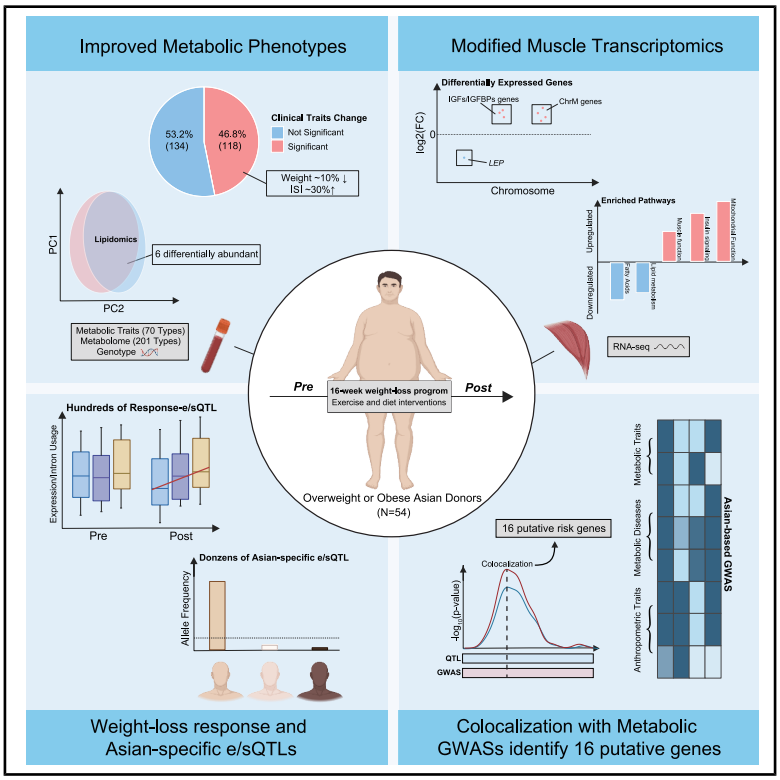
Cell Genomics

Article

Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues

Graphical abstract



Highlights

- Conducted a 16-week lifestyle intervention in 54 East Asians with deep phenotyping
- Lifestyle reshapes muscle transcriptome via mitochondrial and insulin-related pathways
- Hundreds of lifestyle-responsive and dozens of Asian-specific e/sQTLs identified
- e/sQTL colocalization with Asian GWASs reveals 16 metabolic risk gene candidates

Authors

Wenjing Wang, Wei Lin Liew, Shiqi Huang, ..., Eric Yin Hao Khoo, Mei Hui Liu, Boxiang Liu

Correspondence

fstlmh@nus.edu.sg (M.H.L.), boxiangliu@nus.edu.sg (B.L.)

In brief

Wang et al. conducted a 16-week lifestyle intervention in obese East Asians and profiled paired skeletal muscle transcriptomes. They identified hundreds of lifestyle-responsive and Asian-specific eQTLs and sQTLs and linked regulatory variants to metabolic traits through colocalization with GWAS, highlighting gene-lifestyle interactions in human skeletal muscle.

Wang et al., 2025, Cell Genomics *5*, 100951
September 10, 2025
https://doi.org/10.1016/j.xgen.2025.100951

CellPress

Cell Genomics

CellPress
OPEN ACCESS

Article

Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues

Wenjing Wang,[1] Wei Lin Liew,[2] Shiqi Huang,[2] Edmund Chan,[2] Amelia Li Min Tan,[2,8] Chi Tian,[1] Yihan Tong,[1] Yuntian Zhang,[3] Fei Liu,[1] Yixian Qin,[1] Sean Jun Leong Ou,[2,4] Suresh Anand Sadananthan,[5] Sambasivam Sendhil Velan,[5,6] Kavita Venkataraman,[7] Sarah R. Langley,[8,9] Petretto Enrico,[8] Shawn Hoon,[10] Kwang Wei Tham,[11] Yap Seng Chong,[5,12] Yung Seng Lee,[5,13] Melvin Khee-Shing Leow,[5,8] Xueling Sim,[7] Chin Meng Khoo,[2,14] E. Shyong Tai,[2,8,14] Eric Yin Hao Khoo,[2] Mei Hui Liu,[4,*] and Boxiang Liu[1,3,15,16,17,18,19,*]

[1]Department of Pharmacy and Pharmaceutical Sciences, Faculty of Science, National University of Singapore, Singapore 117559, Singapore
[2]Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore
[3]Department of Biomedical Informatics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore
[4]Department of Food Science and Technology, National University of Singapore, Singapore 117543, Singapore
[5]Institute for Human Development and Potential (IHDP), Agency for Science, Technology and Research (A*STAR), Singapore 117609, Singapore
[6]Human Magnetic Resonance Center (hMRC), Institute for Applied Life Sciences, University of Massachusetts, Amherst, Amherst, MA 01003, USA
[7]Saw Swee Hock School of Public Health, National University of Singapore and National University Health System, Singapore 117549, Singapore
[8]Duke-National University of Singapore Medical School, Singapore 169857, Singapore
[9]School of Bioscience, Cardiff University, Cardiff CF10 3AX, UK
[10]Molecular Engineering Laboratory, IMCB, Agency for Science, Technology and Research (A*STAR), Singapore 138673, Singapore
[11]Department of Endocrinology, Woodlands Health, National Healthcare Group, Singapore 737628, Singapore
[12]Department of Obstetrics and Gynaecology, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore
[13]Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 119228, Singapore
[14]Division of Endocrinology, Department of Medicine, National University Health System, Singapore 119228, Singapore
[15]Genome Institute of Singapore (GIS), Agency for Science, Technology and Research (A*STAR), Singapore 138672, Singapore
[16]Precision Medicine Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117596, Singapore
[17]NUS Centre for Cancer Research, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117599, Singapore
[18]Cardiovascular-Metabolic Disease Translational Research Programme, Yong Loo Lin School of Medicine, National University of Singapore, Singapore 117599, Singapore
[19]Lead contact
*Correspondence: fstlmh@nus.edu.sg (M.H.L.), boxiangliu@nus.edu.sg (B.L.)
https://doi.org/10.1016/j.xgen.2025.100951

SUMMARY

Weight loss through exercise and diet reduces the risk of type 2 diabetes, but the genetic regulation of gene expression and splicing in response to weight loss remains unclear in humans. We collected clinical data and skeletal muscle biopsies from 54 overweight/obese Asian individuals before and after a 16-week lifestyle intervention, which resulted in an average of ~10% weight loss, accompanied by an ~30% increase in insulin-stimulated glucose uptake. Improvements were observed in 118 of 252 clinical traits and six blood lipids. Transcriptomic analysis of paired skeletal muscle biopsies identified 505 differentially expressed genes enriched in mitochondrial function and insulin sensitivity. Thousands of muscle-specific expression/splicing quantitative trait loci (e/sQTLs) were detected pre- and post-intervention, including hundreds of lifestyle-responsive e/sQTLs. Notably, approximately 4.2% of eQTLs and 7.3% of sQTLs showed Asian specificity. Joint analysis with genome-wide association study (GWAS) identified 16 putative metabolic risk genes. Our study reveals gene-by-lifestyle interactions and how lifestyle modulates gene regulation in skeletal muscle.

INTRODUCTION

Type 2 diabetes (T2D) is influenced by a complex interplay of genetics, environment, and lifestyle factors. Genetic predisposition accounts for 25%–75% of susceptibility,[1,2] and genome-wide association studies (GWASs) have identified thousands of genetic loci associated with T2D.[3–6] While genetic predispositions are largely non-modifiable, lifestyle modifications as a

Please cite this article in press as: Wang et al., Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.100951

prevention or treatment for T2D are under active investigation.[7] Obesity is a well-known risk factor for T2D, and substantial evidence indicates that weight loss through a combination of exercise and dietary adjustments is the most effective strategy for T2D prevention.[8,9] These lifestyle interventions have shown marked improvements in β-cell function, insulin sensitivity, and fasting blood glucose.[10,11] However, the effectiveness of lifestyle intervention varies across individuals, suggesting that genetic factors may modulate the response to lifestyle changes.[12] Gene-by-lifestyle (GxL) interactions are critical in human health, shaping individual susceptibility and treatment responses to complex diseases like T2D.[13–15] Previous human studies have revealed that genetic risk scores can predict individual responses to lifestyle interventions[16] and interventions can influence gene regulation through epigenetic modifications,[17] suggesting a potential molecular mechanism through which lifestyle factors influence disease risk. However, the tissue-specific molecular mechanisms underlying GxL interactions, particularly in key metabolically active tissues that drive obesity and T2D pathogenesis, remain poorly understood.

The interactions between genetics and lifestyle have been explored in animal models such as rats and baboons, showcasing the potential for understanding how exercise and diet might influence genetic regulation.[18–21] Nair et al.[18] studied the transcriptional impacts of an 8-week endurance exercise training program in Fischer 344 rats, providing insights into how controlled physical activity can affect gene expression and molecular traits within a homogeneous genetic background. However, the lack of genetic diversity in such models poses significant limitations on their applicability to the genetically diverse human population. Lin et al.[21] investigated the effects of a 2-year high dietary fat and cholesterol intervention on genetic expression in baboons. Baboons offer a closer physiological resemblance to humans, potentially making the insights gained more relevant to human health. However, despite this similarity, the translation of findings from baboons to humans is not straightforward due to species-specific differences in metabolism and genetic responses, as well as the complexity of human-environmental interactions.[22] The current lack of human studies is partially due to difficulties in cohort collection. Two primary challenges remain. First, a substantial proportion of participants will drop out at various stages during the lifestyle intervention.[23] Second, for participants who complete the study, unlike in captive animal studies, standardized exercise and dietary intervention on humans are confounded by low compliance and intrinsic variability in incumbent lifestyles.[24] While there is a clear need to understand the molecular mechanisms underlying transcriptional changes, including gene expression and alternative splicing, in response to lifestyle modifications, these challenges have precluded such studies in humans.

Skeletal muscle plays a crucial role in insulin resistance, making it a key tissue for understanding T2D pathophysiology.[25,26] In this study, we utilize matched skeletal muscle biopsies from overweight or obese individuals before and after a 16-week lifestyle intervention to investigate the effect of dietary and exercise-induced weight loss on the skeletal muscle transcriptome. Paired with genotypes, we further investigate the effects of exercise and diet on the genetic effects of molecular traits in skeletal muscles and link them to T2D GWAS signals. Following the intervention, participants experienced ∼10% weight loss, accompanied by ∼30% improvement in insulin-stimulated glucose uptake. Significant changes were observed in 118 out of 252 clinical traits. RNA-sequencing (RNA-seq) analysis revealed 505 differentially expressed genes enriched in mitochondrial and insulin signaling pathways. In expression quantitative trait loci (eQTL) analysis, We identified 1,217 and 1,105 eGenes in the pre- and post-intervention conditions, of which 621 were shared, while 596 were pre-intervention-biased and 484 were post-intervention-biased. In splicing quantitative trait loci (sQTL) analysis, 547 and 347 sIntrons were identified in the pre- and post-intervention conditions, of which 263 were shared, 284 were pre-intervention-biased, and 84 were post-intervention-biased. Furthermore, we discovered dozens of Asian-specific e/sQTLs. Colocalization analysis between our e/sQTLs and metabolism-associated GWAS loci revealed 16 putative risk genes. Among these, we identified a pre-biased and Asian-specific eQTL for *DHRS4L2* that showed strong colocalization with T2D. The genetic effect of *ANK1* on T2D susceptibility was substantially attenuated following intervention. Our findings provide insights into how lifestyle interventions modulate genetic regulation in skeletal muscle, with implications for personalized T2D prevention and treatment strategies.

RESULTS

The Singapore Adult Metabolism Study lifestyle intervention cohort

The Singapore Adult Metabolism Study – phase 2 (SAMS2) intervention program recruited 265 overweight or obese individuals (BMI range: 23 to 35) of East Asian descent to study the effects of lifestyle modification on obesity (Figure S1). As BMI cutoffs for overweight and obesity differ across ethnic groups, we followed the classification for Asian populations (BMI $\geq$ 23 kg/m^2 as overweight and $\geq$ 27.5 kg/m^2 as obese) in our recruitment criteria.[27] After initial screening, 189 eligible participants enrolled and completed baseline measurements. The study progressed through four phases: Phase 1 was a run-in period, during which participants were instructed to stabilize their physiological measurements and reach baseline levels in daily activities. A total of 152 participants remained after phase 1. Phase 2 was a 16-week diet and exercise intervention, after which 132 participants remained. We took post-intervention assessments in phase 3, in which many participants dropped out due to aversion to repeated skeletal muscle biopsies, and 54 participants remained. Phase 4 consisted of a 12-month weight maintenance period (n = 54). We performed comprehensive pre- and post-intervention clinical and molecular measurements on the 54 participants who completed the entire program (Figure 1A): Clinical measurements included anthropometrics, 24-h food recalls, indirect calorimetry, fasting clinical evaluations, insulin response, magnetic resonance imaging (MRI)-based estimates of liver, muscle, abdominal fat, and dual-energy X-ray absorptiometry (DEXA) scans. Molecular measurements included RNA-seq profiling of skeletal muscle biopsies, genotyping, and lipidomics profiling from peripheral blood.

Please cite this article in press as: Wang et al., Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.100951

CellPress
OPEN ACCESS

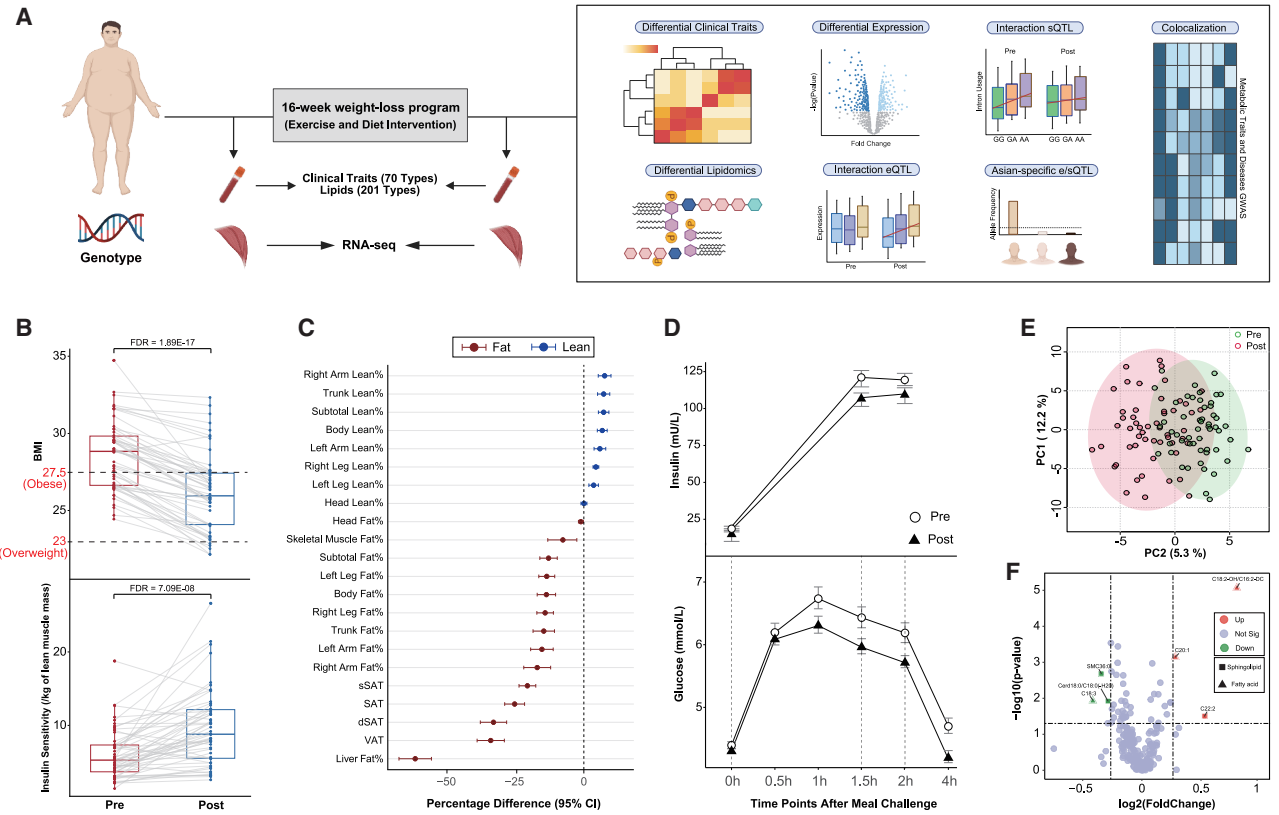


Figure 1. Overview of SAMS2 study and its outcomes

(A) SAMS2 participants underwent a 16-week lifestyle intervention program consisting of structured exercise sessions and diet restriction. Clinical traits, lipid profiles, skeletal muscle transcriptomics, QTL analysis, and colocalization were conducted both before and after the intervention.

