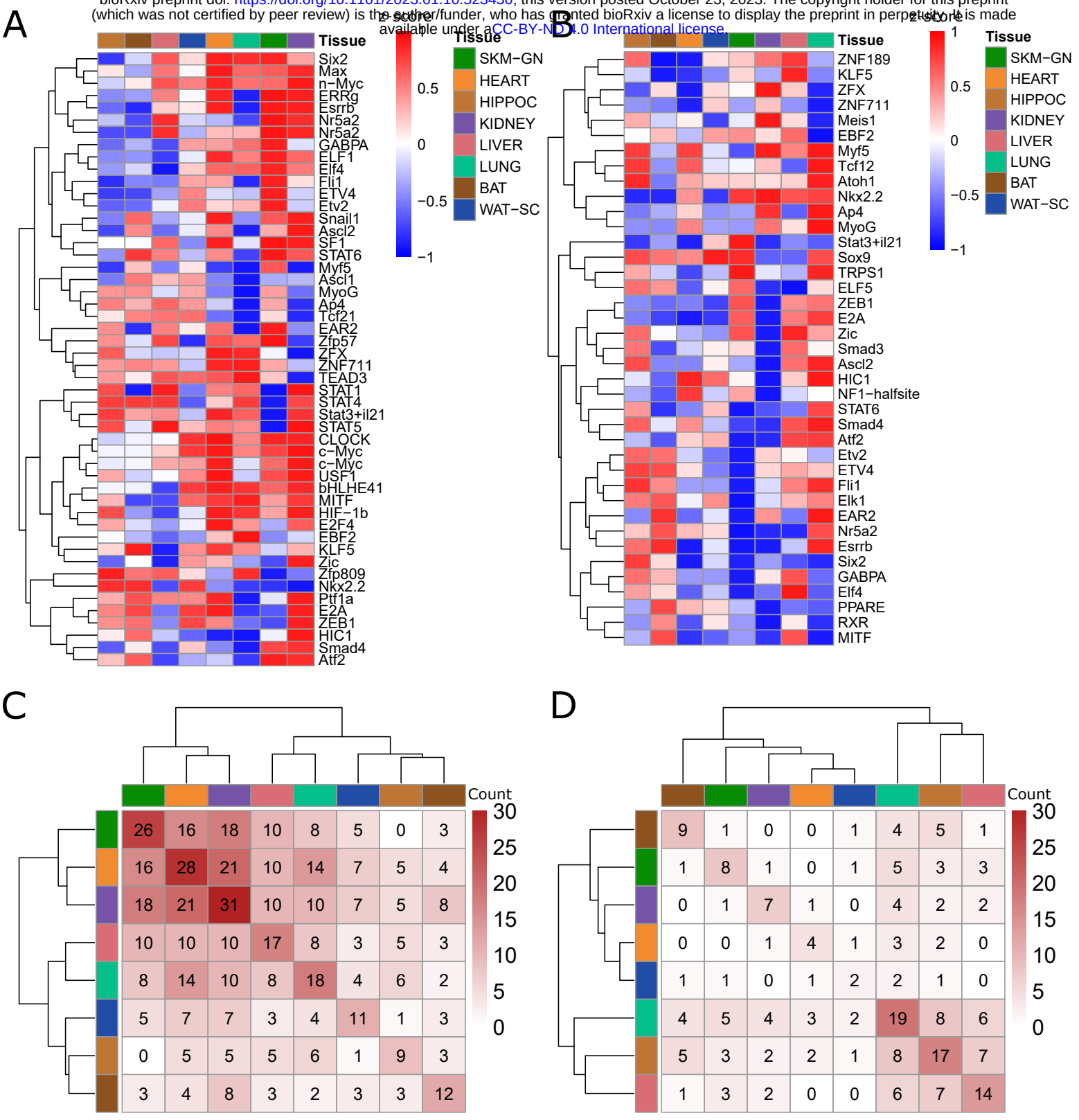

Figure 7: Tissues shared sets of TFs enriched in the promoter regions of genes that were up- or down-regulated after 8 weeks of training. (a-b) Heatmaps of relative enrichment of TFs among the promoter peaks of up- (a) or down-regulated genes (b) across tissues. (c-d) Number of enriched TFs shared between pairs of tissues in the promoter peaks of up- (c) or down-regulated genes (d).