(B) Boxplots showing significant changes in body mass index (BMI) and insulin sensitivity index (ISI) due to the intervention. Boxplots indicate the median and interquartile range. Red boxplots represent the pre-intervention while blue boxplots represent the post-intervention. According to the Asian BMI classification, a BMI greater than 23 is considered overweight, and a BMI greater than 27.5 is classified as obese. These two thresholds are marked with red dashed lines.

(C) Forest plot illustrating the changes in the percentage of fat depots and lean muscle mass across various body regions. Effect sizes and their 95% confidence intervals are shown. Significance is determined if the confidence interval does not cross zero.

(D) Graphs displaying reductions in fasting glucose and insulin levels measured by an *ad libitum* meal challenge after the intervention. Bars represent mean ± standard deviation (SD). Open circles represent the pre-intervention condition, and solid triangles represent the post-intervention condition.

(E) PCA plot displaying the separation of blood lipid profiles before (red) and after (green) the intervention. This analysis revealed a variance of 5.3% attributable to the intervention.

(F) Volcano plot showing the differentially expressed lipid species pre- and post-intervention. Lipid species demonstrating significant upregulation or downregulation are highlighted with red and green to indicate changes in fatty acid and sphingolipid metabolism. Significance was defined as a p value < 0.05 and an absolute log2 fold change $\geq$ 1.2.

During the 16-week intervention, participants attended an average of 37.3 ± 11.5 (out of 48) structured exercise sessions (Figure S2A). With dietary intervention, we observed a significant decrease in energy, carbohydrate, cholesterol, and fat intake (Figure S2B; Table S1), without a significant change in protein intake (false discovery rate [FDR] = 0.098), consistent with our interventional objective to reduce weight while maintaining muscle mass. We observed weight loss in 98% of participants, with an 8.7% ± 4.4% loss (FDR = 4.47×10^{-14}; range: 1.1%–23.6%) from their initial body weight, and 84% of participants experienced improved insulin sensitivity (Figure 1B; FDR = 7.09×10^{-8}). MRI and DEXA imaging scan data revealed whole-body fat reduction from multiple fat depots, where the percentage decrease in liver fat was highest among all depots (Figure 1C). In contrast, lean muscle mass percentage significantly increased across various body regions (FDR < 0.05). Consistent with MRI and DEXA scans, anthropometric measurements, such as waist-hip ratio (WHR), also showed significant decreases (FDR = 3.50×10^{-6}; Figure S3). The reduction in body weight can be attributed to the loss of whole-body fat (FDR = 1.64×10^{-14}), accounting for 65.4% ± 29.3% of total weight loss (Table S2; Figure S4). Additionally, the decrease in fat percentage, particularly liver fat, shows a significant negative correlation with the increase in insulin sensitivity index (FDR = 7.05×10^{-3}; Figure S4). Overall, out of the 252 measured clinical traits, 118 showed significant differences before and after the intervention (FDR < 0.05; Table S2). During the 12-month weight maintenance phase, the body weight (Figure S5A) and BMI (Figure S5B) of the 54 participants showed a slight regain but remained significantly lower than baseline

Table 1. Summary statistics of eight key phenotypes measured in 54 participants at three time points

	Pre		Post		Post WMP	
Clinical variables	Mean	SD	Mean	SD	Mean	SD
Weight (kg)	86.81	8.06	79.14	9.00	83.42	10.42
BMI (kg/m^2)	28.50	2.08	26.21	2.55	27.51	2.73
Fasting plasma glucose (mmol/L)	4.61	0.35	4.50	0.27	4.75	0.40
Insulin at 0 h (mU/L)	15.18	8.26	9.58	4.89	13.32	9.60
HOMA-IR	3.05	1.69	1.93	1.01	2.93	2.13
ISI clamps (/kg of lean muscle mass)	5.66	2.43	9.39	4.73	NA	NA
Visceral adipose tissue (cm^3)	2258.85	738.59	1518.68	697.51	1718.17	728.32
Total whole-body fat (%)	28.34	4.01	24.46	4.29	25.46	4.70

These clinical phenotypes are associated with the effects of a lifestyle intervention and reflect key metabolic and physiological responses. Values represent the mean and standard deviation (SD) for each phenotype at each time point.

($P < 1 \times 10^{-4}$; Figures S5C and S5D), which underscored the long-term efficacy of the program (Table 1).

Enhanced insulin sensitivity and metabolic health following lifestyle intervention

To further confirm the intervention's effectiveness, we assessed glucose and insulin dynamics via an *ad libitum* meal test and hyperinsulinemic-euglycemic clamp. After intervention, participants showed significantly lower postprandial glucose and insulin levels (FDR < 0.05; Figure 1D). Both the glucose disposition index (DI) and insulin sensitivity index (ISI; per kg lean muscle mass) increased significantly (FDR = 2.00×10^{-7} and FDR = 7.09×10^{-8}, respectively). Correspondingly, the Homeostasis Model Assessment of Insulin Resistance (HOMA-IR) decreased from 3.05 to 1.93 after intervention (FDR = 1.03×10^{-5}), indicating enhanced insulin metabolism. Elevated plasma citrate levels (FDR = 1.56×10^{-3}) suggested enhanced triboxylic acid (TCA) cycle activity and mitochondrial energy output.[28] This improved anaplerotic response likely reflects the adaptation to meet exercise-induced metabolic demands[29] (Table S2).

Previous studies have demonstrated that dysfunctional lipid metabolism contributes to insulin resistance.[30] Lipidomic profiling showed that 5.3% of the variances in blood lipids were attributed to the intervention (Figure 1E). Out of 201 lipid species analyzed, six were identified as differentially abundant (fold change ≥ 1.2, $p < 0.05$), three were upregulated, and three were downregulated (Figure 1F; Table S3). Upregulated lipids suggested improved lipid handling and reduced inflammation,[31,32] while downregulated species indicated ceramide synthesis, potentially reducing lipotoxicity.[33,34]

Lifestyle intervention-induced muscle transcriptional responses

Following quality control, we retained RNA-seq data from 108 skeletal muscle paired samples (pre- and post-intervention) across 54 participants and 15,658 genes for downstream analyses (Figure S6). We identified 505 differentially expressed genes (DEGs, $\log_2|\text{foldchange}| > 0.3$ and FDR < 0.05) before and after intervention (Figure 2A; Table S4). Of these, 141 genes were downregulated, while 364 were upregulated post-intervention. For example, *ABCB5*, involved in cellular energy homeostasis,[35] was significantly upregulated (log2FC = 2.12, FDR = 6.77×10^{-13}), suggesting enhanced metabolic efficiency. In contrast, *LEP*, encoding leptin, which plays a key regulator of energy homeostasis,[36] was significantly downregulated (log2FC = −0.87, FDR = 0.015), consistent with fat mass reduction.[37] Over-representation analysis (ORA) identified 20 biological pathways significantly enriched with the DEGs (FDR < 0.05; Table S5). These pathways in DEGs fall into five categories: insulin signaling, metabolism, skeletal muscle, diet, and others (Figure 2B). Notably, within insulin signaling pathways, two glucose uptake pathways in skeletal muscle[38,39] show distinct patterns: the insulin-independent AMPK pathway was enriched with downregulated DEGs. In contrast, the insulin-dependent Akt/PKB pathway, particularly via PI3K-Akt signaling, was enriched with upregulated DEGs. Activation of the PI3K-Akt pathway follows insulin receptor stimulation, leading to increased glucose uptake through GLUT4 translocation.[40] This observation corresponds to increased insulin sensitivity and a lowered reliance on energy deprivation stress responses post-intervention.[41,42] Additionally, genes related to insulin metabolism, such as insulin-like growth factors (IGFs) and their binding proteins (IGFBPs)[43] were significantly upregulated (Table S5). Mitochondrial function and energy metabolism pathways were also enhanced, as reflected by notable upregulation of mitochondrial-related genes.

We further employed weighted gene co-expression network analysis (WGCNA)[44] to identify gene co-expression patterns in response to the intervention. Among the identified modules (Figure S7A), the Module Eigengene Gray exhibited the strongest correlation with intervention (correlation = 0.57, $P = 1 \times 10^{-10}$), showing a significant positive correlation with the post-intervention condition (Figure S7B). Network analysis of the Module Eigengene Gray revealed strong co-expression relationships (edge weight > 0.2) among several mitochondrial genes, including *MT-ND2*, *MT-ND5*, *MT-ND6*, and *MT-ATP8* (Figure S7C; Table S6). These genes are essential components of the mitochondrial electron transport chain (ETC) complex I[45] (Figure S8). They play a critical role in the oxidation of NADH to NAD^+, facilitating the release of electrons necessary for driving ATP synthesis.[46] Additionally, substantial evidence from the previous study indicated that increased levels of NAD+ are associated with enhanced insulin sensitivity and improved mitochondrial function.[11]

Please cite this article in press as: Wang et al., Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.100951

CellPress
OPEN ACCESS

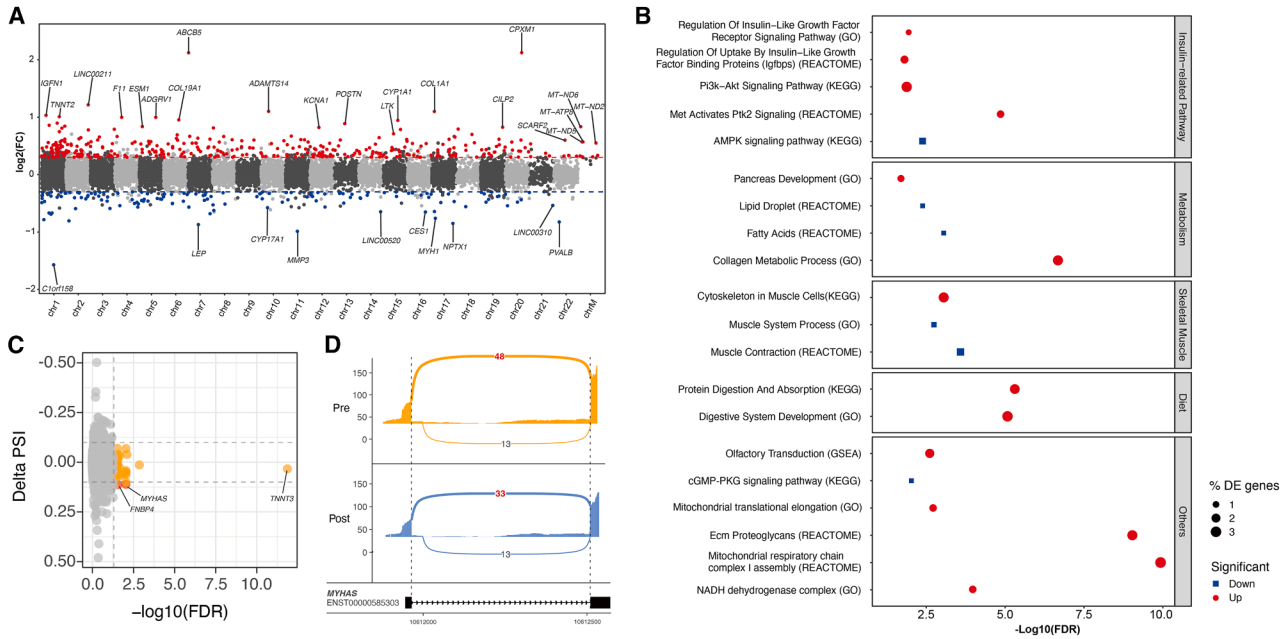


Figure 2. Gene expression changes in mitochondrial function and insulin signaling pre- and post-intervention

(A) Manhattan plot displaying the distribution of differentially expressed genes (DEGs) across chromosomes in skeletal muscle samples after the intervention. Significant DEGs (log2|fold change| > 0.3, FDR < 0.05) are highlighted in red.

(B) Bubble plot shows the top 20 significantly enriched pathways from ORA of differentially expressed genes identified in (A). Significance was determined using an FDR < 0.05. The pathways are categorized into insulin-related pathways, metabolism, skeletal muscle, diet, and others.

(C) Volcano plot showing the relationship between statistical significance (−log10 [FDR]) and the magnitude of splicing changes (Delta PSI) across all detected splicing events. Gray dots represent non-significant events, while orange dots indicate splicing events with FDR < 0.05, and red dots indicate splicing events with FDR < 0.05 and |delta PSI| > 0.1.

(D) Sashimi plot illustrating the alternative splicing pattern of *MYHAS* before and after exercise. The upper and lower panels show the junction reads in pre- and post-intervention, respectively. Numbers indicate junction read counts. The gene structure is shown at the bottom with exons represented as boxes and introns as lines. Genomic coordinates are displayed along the *x* axis.