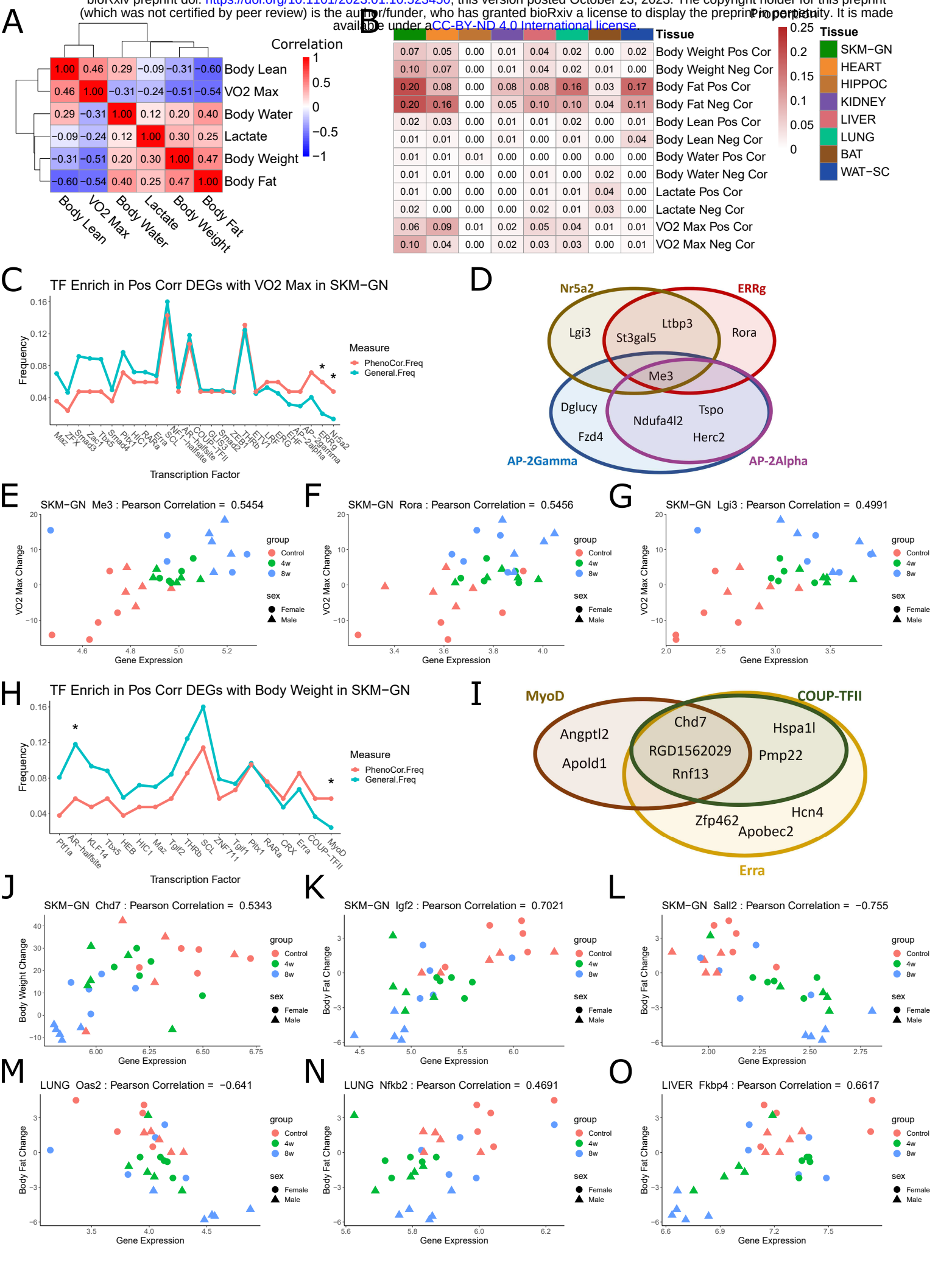

Figure 8: TF enrichment among phenotype-correlated DEGs. (a) Heatmap of Pearson correlation between phenotypic measures. Strong relationships between body lean and VO2 max and body fat and weight gain. (b) Frequency of tissue DEGs positively (> 0.5) or negatively (< -0.5) correlated with each phenotypic 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).

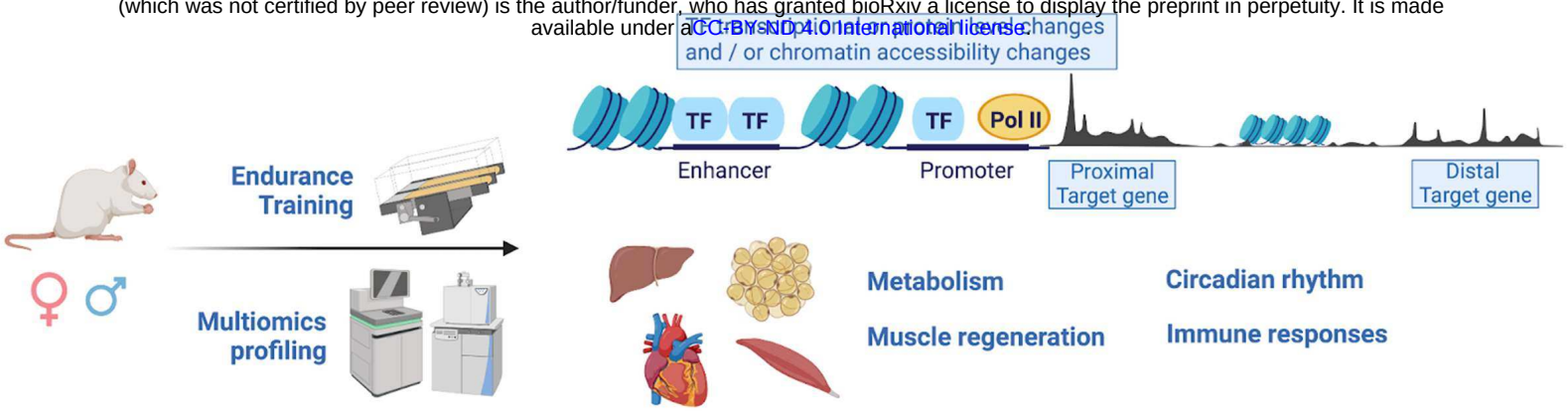

Figure 9: Transcription factors regulate exercise training-induced gene expression via multiple methods. Following eight weeks of endurance training, multiomic analysis across eight tissues have noted gene regulation through direct proximal promoter DAR to DEG relationships, DAR to distant correlated 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.