In addition to differential expression, we analyzed differential alternative splicing (AS) before and after intervention. The analysis identified 17 significant AS events (FDR < 0.05; Figure 2C; Table S7). Among these, *MYHAS* (chr17: 10611945-10612459) and *FNBP4* (chr11:47731561-47732537) exhibited larger changes in splicing patterns, with delta absolute percent spliced in (|PSI|) values greater than 0.1. *MYHAS* encodes a muscle-specific antisense RNA that enables primary mRNA binding activity, which is important to be involved in response to muscle activity and skeletal fiber development.[47] Although the magnitude of change was moderate (ΔPSI = 0.11), the sashimi plot reveals a consistent 1.45-fold increase in the splice junction usage following the intervention, specifically affecting the ENST00000585303 transcript abundance (Figure 2D). *FNBP4* is involved in cytoskeletal regulation and cell migration,[48] splicing changes in *FNBP4* may influence skeletal muscle function and adaptation, but further experimental validation is required. Additionally, we observed significant splicing changes in *TNNT3* (chr11: 1926709-1929810), which showed decreased inclusion after the intervention (FDR = 1.39×10^{-12}). As *TNNT3* encodes fast skeletal muscle troponin T, a key regulator of muscle contraction, these splicing changes might reflect exercise-induced adaptations in muscle contractile properties.[49]

Interactions between genetic effects and lifestyle influence gene expression

We obtained genotypes for each individual using the InfiniumOm2-5Exome single-nucleotide polymorphism (SNP) array followed by imputation with 1000G Phase 3 as the reference panel[50] (STAR Methods; Figure S9). To identify the gene expression alterations across genetic backgrounds in skeletal muscle transcriptomes, we utilized RASQUAL[51] to identify *cis* genetic variants (minor allele frequency [MAF] > 5%) that are associated with gene expression. We leveraged gene-level and allele-specific count information to enhance discovery power, accounting for known covariates and latent sources of gene expression variability. Multiple hypothesis testing was conducted for both pre-intervention and post-intervention using a hierarchical procedure in TreeQTL.[52] At FDR < 0.05, we found 1,217 significant *cis*-eGenes pre-intervention (*cis*-eGenes were defined as a gene whose expression is regulated by a genetic variant within ±1 Mb of its transcription start site [TSS]) and 1,105 *cis*-eGenes post-intervention (Table S8). Lead eSNPs were enriched near TSSs (Figure S10). To assess the reproducibility of our identified eQTLs, we conducted a replication analysis with GTEx skeletal muscle using the $\pi 1$ statistic. The $\pi 1$ estimates was 0.81 and 0.76 for the pre- and post-intervention conditions, suggesting the SAMS2 *cis*-eQTLs replicated

Please cite this article in press as: Wang et al., Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.100951

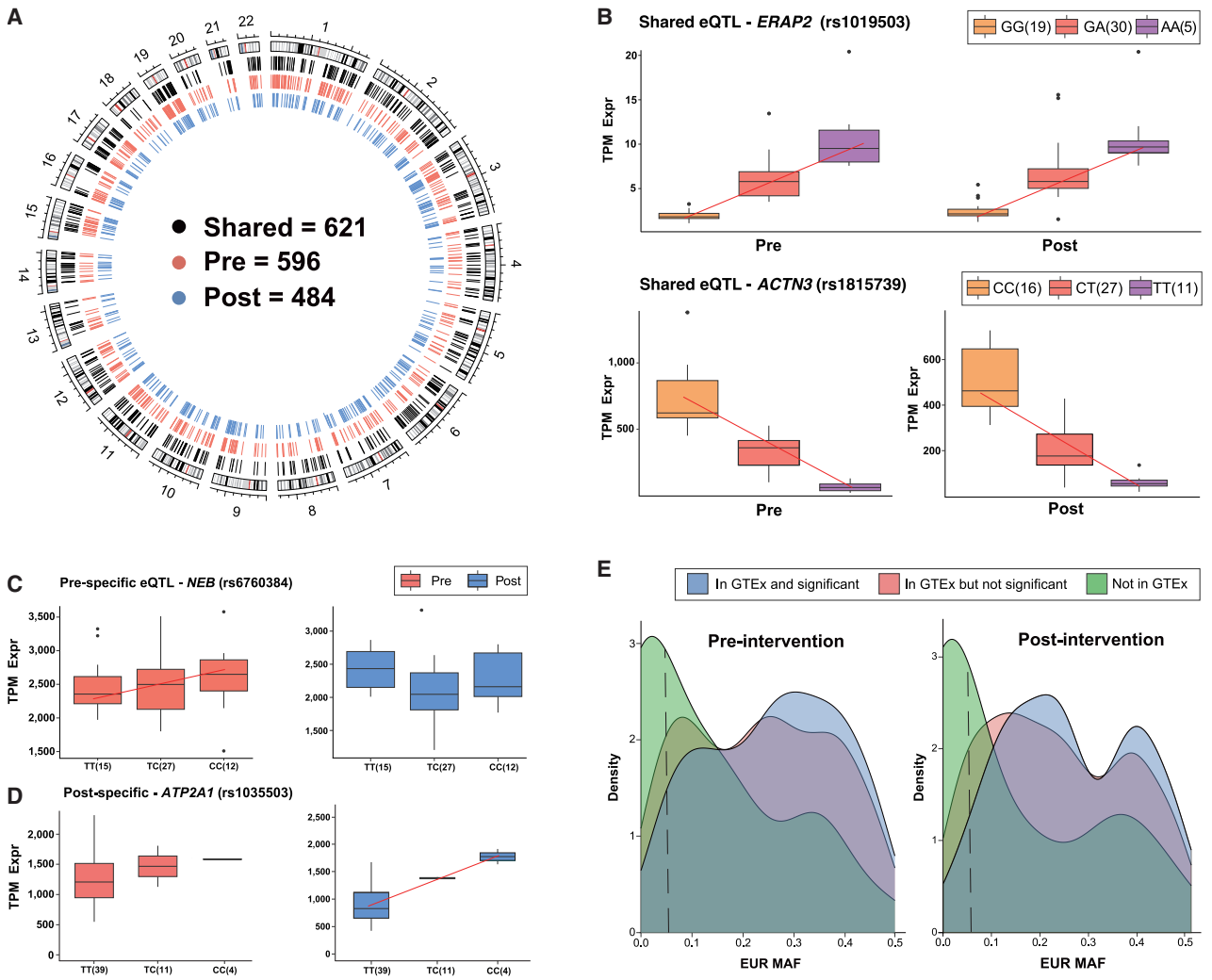


Figure 3. Gene-by-lifestyle interactions reveal differential genetic regulation of gene expression in response to the intervention

(A) The circular plot showing the position of significant eGenes identified pre-biased (red), post-biased (blue), and shared between both conditions (black).

(B) Boxplots depict shared eQTLs in *ERAP2* and *ACTN3*. Boxplots indicate the median and interquartile range.

(C) A pre-biased eQTL in *NEB*. Boxplots indicate the median and interquartile range.

(D) A post-biased eQTL in *ATP2A1*. Boxplots indicate the median and interquartile range.

(E) Density plots illustrate European MAF distribution for eQTLs identified pre- and post-intervention. The plots categorize eQTLs based on the GTEx database, highlighting the Asian-specific eQTL.

well. We further identify lifestyle-responsive genetic effects on gene expression by systematically classifying significant e/sQTLs based on their detection patterns across pre- and post-intervention conditions. Additionally, we applied a linear regression model incorporating a gene-by-lifestyle interaction term to further assess whether shared QTLs exhibited significant interaction effects. Using this approach, we discovered 621 shared eQTLs, 596 pre-biased eQTLs, and 484 post-biased eQTLs (Figure 3A; Table S9).

Shared eQTLs were defined as genetic associations significant both pre- and post-intervention (FDR < 0.05) and exhibited no significant interaction effects between condition and genotype. Their persistence across physiological states may indicate involvement in essential cellular processes or metabolic pathways, and their *cis*-regulatory elements may be unaffected by exercise and dietary weight loss interventions. Two notable examples of shared eQTLs are particularly insightful (Figure 3B). One example was found in the *ERAP2* gene, which involves a shared eQTL ($p = 6.95 \times 10^{-96}$ for pre-intervention and $p = 1.37 \times 10^{-86}$ for post-intervention) with the SNP rs1019503 (chr5:96919113:G:A) previously identified in European populations.[53,54] *ERAP2* is known for its role in amino acid trimming for MHC class I presentation, and the variant rs1019503 has been linked to glycemic response, especially 2-h glucose level measured during an oral glucose tolerance test (OGTT) in previous GWAS studies,[55,56] indicating its potential influence on postprandial glucose metabolism. Another significant shared eQTL for the *ACTN3* ($p = 2.99 \times 10^{-132}$ for pre-intervention

Please cite this article in press as: Wang et al., Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.100951

and $p = 1.50 \times 10^{-96}$ for post-intervention), whose proxy SNP rs1815739 (linkage disequilibrium [LD] > 0.8 with the lead SNP, chr11:66560624:C:T) is exonic.[57] The alternative allele of rs1815739 creates a stop-gained codon, leading to nonsense-mediated delay of this fast twitch fiber-specific gene.[58] The R577X mutation in *ACTN3* has been previously documented for its influence on skeletal muscle function and metabolism.[59] Although *ACTN3* expression significantly decreased after the intervention ($p = 1.16 \times 10^{-5}$), the eQTL for this gene remains present both before and after the intervention with a consistent effect size. The variation rs1815739 in this gene affects muscle performance and has been widely studied in the context of muscle strength, highlighting its importance in muscle physiology.[60,61] We asked whether rs1815739 is correlated with lifestyle intervention outcomes. Initial analysis using an additive genetic model showed a trend toward genotype-dependent weight loss ($p = 0.0673$). Further investigation revealed that the association followed a recessive inheritance pattern, with the recessive model providing a better fit ($p = 0.0202$). Specifically, homozygous carriers of the X allele (TT genotype) showed significantly greater weight loss compared with both RR (CC) and RX (CT) carriers (TT vs. CC: −4.42%, $p = 0.0441$; TT vs. CT: −4.39%, $p = 0.0304$), and no significant difference was observed between RR and RX carriers ($p = 0.9850$) (Figure S11). This enhanced weight loss in TT carriers might be attributed to their altered muscle metabolism, as the absence of α-actinin-3 has been associated with improved oxidative capacity,[62] potentially facilitating more efficient fat utilization during exercise.

We further explored eQTLs that were biased toward either the pre- or post-intervention condition, defined as genetic associations that were significant in one condition but not the other (FDR < 0.05), or those that remained significant in both conditions but exhibited significant genotype-by-lifestyle interaction effects, indicating a bias toward either the pre- or post-intervention state. An example pre-biased eQTL (chr2:151731300:T:C, $p = 1.2 \times 10^{-35}$) was found in the *NEB* gene (Figure 3C), which encodes nebulin and is essential for the proper function and structure of sarcomeres in skeletal muscle fibers.[63] Mutations in this gene are often associated with various forms of myopathies.[64] Similarly, the *NRAP* eQTL (chr10:113657530:G:A, $p = 3.44 \times 10^{-34}$) is also biased toward the pre-intervention condition (Figure S12). The *NRAP* gene encodes the nebulin-related anchoring protein, which interacts with nebulin and is involved in myofibril assembly.[65] Conversely, the emergence of post-biased eQTLs indicates that therapeutic lifestyle changes, including increased exercise and improved diet, could establish new gene expression regulation patterns (Figures 3D and S12). The *ATP2A1* gene is particularly noteworthy, as it plays a critical role in calcium homeostasis within skeletal muscle.[66] *ATP2A1* expression was significantly downregulated following the weight-loss intervention ($p = 1.37 \times 10^{-6}$), reflecting potential physiological adaptations in muscle tissue, such as enhanced efficiency and change in muscle mass. Its eQTL (chr16:27935099:T:C, $p = 7.17 \times 10^{-7}$) emerged exclusively post-intervention, suggesting that the intervention activated *cis*-regulatory elements influencing *ATP2A1* expression.

Existing skeletal muscle transcriptomic and QTL datasets are biased toward European populations.[54,67,68] As such, we investigated the presence of Asian-specific QTLs among the significant eQTLs identified by comparing our East Asian-derived eQTL results and those from the GTEx database, which predominantly comprises European individuals. Our analysis identified 46 Asian-specific eQTLs pre-intervention and 52 post-intervention (Table S10). We hypothesized that these Asian-specific eQTLs might arise from elevated minor allele frequencies of certain SNPs in Asians relative to other populations. To test this hypothesis, we examined the MAF of East Asian-specific eQTLs in European populations using data from the 1000 Genomes Project.[69] We observed that East Asian-derived eQTLs not replicated in the GTEx database exhibited the highest density at MAF < 0.05 in European populations (Figure 3E). Conversely, the MAFs of SNPs within eQTLs validated in the GTEx database correspond to levels typical of population-specific common SNPs. Both observations suggested that a majority of East Asian-specific eQTLs not replicated in the GTEx dataset could be attributed to allele frequency differences between Europeans and East Asian populations.

Genetic effects on alternative splicing in response to lifestyle intervention

Similar to eQTL analysis, we next explored the impact of genetic variation on alternative splicing in skeletal muscle both before and after the intervention. We employed QTLtools[70] for mapping independent splicing quantitative trait loci (sQTLs) at the intron-cluster level. After adjusting for FDR < 0.05, we identified 547 *cis*-sIntrons (defined as an intron whose splicing is regulated by a nearby genetic variant) pre-intervention and 347 sIntrons post-intervention (Table S11). We identified 263 shared, 284 pre-biased, and 84 post-biased sQTLs (Figure 4A; Table S12). Lead sQTL SNPs were enriched near splice sites, introns, and splice regions (Figure S13). We assessed the reproducibility of our sQTLs by replicating within GTEx skeletal muscle sQTLs. The resulting $\pi 1$ estimates were 1 pre-intervention and 0.91 post-intervention, demonstrating high concordance and robust reproducibility of our sQTL findings for both conditions.

We identified several ATP-related sGenes (defined as a gene whose splicing is regulated by a genetic variant within ±1 Mb of the gene), including *ATP5PF*, *ATP5MD*, and *ATP5MC2*, which are key components of the mitochondrial ATP synthase complex (Figure 4B). These genes exhibited significant genetic effects on splicing regulation both before and after intervention (FDR < 0.05, $|\beta| > 1$), suggesting their critical role in ATP metabolism and utilization in muscle tissue.[71] Moreover, we also identified other shared sQTLs in muscle-relevant genes (Figure S14), including *TRDN* intron (chr6:123331929-123352539). *TRDN* encodes triadin, which is essential for excitation-contraction coupling in skeletal muscle through its regulation of calcium release from the sarcoplasmic reticulum.[72] Among these shared sQTLs, we found sGene *ANK1* to be particularly interesting (Figure 4B). *ANK1* has been identified as a T2D risk locus in multiple GWAS studies.[3–5] Through alternative splicing, *ANK1* generates tissue-specific isoforms, with long isoforms predominantly in erythrocytes and short isoforms abundant in muscle tissue.[73] In muscle, the short isoforms represent the primary expressed form.[74] While *ANK1* showed consistently significant sQTLs across both pre- and post-intervention, the strength of

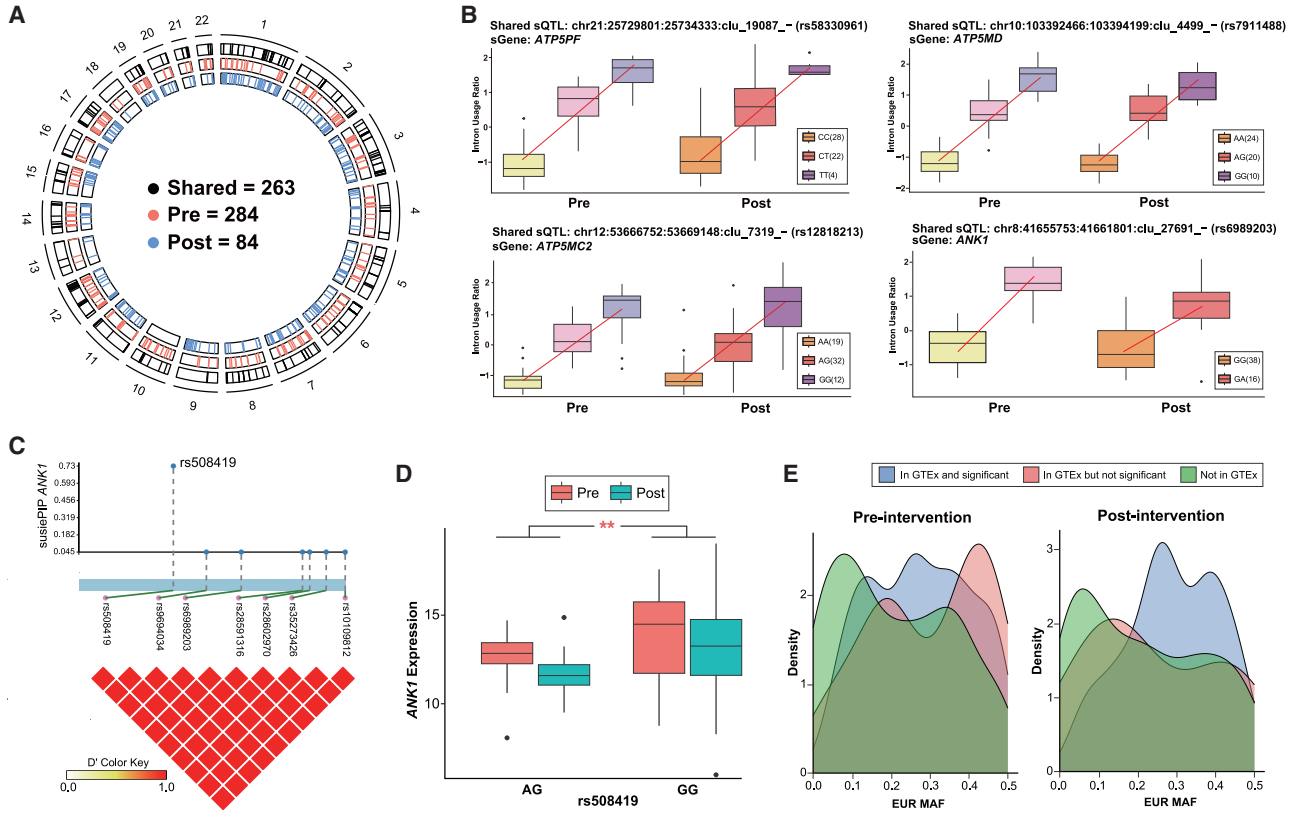


Figure 4. Dynamical sQTLs before and after the intervention

(A) Circular plot showing significant sIntrons identified only pre-intervention, or only in post-intervention and both conditions.

(B) Boxplots depict four shared eQTLs in ATP-related genes and *ANK1* example. Boxplots indicate the median and interquartile range.

(C) Fine-mapping analysis pinpointing the potential causal SNP rs508419 for the sQTL of the *ANK1* gene.

(D) These boxplots present the *ANK1* expression levels across different genotypes of the causal SNP (rs508419). The comparison shows a significant difference in *ANK1* expression between AG and GG genotypes, highlighting the gene-environment interaction effect. Boxplots indicate the median and interquartile range. Statistical significance was assessed using ANOVA. Asterisks denote significance, with $p = 0.01$.

(E) Density plots illustrate European MAF distribution for sQTLs identified pre- and post-intervention. The plots categorize sQTLs based on the GTEx dataset, highlighting the Asian-specific sQTL.

its genetic association was notably attenuated following the intervention (FDR = 0.018). Specifically, we identified that the sQTLs affect *ANK1* intron usage (chr8:41655753-41661801) and further influence the abundance of ENST00000522543, a muscle-specific short isoform. Although the genetic effects of *ANK1* splicing in muscle tissue have been reported,[54,75] no studies have yet explored the dynamic changes in its genetic effect in response to lifestyle modifications. The sashimi plot for the pre-intervention highlights these variations in splicing event frequencies associated with the SNP mutation (rs6989203). Specifically, the mutation from G to A at rs6989203 led to a 2.5-fold increase in sequencing read across the affected intron. After intervention, the genetic effect of rs6989203 was notably reduced, leading to a 1.6-fold increase (Figure S15). The lead SNP rs6989203 of *ANK1* sQTL is in high linkage disequilibrium (LD > 0.95) with several other SNPs. Fine-mapping with SuSiE suggested that rs508419 was the likely causal SNP for this sQTL (posterior inclusion probability [PIP] = 0.73) (Figure 4C). By examining *ANK1* general expression changes pre- and post-intervention according to the genotype of rs508419, we found that individuals with the AG genotype exhibited a more pronounced response to the intervention compared with those with the GG genotype (ANOVA $p = 0.01$; Figure 4D).

In addition to shared sQTLs, we identified context-biased sQTLs that underscore the regulatory changes in skeletal muscle during the intervention (Figure S14). We observed a significant pre-biased sQTL ($p = 2.22 \times 10^{-9}$, $\beta = 0.95$) in *ART3* intron usage (chr4:7607914-76081824) with the lead SNP rs6816425. The splicing event affects an intron that modulates the mono-ADP-ribosyltransferase activity of *ART3*, which plays a crucial role in muscle metabolism and energy homeostasis.[76] Following the intervention, we identified notable post-biased sQTLs. The lead SNP rs11129942 modulated *TRAK1* intron usage (chr3:42115484-42125420; $p = 3.77 \times 10^{-11}$, $\beta = -1.06$). *TRAK1* is essential for mitochondrial trafficking and dynamics, and this splicing affects its interaction with mitochondrial proteins.[77,78] The negative effect size indicates reduced splicing efficiency post-intervention, suggesting a regulated adjustment in mitochondrial dynamics to enhance energy metabolism efficiency during exercise adaptation.

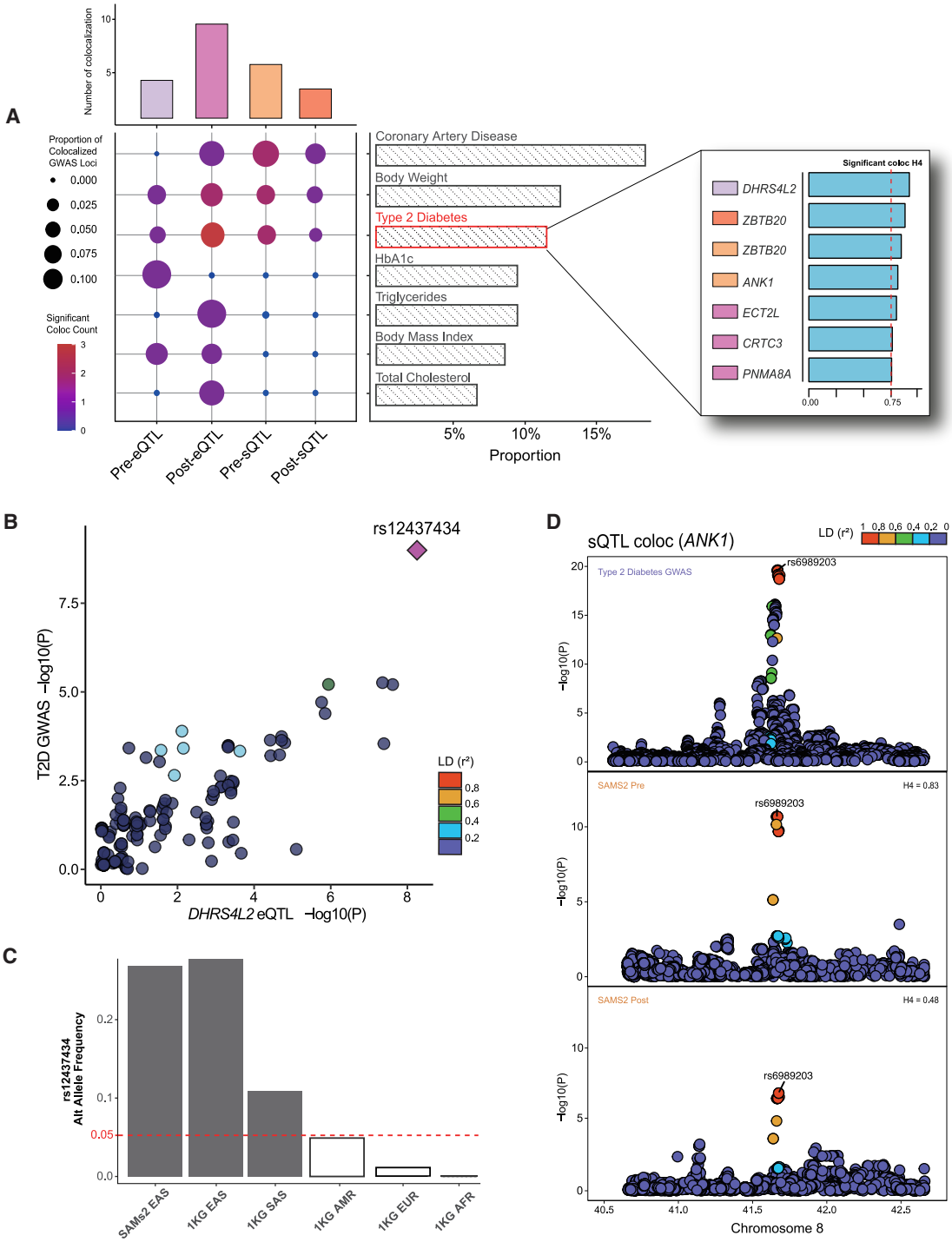


Figure 5. Colocalization of QTLs and metabolic traits or disease

(A) This panel displays the colocalization results of QTLs and GWASs. The *x* axis represents QTLs under different contexts, and the *y* axis represents the 12 metabolic traits or diseases. The size of the circles indicates the proportion of colocalized GWAS loci, while the color gradient (from purple to red) represents the number of significant colocalizations. Significance defined as colocalization posterior probability (PP4 $\geq$ 0.75).

(B) This plot shows the LocusCompare of pre-biased eQTLs for the gene *DHRS4L2* with the SNP rs12437434 from a T2D GWAS. The *x* axis represents the *DHRS4L2* eQTL -$\log_{10}(P)$, and the *y* axis represents T2D GWAS -$\log_{10}(P)$. The linkage disequilibrium (LD) between SNPs and the target SNP is indicated by different colors, representing five levels of correlation.

(legend continued on next page)

Please cite this article in press as: Wang et al., Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.100951

Finally, employing an analogous methodology to identifying Asian-specific eQTLs, we compared our sQTL results against those reported in the GTEx database. This revealed 35 Asian-specific sQTLs pre-intervention and 30 post-intervention. For sQTLs shared between SAMS2 and GTEx, the MAF of the sSNP tended to be higher in European populations. While for sQTLs significant in SAMS2 but not significant in GTEx, the MAF of the sSNP tended to be lower. (Figure 4E; Table S13).

Identifying putative risk genes for metabolism-related traits

To identify putative risk genes underlying metabolism-related diseases and traits, we performed QTL-based gene prioritization using GWAS signals from 12 metabolism-related disease/trait GWAS datasets derived from the East Asian population[4,79,80] (Table S14). These datasets encompass metabolic diseases (T2D and coronary artery disease [CAD]), metabolic traits (glycated hemoglobin levels [HbA1c], blood glucose levels, total cholesterol, high-density lipoprotein [HDL], low-density lipoprotein [LDL], and triglycerides [TG]), and anthropometric traits (BMI, body weight, waist circumference, and hip circumference). Using colocalization analyses to assess overlaps in GWAS and QTL signals (H4 > 0.75) in both pre- and post-intervention conditions,[81] generally, eQTLs exhibited more colocalization events than sQTLs (Figure 5A; Table S15). The e/sQTLs colocalized with metabolic diseases (CAD and T2D) and metabolic traits (such as body weight and HbA1c) (Figure 5A; Table S16). In total, we identified 16 putative target genes (Table 2).

Leveraging summary statistics from Asian T2D GWAS, we identified a significant eQTL-GWAS colocalization locus (H4 = 0.92) involving the gene *DHRS4L2* and lead SNP rs12437434, which is statistically significant only pre-intervention (FDR = 4×10^{-4}) but not post-intervention (FDR = 0.1528; Figures 5B and S16). The lead SNP rs12437434 is in low LD with nearby SNPs, and fine-mapping analysis with SuSiE[82] revealed that it is the most likely causal variant (PIP = 0.96). Notably, rs12437434 is an Asian-biased variant with a higher MAF in East Asian individuals (MAF[SAMS2] = 0.269; MAF[$1KG_{EAS}$] = 0.278) compared with other ethnicities and has an MAF of 0.0109 for European populations (Figure 5C). While the functional mechanism of the *DHRS4L2* gene needs further experimental validation, this finding underscores the importance of population-specific genetic variants in T2D GWAS interpretation and provides insights into the genetic architecture of T2D in Asian populations.

We also found a shared eQTL for *FN3KRP* (with lead SNP rs2256833) that remained significant across both conditions, but its genetic effects exhibited condition-dependent differences (FDR = 0.012). This shared eQTL also demonstrated stronger colocalization with HbA1c levels pre-intervention (H4 = 0.87) than post-intervention (H4 = 0.49; Figure S17), suggesting that lifestyle modifications might modulate the genetic influence of *FN3KRP* on glycemic control. This SNP rs2256833 has been linked to HbA1c across multiple ethnic groups and replicated in several GWASs.[79,83] *FN3KRP* encodes the fructosamine-3-kinase-related protein, which plays a crucial role in protein deglycation and protection against protein damage caused by excess glucose.[84] This protein deglycation system is particularly relevant to T2D as it helps prevent the accumulation of glycated proteins, which is a hallmark of chronic hyperglycemia.[85] Conversely, we identified a post-biased eQTL for *CRTC3* (with lead SNP rs8026714) that showed stronger colocalization with T2D. *CRTC3* encodes the cAMP-responsive element binding protein (CREB)-regulated transcription coactivator 3, which functions as a critical regulator of energy homeostasis and glucose metabolism through its role in mediating cAMP-dependent gene transcription.[86] The involvement of *CRTC3* in energy metabolism and glucose homeostasis makes it particularly relevant to T2D pathogenesis through its role in insulin resistance and metabolic dysfunction. Another post-biased eQTL was identified for *DOK7* (with lead SNP rs1203111) showing strong colocalization with both TG (H4 = 0.84) and TC (H4 = 0.80) levels post-intervention. Although *DOK7* is primarily known for its role in neuromuscular junction formation, it has also been previously associated with blood lipid levels.[87] Our findings suggest a potential link between *DOK7* expression and lipid metabolism in response to exercise intervention.

Similarly, we observed several sQTL-GWAS colocalization events. The sQTL of *ANK1* showed strong colocalization (H4 = 0.83) with East Asian T2D GWAS pre-intervention, consistent with previous findings in European populations.[54] This colocalization significantly reduced post-intervention (H4 = 0.48), suggesting that lifestyle intervention may modulate *cis*-regulatory elements of *ANK1* sQTL and impact its influence on T2D risk (Figure 5D). In contrast, a post-biased sQTL for *YBX3* (rs2886722) demonstrated stronger colocalization with CAD post-intervention (H4 = 0.98). *YBX3* has been previously implicated in cardiovascular development. The stronger colocalization following intervention may modulate *YBX3* genetic regulation on splicing in skeletal muscle. As *YBX3* is highly expressed in both heart and skeletal muscle tissues[76] and has been previously implicated in cardiovascular development,[88] this finding highlights a potential cross-tissue mechanism through weight loss-induced changes (Figure S18). Additionally, we observed a shared sQTL for *ZBTB20* (rs6806156) that maintained significant colocalization with T2D GWAS both pre-intervention (H4 = 0.87) and post-intervention (H4 = 0.89) (Figure S18), indicating a stable genetic effect that persists regardless of lifestyle changes. *ZBTB20* encodes the zinc finger and BTB domain-containing protein 20, a transcriptional repressor that plays important roles in glucose homeostasis and metabolic regulation.[76] Its persistent colocalization suggests that certain genetic effects on T2D risk may be independent of environmental modifications. *ZBTB20* has been shown

(C) This histogram displays the MAF of rs12437434 in the SAMS2 dataset and various populations from the 1000 Genomes Project. The *x* axis represents populations from different datasets, and the *y* axis represents the MAF. The red dashed line marks the 0.05 threshold, distinguishing between common and rare SNPs.

(D) The panel illustrates the colocalization of an sQTL for *ANK1*, with both the SNP and splicing intron located in the *ANK1* gene. The *x* axis represents the genomic position, and the *y* axis represents $-\log_{10}(P)$. The LD between SNPs and the target SNP is shown by the color gradient, indicating five levels of correlation.

Please cite this article in press as: Wang et al., Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.100951

CellPress
OPEN ACCESS

Table 2. List of colocalizations between SAMS2 e/sQTLs and metabolic trait GWASs

e/sQTL	Gene	GWAS phenotype	GWAS lead SNP	GWAS lead SNP position	Risk allele	AF	Type	Coloc PP4
eQTL	*RAB29*	body weight	rs823118	chr1:205754444:C:T	T	0.51	both	0.76
eQTL	*DHRS4L2*	type 2 diabetes	rs12437434	chr14:24409164:C:T	C	0.71	pre-biased	0.92
eQTL	*STIMATE*	body mass index	rs1108842	chr3:52686064:A:C	C	0.45	pre-biased	0.75
eQTL	*FN3KRP*	HbA1c	rs2256833	chr17:82731791:C:T	T	0.49	pre-biased	0.87
eQTL	*CRTC3*	type 2 diabetes	rs8026714	chr15:90979023:G:A	A	0.51	post-biased	0.76
eQTL	*ECT2L*	type 2 diabetes	rs9376382	chr6:138884249:T:C	C	0.66	post-biased	0.81
eQTL	*PNMA8A*	type 2 diabetes	rs11083824	chr19:46406873:C:G	G	0.23	post-biased	0.75
eQTL	*MFSD13A*	coronary artery disease body mass index body weight	rs1926032	chr10:103069712:C:T	T	0.28	post-biased	 0.77 0.77 0.76
eQTL	*DOK7*	total cholesterol	rs59950280	chr4:3450618:G:A	A	0.39	post-biased	0.8
eQTL	*DOK7*	triglycerides	rs4690098	chr4:3445429:C:T	T	0.48	post-biased	0.84
sQTL	*ZBTB20*	type 2 diabetes	rs6806156	chr3:115249171:C:T	T	0.57	both	0.89
sQTL	*RNF181*	coronary artery disease	rs2886722	chr2:85515174:A:G	G	0.37	both	0.9
sQTL	*UQCC1*	body weight	rs6142378	chr20:35409488:T:A	A	0.75	pre-biased	0.87
sQTL	*ANK1*	type 2 diabetes	rs6989203	chr8:41666227:G:A	A	0.15	pre-biased	0.83
sQTL	*YBX3*	coronary artery disease	rs138046359	chr12:10718301:A:C	C	0.49	pre-biased	0.99
sQTL	*KLHL24*	body weight	rs6770998	chr3:183627183:C:A	A	0.49	pre-biased	0.85
sQTL	*MROH8*	body weight	rs58664618	chr20:36901572:C:T	T	0.48	post-biased	0.48

This table summarizes loci showing evidence of colocalization between QTL signals identified in the SAMS2 skeletal muscle dataset and Asian GWAS signals for metabolic traits based on the COLOC method. A posterior probability for a shared causal variant (PP4) $\geq$ 0.75 was considered indicative of significant colocalization. Each column represents the following: e/sQTL (the type of identified QTL in the SAMS2 skeletal muscle database), gene (the target gene whose expression or splicing is associated with the QTL variant), GWAS phenotype (the metabolic traits linked to this intervention study); GWAS lead SNP (the lead SNP from the GWAS locus showing the strongest associated with the phenotype); GWAS lead SNP position (genomic coordinate hg38 of the lead GWAS SNP); risk allele (the allele associated with increased risk or higher trait value in the GWAS); AF (the allele frequency of the risk allele); type (indicates whether the QTL is shared or condition-biased); coloc PP4 (posterior probability PP4 from COLOC representing the probability that both the QTL and GWAS signals share a single causal variant).

to regulate hepatic glucose homeostasis and insulin sensitivity through direct transcriptional control of metabolic genes,[89] and its dysfunction has been linked to metabolic disorders in multiple studies. These findings illustrate the complex interplay between genetic and lifestyle factors in determining metabolic disease risks, underscoring the importance of considering both pre- and post-intervention to fully understand the genetic regulation of metabolic health.

DISCUSSION

Our study comprehensively characterizes the genetic regulatory mechanisms governing expression and splicing in response to lifestyle intervention combining diet and exercise in the skeletal muscle of an Asian male cohort. Unlike lifestyle intervention studies in captive rats and baboons,[18,21] our study needed to overcome low compliance and high attrition inherent in human interventional studies, so we constructed an extensive clinical phenotype and paired skeletal muscle transcriptomics resource specifically for an East Asian population undergoing weight-loss intervention. After intervention, body weight and fat percentage significantly decreased, while lean mass percentage, blood glucose control, and insulin sensitivity improved. Despite some weight regain, body weight and BMI remained significantly better than the baseline after 12 months, highlighting the lasting effects of the intervention. Paired skeletal muscle transcriptomes before and after intervention revealed changes in gene expression and splicing, highlighting improved insulin signaling and metabolic function. By mapping e/sQTLs in pre- and post-intervention states, we identified genetic variants shaping muscle gene regulation, including dozens of Asian-specific e/sQTLs absent in GTEx. Importantly, unlike most QTL studies that focus on steady-state conditions, we captured hundreds of gene-by-lifestyle interaction QTLs to reveal both shared and condition-biased regulatory mechanisms, demonstrating how genetic effects on muscle transcriptomic regulation are dynamically modulated by weight loss. Finally, colocalization with 12 metabolism-associated Asian GWASs pinpointed 16 putative risk genes.

Our findings provide a foundation for future mechanistic investigations by identifying key genetic regulatory changes in skeletal muscle following the lifestyle intervention. Notably, we observed significant splicing alterations in *FNBP4* and *TNNT3*, highlighting a rarely explored aspect of weight loss biology. Additionally, the association between the *ACTN3* R577X mutation and greater weight loss suggests potential

Please cite this article in press as: Wang et al., Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.100951

muscle fiber type-driven adaptations to aerobic exercise, warranting further functional validation to establish causality.[90] Through joint analysis with GWAS, we identified 16 putative metabolic risk genes, 13 of which exhibited differential regulatory effects across pre- and post-weight loss states. Genes with consistent regulatory effects across both conditions, such as *RAB29*, *ZBTB20*, and *RNF181*, likely represent core metabolic regulators with stable effects.[85,89,91] In contrast, pre-biased genes (e.g., *DHRS4L2*, *ANK1*, *UQCC1*) may contribute to obesity-associated metabolic dysfunction such as insulin resistance and impaired lipid handling that diminishes after weight loss.[54,92,93] Among them, *ANK1* is a well-established T2D risk locus identified across multiple GWASs, and prior studies have also reported skeletal muscle-specific regulatory mechanisms involving this gene.[4,55,73,75,79] However, our study demonstrated that the genetic regulation of *ANK1* splicing is significantly attenuated following lifestyle intervention. This suggests that weight loss can dynamically modulate genetic effects at T2D-associated loci, providing a mechanistic link between lifestyle changes and the mitigation of inherited metabolic risk. Post-biased genes (e.g., *FN3KRP*, *CRTC3*, *MFSD13A*) appear to gain regulatory significance in the improved metabolic state, potentially reflecting enhanced glycemic control and insulin sensitivity in response to the intervention.[79,86,94] These findings underscore the dynamic nature of gene-lifestyle interactions in metabolic regulation and emphasize the importance of considering ancestry-specific regulatory variants, particularly those enriched in Asian populations such as *DHRS4L2*. By capturing how genetic effects are modulated by lifestyle intervention, our study provides new insights for understanding the molecular mechanisms underlying obesity and T2D and may inform the development of precision strategies for metabolic disease prevention and treatment.

To better understand the long-term effects of lifestyle interventions, future studies should adopt a more structured and monitored approach during the weight maintenance phase, as the absence of exercise and dietary tracking may have influenced long-term outcomes. Incorporating single-cell RNA-seq will provide deeper insights into cell-type-specific responses,[95] particularly immune-muscle interactions that are masked in our bulk RNA-seq analysis. Expanding the study to include female participants is crucial, as metabolic responses to lifestyle interventions differ significantly between sexes due to genetic and hormonal influences.[96] This study serves as a proof of concept, and future research will be extended to women to explore sex-specific metabolic adaptations. Given that Asian individuals develop metabolic diseases at a lower BMI threshold compared with European individuals and remain underrepresented populations in genetic studies,[27] extending this research to more diverse Asian and other underrepresented populations will enhance our understanding of ancestry-specific genetic effects on intervention outcomes. Together, these research directions will enhance the scale and diversity of intervention studies, improve our characterization of complex tissue-specific disease-relevant effects, and facilitate better interpretation of GWAS signals across a wider array of complex traits, ultimately contributing to the development of more targeted interventions and treatments.

Limitations of the study

Although we comprehensively characterize the genetic regulatory mechanisms in response to lifestyle intervention, our results should be interpreted within several limitations. The complexity of maintaining a 16-week intervention study and participants' unwillingness to donate repeated muscle biopsies led to a relatively small final cohort size, which limits our ability to detect associations, particularly those with smaller effect sizes.[97] Additionally, our study cohort consisted exclusively of male participants, as is conventionally done in human exercise-based interventional studies.[98] While this design minimized confounding effects from sex-specific metabolic variations, including menstrual cycle-related metabolic fluctuations, it limits the generalizability of our findings to female populations. Dietary data in this study were based on self-reported food diaries, which are subject to recall bias and inaccuracies in portion estimation. These factors are commonly encountered in nutrition research and may affect the accuracy of caloric intake and nutrition research.[99] Efforts were made to minimize these biases by providing detailed instructions on how to fill up the food diary accurately. Moreover, we did not include lean individuals as controls, primarily because the study was initially designed as a longitudinal intervention study to assess within-individual regulatory changes before and after weight loss. Additionally, obtaining skeletal muscle biopsies from lean and healthy individuals is generally not compliant with institutional review board (IRB) regulations. Although our colocalization analyses identified regulatory variants associated with metabolic traits, we acknowledge that colocalization does not establish causality, nor does it imply that these genes are the primary drivers of metabolic traits. Finally, we acknowledge that functional validation needs to be performed in future studies to confirm these results from our current study.

RESOURCE AVAILABILITY

Lead contact

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Boxiang Liu (boxiangliu@nus.edu.sg).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The SAMS2 study was registered at ClinicalTrials.gov under the identifier NCT01080378. All the bulk RNA-sequencing raw data and the processed file generated from this study have been submitted to the NCBI under the accession number GSE282733. The skeletal muscle e/sQTL data, both pre-intervention and post-intervention, are deposited in Zenodo (doi: https://zenodo.org/records/14040483). The nucleotide sequence of the GRCh38 primary genome assembly and gene annotation are available at https://www.gencodegenes.org/human/release_43.html. The GTEx e/sQTL summary statistics can be downloaded from the GTEx portal (https://gtexportal.org/home/downloads/adult-gtex/qtl). The sources of the GWAS summary statistics used for the colocalization analysis are outlined in Table S14. The source code for the muscle-QTL pipeline is available at GitHub (https://github.com/boxiangliulab/muscle-QTL) and archived on Zenodo for reproducibility: https://doi.org/10.5281/zenodo.15683589. Genotype and clinical phenotypes are accessible through dbGaP under the accession number phs004078.

ACKNOWLEDGMENTS

B.L. is supported by the Ministry of Education, Singapore, under its Academic Research Fund Tier 2 (MOE-T2EP30123-0015) and Academic Research Fund Tier 1 (FY2023; 23-0434-A0001; 22-5800-A0001), and by A*STAR under the Nucleic Acid Therapeutics Initiative (award no.: H24J5a0066). This research is partially supported by the Precision Medicine Translational Research Program Core Funding under NUHSRO/2020/080/MSC/04/PM. The computational work for this article was partially performed on resources of the National Supercomputing Center, Singapore (https://www.nscc.sg) and partially supported by NUS IT's Cloud Credits for Research Program.

AUTHOR CONTRIBUTIONS

W.W., M.H.L., and B.L. conceptualized this study. S.Q.H. and E.C. processed muscle biopsy samples for RNA-seq. X.S. and S.H. carried out the genotyping and RNA-seq experiments. W.W. processed genotyping data. C.T. assisted W.W. in RNA-seq data processing and biological interpretation. A.L.M.T., S.R.L., and P.E. analyzed initial RNA-seq data. W.W. and Y.T. performed QTL analyses. W.W. and Y.Z. performed colocalization. W.W. and F.L. conducted fine-mapping analyses. S.A.S., S.S.V., and K.V. were involved in the data collection of the study. E.Y.H.K., E.S.T., C.M.K., M.K.-S.L., Y.S.L., and K.V. ran the study and collected the clinical data and biopsy samples from all the study participants. E.S.T., E.Y.H.K., M.K.-S.L., and Y.S.C. designed and secured funding for the study. K.W.T. oversaw the design and implementation of the weight loss intervention. W.W. and S.J.L.O. consolidated and analyzed clinical data. W.L.L., M.H.L., and Y.Q. performed validation studies in myotubes. E.Y.H.K., E.S.T., C.M.K., and M.H.L. maintained study data. W.W. performed main data analyses, interpreted the results, and wrote the manuscript. B.L., W.L.L., M.H.L., and E.S.T. reviewed and provided input in the writing of the final manuscript. B.L. and M.H.L. supervised and guided W.W.

DECLARATION OF INTERESTS

The authors declare no competing interests.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - Study design for assessment of weight loss outcomes
  - Weight maintenance phase
  - Anthropometrics
  - Dual-energy X-Ray absorptiometry (DXA)
  - Magnetic Resonance Imaging
  - MR spectroscopy of liver and skeletal muscle
  - Peripheral blood biochemical measurements
  - Accelerometer
  - Doubly labeled water technique for measurement of total energy expenditure (TEE)
  - Indirect calorimetry – Mixed meal tolerance test
  - Differential clinical traits
  - Lipidomes profiling analysis
  - Genotyping and imputation
  - Skeletal muscle RNA sequencing and quality control
  - Skeletal muscle transcriptome profiling and network analysis
  - Quantification of mRNA alternative splicing
  - Differential alternative splicing analysis
  - Expression quantitative trait loci (eQTLs) with RASQUAL
  - Splicing quantitative trait loci (sQTLs) with QTLtools
  - Independent sQTL mapping
  - e/sQTL replication
  - Identification of gene-by-lifestyle interactions e/sQTLs
  - Asian-specific e/sQTLs
  - GWAS summary statistics
  - Colocalization between GWAS and e/sQTLs
  - Fine-mapping

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.xgen.2025.100951.

Received: November 11, 2024
Revised: April 9, 2025
Accepted: June 23, 2025

REFERENCES

1. Florez, J.C., Udler, M.S., and Hanson, R.L. (2018). Genetics of Type 2 Diabetes. In Diabetes in America, 3rd Edition (National Institute of Diabetes and Digestive and Kidney Diseases (US)).
2. Kolb, H., and Martin, S. (2017). Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. BMC Med. *15*, 131.
3. Ishigaki, K., Akiyama, M., Kanai, M., Takahashi, A., Kawakami, E., Sugishita, H., Sakaue, S., Matoba, N., Low, S.K., Okada, Y., et al. (2020). Large-scale genome-wide association study in a Japanese population identifies novel susceptibility loci across different diseases. Nat. Genet. *52*, 669–679.
4. Spracklen, C.N., Horikoshi, M., Kim, Y.J., Lin, K., Bragg, F., Moon, S., Suzuki, K., Tam, C.H.T., Tabara, Y., Kwak, S.H., et al. (2020). Identification of type 2 diabetes loci in 433,540 East Asian individuals. Nature *582*, 240–245.
5. Xue, A., Wu, Y., Zhu, Z., Zhang, F., Kemper, K.E., Zheng, Z., Yengo, L., Lloyd-Jones, L.R., Sidorenko, J., Wu, Y., et al. (2018). Genome-wide association analyses identify 143 risk variants and putative regulatory mechanisms for type 2 diabetes. Nat. Commun. *9*, 2941.
6. Mahajan, A., Spracklen, C.N., Zhang, W., Ng, M.C.Y., Petty, L.E., Kitajima, H., Yu, G.Z., Rüeger, S., Speidel, L., Kim, Y.J., et al. (2022). Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. Nat. Genet. *54*, 560–572.
7. Uusitupa, M., Khan, T.A., Viguiliouk, E., Kahleova, H., Rivellese, A.A., Hermansen, K., Pfeiffer, A., Thanopoulou, A., Salas-Salvadó, J., Schwab, U., and Sievenpiper, J.L. (2019). Prevention of Type 2 Diabetes by Lifestyle Changes: A Systematic Review and Meta-Analysis. Nutrients *11*, 2611.
8. Diabetes Prevention Program DPP Research Group (2002). The Diabetes Prevention Program (DPP): description of lifestyle intervention. Diabetes Care *25*, 2165–2171.
9. Diabetes Prevention Program Research Group; Knowler, W.C., Fowler, S.E., Hamman, R.F., Christophi, C.A., Hoffman, H.J., Brenneman, A.T., Brown-Friday, J.O., Goldberg, R., Venditti, E., and Nathan, D.M. (2009). 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. Lancet *374*, 1677–1686.
10. Legaard, G.E., Lyngbæk, M.P.P., Almdal, T.P., Karstoft, K., Bennetsen, S.L., Feineis, C.S., Nielsen, N.S., Durrer, C.G., Liebetrau, B., Nystrup, U., et al. (2023). Effects of different doses of exercise and diet-induced weight loss on beta-cell function in type 2 diabetes (DOSE-EX): a randomized clinical trial. Nat. Metab. *5*, 880–895.
11. Yoshino, M., Yoshino, J., Smith, G.I., Stein, R.I., Bittel, A.J., Bittel, D.C., Reeds, D.N., Sinacore, D.R., Cade, W.T., Patterson, B.W., et al. (2022). Worksite-based intensive lifestyle therapy has profound cardiometabolic benefits in people with obesity and type 2 diabetes. Cell Metab. *34*, 1431–1441.e5.
12. Papandonatos, G.D., Pan, Q., Pajewski, N.M., Delahanty, L.M., Peter, I., Erar, B., Ahmad, S., Harden, M., Chen, L., Fontanillas, P., et al. (2015). Genetic Predisposition to Weight Loss and Regain With Lifestyle

Intervention: Analyses From the Diabetes Prevention Program and the Look AHEAD Randomized Controlled Trials. Diabetes *64*, 4312–4321.

13. Herrera-Luis, E., Benke, K., Volk, H., Ladd-Acosta, C., and Wojcik, G.L. (2024). Gene–environment interactions in human health. Nat. Rev. Genet. *25*, 768–784.
14. Boye, C., Nirmalan, S., Ranjbaran, A., and Luca, F. (2024). Genotype × environment interactions in gene regulation and complex traits. Nat. Genet. *56*, 1057–1068.
15. Motsinger-Reif, A.A., Reif, D.M., Akhtari, F.S., House, J.S., Campbell, C. R., Messier, K.P., Fargo, D.C., Bowen, T.A., Nadadur, S.S., Schmitt, C.P., et al. (2024). Gene-environment interactions within a precision environmental health framework. Cell Genom. *4*, 100591.
16. Langenberg, C., Sharp, S.J., Franks, P.W., Scott, R.A., Deloukas, P., Forouhi, N.G., Froguel, P., Groop, L.C., Hansen, T., Palla, L., et al. (2014). Gene-Lifestyle Interaction and Type 2 Diabetes: The EPIC InterAct Case-Cohort Study. PLoS Med. *11*, e1001647.
17. Ling, C., and Rönn, T. (2019). Epigenetics in Human Obesity and Type 2 Diabetes. Cell Metab. *29*, 1028–1044.
18. Nair, V.D., Pincas, H., Smith, G.R., Zaslavsky, E., Ge, Y., Amper, M.A.S., Vasoya, M., Chikina, M., Sun, Y., Raja, A.N., et al. (2024). Molecular adaptations in response to exercise training are associated with tissue-specific transcriptomic and epigenomic signatures. Cell Genom. *4*, 100421.
19. Vetr, N.G., Gay, N.R., and MoTrPAC Study Group; and Montgomery, S.B. (2024). The impact of exercise on gene regulation in association with complex trait genetics. Nat. Commun. *15*, 3346.
20. MoTrPAC Study Group , Lead Analysts, MoTrPAC Study Group (2024). Temporal dynamics of the multi-omic response to endurance exercise training. Nature *629*, 174–183.
21. Lin, W., Wall, J.D., Li, G., Newman, D., Yang, Y., Abney, M., VandeBerg, J.L., Olivier, M., Gilad, Y., and Cox, L.A. (2024). Genetic regulatory effects in response to a high-cholesterol, high-fat diet in baboons. Cell Genom. *4*, 100509.
22. Galvani, A.P., Bauch, C.T., Anand, M., Singer, B.H., and Levin, S.A. (2016). Human–environment interactions in population and ecosystem health. Proc. Natl. Acad. Sci. USA *113*, 14502–14506.
23. Ruelas, A.L., Martínez Contreras, T.d.J., Esparza Romero, J., Díaz Zavala, R.G., Candia Plata, M.D.C., Hingle, M., Armenta Guirado, B., and Haby, M.M. (2023). Factors influencing adults to drop out of intensive lifestyle interventions for weight loss. Transl. Behav. Med. *13*, 245–254.
24. Middleton, K.R., Anton, S.D., and Perri, M.G. (2013). Long-Term Adherence to Health Behavior Change. Am. J. Lifestyle Med. *7*, 395–404.
25. DeFronzo, R.A., and Tripathy, D. (2009). Skeletal Muscle Insulin Resistance Is the Primary Defect in Type 2 Diabetes. Diabetes Care *32*, S157–S163.
26. Espino-Gonzalez, E., Dalbram, E., Mounier, R., Gondin, J., Farup, J., Jessen, N., and Treebak, J.T. (2024). Impaired skeletal muscle regeneration in diabetes: From cellular and molecular mechanisms to novel treatments. Cell Metab. *36*, 1204–1236.
27. (2004). Appropriate body-mass index for Asian populations and its implications for policy and intervention strategies. Lancet *363*, 157–163.
28. Martínez-Reyes, I., and Chandel, N.S. (2020). Mitochondrial TCA cycle metabolites control physiology and disease. Nat. Commun. *11*, 102.
29. Gibala, M.J., Young, M.E., and Taegtmeyer, H. (2000). Anaplerosis of the citric acid cycle: role in energy metabolism of heart and skeletal muscle. Acta Physiol. Scand. *168*, 657–665.
30. Savage, D.B., Petersen, K.F., and Shulman, G.I. (2007). Disordered lipid metabolism and the pathogenesis of insulin resistance. Physiol. Rev. *87*, 507–520.
31. Conroy, M.J., Andrews, R.M., Andrews, S., Cockayne, L., Dennis, E.A., Fahy, E., Gaud, C., Griffiths, W.J., Jukes, G., Kolchin, M., et al. (2024). LIPID MAPS: update to databases and tools for the lipidomics community. Nucleic Acids Res. *52*, D1677–D1682.
32. Schooneman, M.G., Vaz, F.M., Houten, S.M., and Soeters, M.R. (2013). Acylcarnitines: reflecting or inflicting insulin resistance? Diabetes *62*, 1–8.
33. Meikle, P.J., and Summers, S.A. (2017). Sphingolipids and phospholipids in insulin resistance and related metabolic disorders. Nat. Rev. Endocrinol. *13*, 79–91.
34. Chaurasia, B., and Summers, S.A. (2015). Ceramides – Lipotoxic Inducers of Metabolic Disorders. Trends Endocrinol. Metab. *26*, 538–550.
35. Tarling, E.J., Vallim, T.Q.D.A., and Edwards, P.A. (2013). Role of ABC transporters in lipid transport and human disease. Trends Endocrinol. Metab. *24*, 342–350.
36. Farooqi, I.S., and O'Rahilly, S. (2024). Leptin: a pivotal regulator of human energy homeostasis. Am. J. Clin. Nutr. *89*, 980S–984S.
37. Zhao, S., Li, N., Xiong, W., Li, G., He, S., Zhang, Z., Zhu, Q., Jiang, N., Ikejiofor, C., Zhu, Y., et al. (2024). Leptin Reduction as a Required Component for Weight Loss. Diabetes *73*, 197–210.
38. Merz, K.E., and Thurmond, D.C. (2020). Role of Skeletal Muscle in Insulin Resistance and Glucose Uptake. Compr. Physiol. *10*, 785–809.
39. O'Neill, H.M. (2013). AMPK and Exercise: Glucose Uptake and Insulin Sensitivity. Diabetes Metab. J. *37*, 1–21.
40. Watson, R.T., and Pessin, J.E. (2006). Bridging the GAP between insulin signaling and GLUT4 translocation. Trends Biochem. Sci. *31*, 215–222.
41. Mackenzie, R.W., and Elliott, B.T. (2014). Akt/PKB activation and insulin signaling: a novel insulin signaling pathway in the treatment of type 2 diabetes. Diabetes Metab. Syndr. Obes. *7*, 55–64.
42. Dos Santos, J.M., Moreli, M.L., Tewari, S., and Benite-Ribeiro, S.A. (2015). The effect of exercise on skeletal muscle glucose uptake in type 2 diabetes: An epigenetic perspective. Metabolism *64*, 1619–1628.
43. Rajpathak, S.N., Gunter, M.J., Wylie-Rosett, J., Ho, G.Y.F., Kaplan, R.C., Muzumdar, R., Rohan, T.E., and Strickler, H.D. (2009). The role of insulin-like growth factor-I and its binding proteins in glucose homeostasis and type 2 diabetes. Diabetes. Metab. Res. Rev. *25*, 3–12.
44. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinf. *9*, 559.
45. Nolfi-Donegan, D., Braganza, A., and Shiva, S. (2020). Mitochondrial electron transport chain: Oxidative phosphorylation, oxidant production, and methods of measurement. Redox Biol. *37*, 101674.
46. Xiao, W., Wang, R.-S., Handy, D.E., and Loscalzo, J. (2018). NAD(H) and NADP(H) Redox Couples and Cellular Energy Metabolism. Antioxid. Redox Signal. *28*, 251–272.
47. Pandorf, C.E., Haddad, F., Owerkowicz, T., Carroll, L.P., Baldwin, K.M., and Adams, G.R. (2020). Regulation of myosin heavy chain antisense long noncoding RNA in human vastus lateralis in response to exercise training. Am. J. Physiol. Cell Physiol. *318*, C931–C942.
48. Zheng, K.-W., Zhang, C.H., Wu, W., Zhu, Z., Gong, J.P., and Li, C.M. (2023). FNBP4 is a Potential Biomarker Associated with Cuproptosis and Promotes Tumor Progression in Hepatocellular Carcinoma. Int. J. Gen. Med. *16*, 467–480.
49. Stefancsik, R., Randall, J.D., Mao, C., and Sarkar, S. (2003). Structure and sequence of the human fast skeletal troponin T (TNNT3) gene: insight into the evolution of the gene and the origin of the developmentally regulated isoforms. Comp. Funct. Genom. *4*, 609–625.
50. Das, S., Forer, L., Schönherr, S., Sidore, C., Locke, A.E., Kwong, A., Vrieze, S.I., Chew, E.Y., Levy, S., McGue, M., et al. (2016). Next-generation genotype imputation service and methods. Nat. Genet. *48*, 1284–1287.
51. Kumasaka, N., Knights, A.J., and Gaffney, D.J. (2016). Fine-mapping cellular QTLs with RASQUAL and ATAC-seq. Nat. Genet. *48*, 206–213.
52. Peterson, C.B., Bogomolov, M., Benjamini, Y., and Sabatti, C. (2016). TreeQTL: hierarchical error control for eQTL findings. Bioinformatics *32*, 2556–2558.

53. Keildson, S., Fadista, J., Ladenvall, C., Hedman, Å.K., Elgzyri, T., Small, K.S., Grundberg, E., Nica, A.C., Glass, D., Richards, J.B., et al. (2014). Expression of Phosphofructokinase in Skeletal Muscle Is Influenced by Genetic Variation and Associated With Insulin Sensitivity. Diabetes *63*, 1154–1165.

54. Scott, L.J., Erdos, M.R., Huyghe, J.R., Welch, R.P., Beck, A.T., Wolford, B.N., Chines, P.S., Didion, J.P., Narisu, N., Stringham, H.M., et al. (2016). The genetic regulatory signature of type 2 diabetes in human skeletal muscle. Nat. Commun. *7*, 11764.

55. Scott, R.A., Lagou, V., Welch, R.P., Wheeler, E., Montasser, M.E., Luan, J., Mägi, R., Strawbridge, R.J., Rehnberg, E., Gustafsson, S., et al. (2012). Large-scale association analyses identify new loci influencing glycemic traits and provide insight into the underlying biological pathways. Nat. Genet. *44*, 991–1005.

56. Sajuthi, S.P., Sharma, N.K., Chou, J.W., Palmer, N.D., McWilliams, D.R., Beal, J., Comeau, M.E., Ma, L., Calles-Escandon, J., Demons, J., et al. (2016). Mapping adipose and muscle tissue expression quantitative trait loci in African Americans to identify genes for type 2 diabetes and obesity. Hum. Genet. *135*, 869–880.

57. Sherry, S.T., Ward, M.H., Kholodov, M., Baker, J., Phan, L., Smigielski, E. M., and Sirotkin, K. (2001). dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. *29*, 308–311.

58. Vincent, B., De Bock, K., Ramaekers, M., Van den Eede, E., Van Leemputte, M., Hespel, P., and Thomis, M.A. (2007). ACTN3 (R577X) genotype is associated with fiber type distribution. Physiol. Genomics *32*, 58–63.

59. North, K.N., Yang, N., Wattanasirichaigoon, D., Mills, M., Easteal, S., and Beggs, A.H. (1999). A common nonsense mutation results in α-actinin-3 deficiency in the general population. Nat. Genet. *21*, 353–354.

60. Seto, J.T., Roeszler, K.N., Meehan, L.R., Wood, H.D., Tiong, C., Bek, L., Lee, S.F., Shah, M., Quinlan, K.G.R., Gregorevic, P., et al. (2021). ACTN3 genotype influences skeletal muscle mass regulation and response to dexamethasone. Sci. Adv. *7*, eabg0088.

61. Kikuchi, N., Yoshida, S., Min, S.k., Lee, K., Sakamaki-Sunaga, M., Okamoto, T., and Nakazato, K. (2015). The ACTN3 R577X genotype is associated with muscle function in a Japanese population. Appl. Physiol. Nutr. Metab. *40*, 316–322.

62. Chan, S., Seto, J.T., MacArthur, D.G., Yang, N., North, K.N., and Head, S. I. (2008). A gene for speed: contractile properties of isolated whole EDL muscle from an α-actinin-3 knockout mouse. Am. J. Physiol. Cell Physiol. *295*, C897–C904.

63. Yuen, M., and Ottenheijm, C.A.C. (2020). Nebulin: big protein with big responsibilities. J. Muscle Res. Cell Motil. *41*, 103–124.

64. Labeit, S., Ottenheijm, C.A.C., and Granzier, H. (2011). Nebulin, a major player in muscle health and disease. FASEB j. *25*, 822–829.

65. Luo, G., Herrera, A.H., and Horowits, R. (1999). Molecular Interactions of N-RAP, a Nebulin-Related Protein of Striated Muscle Myotendon Junctions and Intercalated Disks. Biochemistry *38*, 6135–6143.

66. Jin, S., Kim, J., Willert, T., Klein-Rodewald, T., Garcia-Dominguez, M., Mosqueira, M., Fink, R., Esposito, I., Hofbauer, L.C., Charnay, P., and Kieslinger, M. (2014). Ebf factors and MyoD cooperate to regulate muscle relaxation via Atp2a1. Nat. Commun. *5*, 3793.

67. GTEx Consortium (2017). The impact of rare variation on gene expression across tissues. Nature *550*, 239–243.

68. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330.

69. 1000 Genomes Project Consortium; Auton, A., Brooks, L.D., Durbin, R. M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). A global reference for human genetic variation. Nature *526*, 68–74.

70. Delaneau, O., Ongen, H., Brown, A.A., Fort, A., Panousis, N.I., and Dermitzakis, E.T. (2017). A complete tool set for molecular QTL discovery and analysis. Nat. Commun. *8*, 15452.

71. Hood, D.A., Memme, J.M., Oliveira, A.N., and Triolo, M. (2019). Maintenance of Skeletal Muscle Mitochondria in Health, Exercise, and Aging. Annu. Rev. Physiol. *81*, 19–41.

72. Oddoux, S., Brocard, J., Schweitzer, A., Szentesi, P., Giannesini, B., Brocard, J., Fauré, J., Pernet-Gallay, K., Bendahan, D., Lunardi, J., et al. (2009). Triadin Deletion Induces Impaired Skeletal Muscle Function. J. Biol. Chem. *284*, 34918–34929.

73. Gallagher, P.G., and Forget, B.G. (1998). An Alternate Promoter Directs Expression of a Truncated, Muscle-specific Isoform of the Human Ankyrin 1 Gene. J. Biol. Chem. *273*, 1339–1348.

74. Bagnato, P., Barone, V., Giacomello, E., Rossi, D., and Sorrentino, V. (2003). Binding of an ankyrin-1 isoform to obscurin suggests a molecular link between the sarcoplasmic reticulum and myofibrils in striated muscles. J. Cell Biol. *160*, 245–253.

75. Yan, R., Lai, S., Yang, Y., Shi, H., Cai, Z., Sorrentino, V., Du, H., and Chen, H. (2016). A novel type 2 diabetes risk allele increases the promoter activity of the muscle-specific small ankyrin 1 gene. Sci. Rep. *6*, 25105.

76. Stelzer, G., Rosen, N., Plaschkes, I., Zimmerman, S., Twik, M., Fishilevich, S., Stein, T.I., Nudel, R., Lieder, I., Mazor, Y., et al. (2016). The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. Curr. Protoc. Bioinformatics *54*, 1.30.1–1.30.33.

77. Blue, R.E., Koushik, A., Engels, N.M., Wiedner, H.J., Cooper, T.A., and Giudice, J. (2018). Modulation of alternative splicing of trafficking genes by genome editing reveals functional consequences in muscle biology. Int. J. Biochem. Cell Biol. *105*, 134–143.

78. Gilbert, S.L., Zhang, L., Forster, M.L., Anderson, J.R., Iwase, T., Soliven, B., Donahue, L.R., Sweet, H.O., Bronson, R.T., Davisson, M.T., et al. (2006). Trak1 mutation disrupts GABAA receptor homeostasis in hypertonic mice. Nat. Genet. *38*, 245–250.

79. Sakaue, S., Kanai, M., Tanigawa, Y., Karjalainen, J., Kurki, M., Koshiba, S., Narita, A., Konuma, T., Yamamoto, K., Akiyama, M., et al. (2021). A cross-population atlas of genetic associations for 220 human phenotypes. Nat. Genet. *53*, 1415–1424.

80. Nam, K., Kim, J., and Lee, S. (2022). Genome-wide study on 72,298 individuals in Korean biobank data for 76 traits. Cell Genom. *2*, 100189.

81. Giambartolomei, C., Vukcevic, D., Schadt, E.E., Franke, L., Hingorani, A. D., Wallace, C., and Plagnol, V. (2014). Bayesian Test for Colocalisation between Pairs of Genetic Association Studies Using Summary Statistics. PLoS Genet. *10*, e1004383.

82. Wang, G., Sarkar, A., Carbonetto, P., and Stephens, M. (2020). A Simple New Approach to Variable Selection in Regression, with Application to Genetic Fine Mapping. J. R. Stat. Soc. Series B Stat. Methodol. *82*, 1273–1300.

83. Sakashita, T., Nakamura, Y., Sutoh, Y., Shimizu, A., Hachiya, T., Otsuka-Yamasaki, Y., Takashima, N., Kadota, A., Miura, K., Kita, Y., et al. (2023). Comparison of the loci associated with HbA1c and blood glucose levels identified by a genome-wide association study in the Japanese population. Diabetol. Int. *14*, 188–198.

84. Collard, F., Wiame, E., Bergans, N., Fortpied, J., Vertommen, D., Vanstapel, F., Delpierre, G., and Van Schaftingen, E. (2004). Fructosamine 3-kinase-related protein and deglycation in human erythrocytes. Biochem. J. *382*, 137–143.

85. Rabbani, N., and Thornalley, P.J. (2021). Protein glycation - biomarkers of metabolic dysfunction and early-stage decline in health in the era of precision medicine. Redox Biol. *42*, 101920.

86. Song, Y., Altarejos, J., Goodarzi, M.O., Inoue, H., Guo, X., Berdeaux, R., Kim, J.H., Goode, J., Igata, M., Paz, J.C., et al. (2010). CRTC3 links catecholamine signalling to energy balance. Nature *468*, 933–939.

87. Global Lipids Genetics Consortium (2013). Discovery and refinement of loci associated with lipid levels. Nat. Genet. *45*, 1274–1283.

88. Awad, S., Skipper, W., Vostrejs, W., Ozorowski, K., Min, K., Pfuhler, L., Mehta, D., and Cooke, A. (2024). The YBX3 RNA-binding protein

posttranscriptionally controls SLC1A5 mRNA in proliferating and differentiating skeletal muscle cells. J. Biol. Chem. *300*, 105602.

89. Liu, G., Zhou, L., Zhang, H., Chen, R., Zhang, Y., Li, L., Lu, J.Y., Jiang, H., Liu, D., Qi, S., et al. (2017). Regulation of hepatic lipogenesis by the zinc finger protein Zbtb20. Nat. Commun. *8*, 14824.
90. El Ouali, E.M., Barthelemy, B., Del Coso, J., Hackney, A.C., Laher, I., Govindasamy, K., Mesfioui, A., Granacher, U., and Zouhal, H. (2024). A Systematic Review and Meta-analysis of the Association Between ACTN3 R577X Genotypes and Performance in Endurance Versus Power Athletes and Non-athletes. Sports Med. Open *10*, 37.
91. Sánchez-Garrido, M.A., Serrano-López, V., Ruiz-Pino, F., Vázquez, M.J., Rodríguez-Martín, A., Torres, E., Velasco, I., Rodríguez, A.B., Chicano-Gálvez, E., Mora-Ortiz, M., et al. (2024). Superior metabolic improvement of polycystic ovary syndrome traits after GLP1-based multi-agonist therapy. Nat. Commun. *15*, 8498.
92. Su, Z.-J., Zhang, Q.X., Liu, G.F., Song, X.H., Li, Q., Wang, R.J., Chen, H. B., Xu, X.Y., Sui, X.X., and Huang, D.Y. (2010). Bioinformatic analysis of the human DHRS4 gene cluster and a proposed mechanism for its transcriptional regulation. BMC Mol. Biol. *11*, 43.
93. Liang, C., Zhang, S., Robinson, D., Ploeg, M.V., Wilson, R., Nah, J., Taylor, D., Beh, S., Lim, R., Sun, L., et al. (2022). Mitochondrial microproteins link metabolic cues to respiratory chain biogenesis. Cell Rep. *40*, 111204.
94. Sarnowski, C., Leong, A., Raffield, L.M., Wu, P., de Vries, P.S., DiCorpo, D., Guo, X., Xu, H., Liu, Y., Zheng, X., et al. (2019). Impact of Rare and Common Genetic Variants on Diabetes Diagnosis by Hemoglobin A1c in Multi-Ancestry Cohorts: The Trans-Omics for Precision Medicine Program. Am. J. Hum. Genet. *105*, 706–718.
95. De Micheli, A.J., Spector, J.A., Elemento, O., and Cosgrove, B.D. (2020). A reference single-cell transcriptomic atlas of human skeletal muscle tissue reveals bifurcated muscle stem cell populations. Skelet. Muscle *10*, 19.
96. Pataky, M.W., Dasari, S., Michie, K.L., Sevits, K.J., Kumar, A.A., Klaus, K. A., Heppelmann, C.J., Robinson, M.M., Carter, R.E., Lanza, I.R., and Nair, K.S. (2023). Impact of biological sex and sex hormones on molecular signatures of skeletal muscle at rest and in response to distinct exercise training modes. Cell Metab. *35*, 1996–2010.e6.
97. Cadmus-Bertram, L., Irwin, M., Alfano, C., Campbell, K., Duggan, C., Foster-Schubert, K., Wang, C.Y., and McTiernan, A. (2014). Predicting Adherence of Adults to a 12-Month Exercise Intervention. J. Phys. Act. Health *11*, 1304–1312.
98. Costello, J.T., Bieuzen, F., and Bleakley, C.M. (2014). Where are all the female participants in Sports and Exercise Medicine research? Eur. J. Sport Sci. *14*, 847–851.
99. Ottaviani, J.I., Sagi-Kiss, V., Schroeter, H., and Kuhnle, G.G.C. (2024). Reliance on self-reports and estimated food composition data in nutrition research introduces significant bias that can only be addressed with biomarkers. eLife *13*, RP92941.
100. Sadananthan, S.A., Prakash, B., Leow, M.K.S., Khoo, C.M., Chou, H., Venkataraman, K., Khoo, E.Y.H., Lee, Y.S., Gluckman, P.D., Tai, E.S., and Velan, S.S. (2015). Automated segmentation of visceral and subcutaneous (deep and superficial) adipose tissues in normal and overweight men. J. Magn. Reson. Imaging. *41*, 924–934.
101. Cowin, G.J., Jonsson, J.R., Bauer, J.D., Ash, S., Ali, A., Osland, E.J., Purdie, D.M., Clouston, A.D., Powell, E.E., and Galloway, G.J. (2008). Magnetic resonance imaging and spectroscopy for monitoring liver steatosis. J. Magn. Reson. Imaging. *28*, 937–945.
102. Buuren, S.V., and Groothuis-Oudshoorn, K. (2011). mice : Multivariate Imputation by Chained Equations in R. J. Stat. Soft. *45*, 1–67.
103. Haynes, W. (2013). Benjamini–hochberg method. In Encyclopedia of Systems Biology, W. Dubitzky, O. Wolkenhauer, K.-H. Cho, and H. Yokota, eds. (New York: NY: Springer New York), p. 78. https://doi.org/10.1007/978-1-4419-9863-7_1215.
104. Makowski, D., Ben-Shachar, M., Patil, I., and Lüdecke, D. (2020). Methods and Algorithms for Correlation Analysis in R. JOSS *5*, 2306.
105. Villanueva, R.A.M., and Chen, Z.J. (2019). ggplot2: Elegant Graphics for Data Analysis (2nd ed.). Measurement: Interdisciplinary Research and Perspectives *17*, 160–167.
106. Chong, J., and Xia, J. (2018). MetaboAnalystR: an R package for flexible and reproducible analysis of metabolomics data. Bioinformatics *34*, 4313–4314.
107. Weeks, J.P. (2010). plink : An R Package for Linking Mixed-Format Tests Using IRT-Based Methods. J. Stat. Soft. *35*, 1–33.
108. Zhao, S., Jing, W., Samuels, D.C., Sheng, Q., Shyr, Y., and Guo, Y. (2018). Strategies for processing and quality control of Illumina genotyping arrays. Brief. Bioinform. *19*, 765–775.
109. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.
110. Kaminow, B., Yunusov, D., and Dobin, A. (2021). STARsolo: Accurate, Fast and Versatile Mapping/Quantification of Single-Cell and Single-Nucleus RNA-Seq Data. bioRxiv. https://doi.org/10.1101/2021.05.05.442755.
111. Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: The reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.
112. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.
113. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.
114. Xie, C., Jauhari, S., and Mora, A. (2021). Popularity and performance of bioinformatics software: the case of gene set analysis. BMC Bioinf. *22*, 191.
115. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. *13*, 2498–2504.
116. Li, Y.I., Knowles, D.A., Humphrey, J., Barbeira, A.N., Dickinson, S.P., Im, H.K., and Pritchard, J.K. (2018). Annotation-free quantification of RNA splicing using LeafCutter. Nat. Genet. *50*, 151–158.
117. Garrido-Martín, D., Palumbo, E., Guigó, R., and Breschi, A. (2018). ggsashimi: Sashimi plot revised for browser- and annotation-independent splicing visualization. PLoS Comput. Biol. *14*, e1006360.
118. Stegle, O., Parts, L., Piipari, M., Winn, J., and Durbin, R. (2012). Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat. Protoc. *7*, 500–507.
119. Storey, J.D. (2002). A Direct Approach to False Discovery Rates. J. Roy. Stat. Soc. B Stat. Methodol. *64*, 479–498.
120. Liu, B., Gloudemans, M.J., Rao, A.S., Ingelsson, E., and Montgomery, S. B. (2019). Abundant associations with gene expression complicate GWAS follow-up. Nat. Genet. *51*, 768–769.
121. Machiela, M.J., and Chanock, S.J. (2015). LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics *31*, 3555–3557.

Please cite this article in press as: Wang et al., Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.100951

CellPress
OPEN ACCESS

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Human skeletal muscle samples	National University of Singapore; NUH Tissue Repository (NUH TR)	N/A
Deposited data		
Clinical Phenotypes	This paper	dbGaP: phs004078
Genotype	This paper	dbGaP: phs004078
RNA-seq data	This paper	GEO: GSE282733
Clinical data	This paper	https://zenodo.org/records/14934808
e/sQTLs summary statistics	This paper	https://zenodo.org/records/14040483
GTEx data	GTEx	https://gtexportal.org/home/datasets
GWAS Catalog summary statistics	NHGRI-EBI GWAS Catalog	https://www.ebi.ac.uk/gwas/
Codes and Scripts	This paper	https://zenodo.org/records/15683589
Software and algorithms		
PLINK v1.9	Chang et al.	https://www.cog-genomics.org/plink/
FASTP	Chen et al.	https://github.com/OpenGene/fastp
STAR	Dobin et al.	https://github.com/alexdobin/STAR
featureCounts	Liao et al.	https://subread.sourceforge.net/
SHAPEIT4	Delaneau et al.	https://odelaneau.github.io/shapeit4/
DESeq2	Love et al.	https://github.com/thelovelab/DESeq2
WGCNA	Langfelder et al.	https://github.com/cran/WGCNA
MetaboAnalystR	Pang et al.	https://github.com/xia-lab/MetaboAnalystR
QTLtools	Delaneau et al.	https://qtltools.github.io/qtltools/
RASQUAL	Kumasaka et al.	https://github.com/natsuhiko/rasqual
susieR	Wang et al.	https://stephenslab.github.io/susieR/articles/finemapping.html
COLOC	Wallace et al.	https://chr1swallace.github.io/coloc/index.html
Locuscompare	Liu et al.	https://github.com/boxiangliu/locuscomparer
qvalue	Storey et al.	https://www.bioconductor.org/packages/release/bioc/html/qvalue.html
Analysis code generated for this study	This paper	https://github.com/boxiangliulab/muscle-QTL

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

SAMS2 was an interventional study where hundreds of donors aged 21–45 were recruited to participate in a 16-week weight loss program. Study individuals selected (see below for more detailed selection standards) were sedentary (exercise 1 or fewer times a week), obese or overweight with a body fat mass greater than 24% and a BMI between 23 and 35 kg/m^2. For this study, we adjusted BMI definitions for the Asian population based on WHO Consultation 2002.[27] Therefore, the BMI cut-off for overweight is 23 kg/m^2 and 27.5 kg/m^2 for obese. The weight loss program included a combination of dietary interventions, structured exercise sessions, and additional physical activity performed in participants' own time. Energy and protein requirements were calculated based on each participant's weight, height, and physical activity level, with the goal of achieving a 40% calorie deficit. Participants' calorie intake was tracked using food recalls and questionnaires. Additionally, subjects attended structured exercise sessions at least three times per week, supervised by a coach. Each session consisted of 90 min of aerobic and strength training exercises, designed to burn approximately 500 kcal per session. To monitor daily physical activity, participants wore pedometers throughout the study. In total,

the exercise sessions (500 kcal per session) and daily physical activity (targeting an additional 500 kcal) were aimed at achieving a total caloric expenditure of 2000 kcal per week.

The following is a list of further exclusion criteria for study participants.

(1) Participants were excluded if they suffered from ischemic heart disease, epilepsy or insulin allergy, as part of the study involved using glucose clamps and insulin infusion protocols to avoid the risk of precipitating epileptic attacks or aggravating such underlying conditions.
(2) Participants with bleeding diathesis or inaccessible veins were excluded as phlebotomy and venipunctures for venous access are required for the study.
(3) Participants on treatment for hypertension, dyslipidemia, and diabetes mellitus were excluded as the study's purpose was to examine the degree of physiological and metabolic changes among an overtly healthy group, and being on treatment would interfere with the interpretation of these parameters. Participants with metabolic syndrome were included if they had not been treated for the abovementioned conditions. These participants were included as it is expected that a significant number of overweight and obese participants with metabolic syndrome may be asymptomatic and thus appear overtly healthy. Capturing such variations in the population would be useful for studying the effects of weight loss.
(4) Participants with significant recent changes in body weight or actively attempting to lose weight through dieting, bariatric surgery, or anti-obesity drugs were excluded, given the need to recruit participants with stable weight of a given narrow distribution of body mass index.
(5) Participants on glucocorticoids or any drugs known to alter insulin resistance (e.g., ACE inhibitors, protease inhibitors, metformin, etc.) were excluded to avoid confounding factors in the measurement of insulin sensitivity and insulin resistance required in this study.
(6) Participants on an investigational drug in the past 6 months were excluded as it would not be predictable if the agent affected insulin resistance.
(7) Participants with contraindications to Dual X-ray Absorptiometry (DEXA) and Magnetic resonance imaging (MRI) (e.g., Cardiac pacemaker or metallic implants or suffering from claustrophobia) as the study entails the use of the DEXA and MRI.
(8) Participants with psychological, eating, or motivational disorders were excluded from the study to increase weight loss rates during the intervention.
(9) Participants whose birth records were unavailable or whose mothers are deceased and hence unable to provide accurate birth weight data.
(10) Participants from the extreme birthweight range (below the 5^{th} percentile or above the 95^{th} percentile) were excluded as extreme birth weights may impact the risk of diabetes.

METHOD DETAILS

Study design for assessment of weight loss outcomes

In the SAMS2 run-in phase, participants were instructed to maintain or discontinue specific behaviors to stabilize baseline physiological measurements and prepare them for the intervention study. Strenuous physical activities, such as long-distance walking or swimming, were restricted for 3–7 days before the study to prevent effects on lipid levels. Participants completed a 3-day food diary and visited the study center for measurements. Pre-intervention baseline measurements were taken from the subjects. This included basic anthropometrics (weight, height, waist circumference, hip circumference), 24-h food recalls, and indirect calorimetry (for resting energy expenditure and a mixed meal tolerance test). DEXA, MRI, MRS and Tanita body composition analyzer (bioimpedance) were used to measure the body composition. The euglycemic hyperinsulinemic glucose clamp was done to determine the insulin sensitivity index of the subjects. Tissue samples were taken, including fasting blood sample, urine samples, buccal smear, and a percutaneous skeletal muscle biopsy using a BioPince full core biopsy gun. The muscle biopsy was split into two parts; one was stored in PBS in a Falcon tube used for myoblast culture, while the additional 3–4 full cores were flash-frozen in liquid nitrogen Dewer flask for RNA extraction and sequencing. The same measurements and biopsies were taken post-weight loss to assess the outcomes of the weight loss intervention. The Student's t test for paired samples was used to assess the statistical significance of differences in clinical data before and after the intervention.

Weight maintenance phase

After completing the weight loss intervention, participants who met the criteria were invited back for the 12-month weight maintenance phase. During this phase, no interventions, exercise, or dietary modifications were implemented. Participants were followed up with visits every three months (at the 3-, 6-, 9-, and 12-month marks) to monitor their progress. At each visit, participants underwent fasting clinical evaluations, including body composition assessments using Tanita and DEXA, as well as the MRI of the abdomen. Additionally, Magnetic Resonance Spectroscopy (MRS) was performed on the liver and muscle, and fasting glucose and insulin levels were measured. It is important to note that the clinical evaluations conducted during this phase represented a narrowed subset of those performed at the pre- and post-intervention visits, and participants were not phenotyped as comprehensively as in the earlier stages of the study.

Anthropometrics

The height and weight of subjects were measured using a wall-mounted stadiometer and digital scale respectively. Their BMI was then calculated using the subject's weight (in kilograms) divided by the square of his height (in meters). We also measured waist circumference as the midpoint between the lower costal margin and iliac crest during mid-respiration. Hip circumference measurements were taken around the widest part of the buttocks.

Dual-energy X-Ray absorptiometry (DXA)

Body composition was quantified using the Hological Discovery Wi DXA scanner. In the Hologic machine, subjects were positioned according to the standard protocol as they all fit within the measurement area. After measurement, three trained dual-energy X-ray absorptiometry technologists certified by the International Society of Clinical Densitometry (ISCD) conducted manual analyses to identify regional soft tissue demarcation.

Magnetic Resonance Imaging

Abdominal MRI was performed using Siemens Tim Trio 3T MR scanner. Axial images of the abdomen with 3 mm slice thickness, interslice gap of 0.6 mm, and an in-plane resolution of 1.25 $\times$ 1.25 mm were acquired using a two-point Dixon sequence (repetition time (TR) = 5.28 ms, echo time (TE)1 = 2.45 ms, TE2 = 3.68 ms) and body matrix coil after anatomical localization. The image slices covering the abdominal region between L1 and L5 vertebrae were acquired during a breath-hold of 18–20 s. A fully automatic graph theoretic segmentation algorithm followed by level sets was used to segment and quantify subcutaneous adipose tissue (SAT) – deep (dSAT) and superficial (sSAT) and visceral adipose tissue (VAT) depots.[100]

MR spectroscopy of liver and skeletal muscle

Liver fat was measured by proton magnetic resonance spectroscopy from 2 $\times$ 2 $\times$ 2 cm^3 voxels localized in the left and right lobe of the liver using body matrix coil. Care was taken to avoid visible vessels and the liver boundary. Within each voxel, spectra were acquired using a point-resolved spectroscopy (PRESS) sequence (TR = 2000 ms, TE = 30 ms). Respiratory motion was handled by breath-holding. The acquired spectra were fitted using the linear combination of model spectra (LCModel).[101] The liver fat was determined from the concentration of methyl and methylene groups of lipids and the unsuppressed water signal from each spectrum, corrected for T2 losses and averaged. The muscle spectrum was obtained from a 2 $\times$ 2 $\times$ 2 cm^3 voxel within the soleus muscle using a PRESS sequence (TR = 2000 ms, TE = 30 ms) and knee coil. The spectrum was processed using LCModel and the amount of IMCL was calculated and expressed as the IMCL-to-creatine ratio.

Peripheral blood biochemical measurements

Peripheral blood samples were obtained from subjects after a 10-h overnight fast to survey the lipids, liver function, thyroid-stimulating hormone, sodium, and lactate dehydrogenase (LDH). These biochemical analyses were done in the National University Hospital Referral Laboratory (Singapore) accredited by the College of American Pathologists.

Accelerometer

Study subjects wore the accelerometer under free-living conditions to measure physical activity levels. This data was presented as mega counts per day (mcnts/day).

Doubly labeled water technique for measurement of total energy expenditure (TEE)

To assess energy expenditure, we used the doubly labeled water technique to precisely measure energy expenditure without requiring an in-patient stay, which may modify the participants' energy expenditure. The body water of the subjects was enriched with a stable isotope of hydrogen (^2H – deuterium) and oxygen (^{18}O), which was used to determine the washout kinetics of both isotopes as their concentrations declined exponentially toward natural abundance levels. After providing a baseline urine sample, each subject was given a known quantity of doubly labeled water based on an estimate of total body water and used a mixture of 10% ^{18}O and 5% ^2H (i.e., does of approximately 100g for a 70kg subject with 25% body fat). Urine specimens were collected according to a schedule after the ingestion of the doubly labeled water over 2 weeks. Participants were instructed not to collect the first daily void of urine. All urine samples were transferred into aliquot tubes and frozen. Baseline and post-does urine samples were analyzed for deuterium and ^{18}O isotopic enrichment by gas isotope ratio mass spectrometry. Fractional turnover rates for each isotope and their dilution spaces were computed and used to compute the CO_2 expiration rate (rCO_2). This, together with the oxygen consumption rate (rO_2) and urinary nitrogen (UN) was converted to TEE using the Weir equation as follows: TEE (kcal/day) = 22.4 * (1.106 * rCO_2 +3.941 * rO_2) $-$2.17 * UN.

Indirect calorimetry – Mixed meal tolerance test

The resting metabolic rate (RMR) was determined after an overnight fast and minimal activity for about two days before testing using an indirect calorimeter mounted on a metabolic cart. Activity level was then estimated using TEE – RMR. Using a ventilated hood to assess resting energy expenditure, participants will undergo indirect calorimetry. After this had been established, participants consumed a standard liquid meal that included 30% of energy from fat (10% for saturated fats, 10% from polyunsaturated fats,

and 10% from monounsaturated fats), 15% energy from protein, and 55% energy from carbohydrates. The respiratory quotient (RQ), fat and carbohydrate levels at these time points were measured. To quantify the changes in the levels of these measurements because of the mixed meal measurements, the baseline was subtracted from the measurement at 1.5 h after the mixed meal. The changes in RQ also served as a measure of metabolic flexibility.

Differential clinical traits

Before evaluating the effects of intervention on clinical traits, we addressed the presence of minimal missing data within our dataset. Using the 'mice' package in R,[102] we performed multiple imputations separately for pre-intervention and post-intervention measurements to generate complete datasets for further analysis. This approach allowed us to maintain the integrity and robustness of subsequent statistical tests. To evaluate the effects of the intervention on clinical traits, we conducted a paired t-test for each clinical trait measured before and after the intervention. Due to the many traits analyzed, it was necessary to adjust for multiple tests to avoid an inflation of the false positive rate.[103] Clinical traits that exhibited a q-value of less than 0.05 after correction were considered statistically significant. Further, we computed correlations between the changes in various clinical traits from pre-to post-intervention using the 'cor' function in R.[104] These correlations were visualized using a heatmap.[105]

Lipidomes profiling analysis

To comprehensively analyze the lipidomics profiles within this dataset, we employed Liquid Chromatography-Tandem Mass Spectrometry (LC-MS/MS) to measure a total of 201 distinct lipids. This method allows for the precise separation, identification, and quantification of a wide range of molecular species, including both small and complex lipids. To identify differential lipids, we utilized the MetaboAnalystR package.[106] Significant differential metabolites were selected based on several criteria: A variable importance in projection (VIP) score from the partial least squares-discriminant analysis (PLS-DA) model of at least 1, a p-value below 0.05, and a fold change threshold of 1.2. Subsequently, we visualized these significant metabolites using the volcano plot function in the ggplot2 R package.

Genotyping and imputation

We performed genotyping using InfiniumOm2-5Exome-8v1-3_A1 according to the manufacturer's protocol. We exported genotype data into PLINK (v1.9)[107] data format using GenomeStudio[108] PLINK Input Report Plug-in v2.1.4 and performed sample QC and variant QC. For sample QC, no individuals were removed. We calculated the call ratio and heterozygous variant ratio per sample using PLINK. Call ratios of all individuals are greater than 0.98, which indicates good genotyping quality. Next, we estimated genetic sex using intensity ratios (F index) for X chromosomes. We defined males as $F > 0.8$ and females as $F < 0.2$. Genetic sex and self-reported sex agreed with each other. To infer kinship between samples, we calculated PI_HAT and Z1 using PLINK to identify related samples and there is no relatedness in our samples. For variant QC, we exclude autosomal variants with low genotyping quality. Variants with suboptimal genotyping quality (call ratio <0.95) were removed. We only keep those variants with a minor allele frequency (MAF) > 0.05. We computed the Hardy-Weinberg Equilibrium (HWE) p-values and removed 4 variants that did not meet the HWE threshold of $p < 1 \times 10^{-6}$. A total of 742,597 autosomal variants remained after all variant QC steps. These post-QC variants were further used for genome-wide imputation using the Michigan Imputation Server,[50] and no additional QC on the imputation server was performed. We used all populations in 1000G Phase 3 high-coverage (GRCh38/hg38) as a reference panel. Following imputation, only those imputed variants with imputation quality $R^2 > 0.8$ were retained. Variants with MAF <0.05 and HWE >1 $\times$ 10^{-6} and biallelic were excluded. A total of 5,126,984 SNPs were used for downstream QTL analysis. We merged these SNPs with the 1000 Genomes Project (1KG) (N = 2,504)[69] for PCA analysis. Extensive linkage disequilibrium in dense genotyping data biases principal component-based ancestry estimates. To reduce linkage disequilibrium, we pruned our dataset using a greedy algorithm implemented in PLINK. Finally, we performed PCA on merged SAMS2 and 1KG samples using PLINK.

Skeletal muscle RNA sequencing and quality control

Total RNA was isolated from the snap frozen samples of skeletal muscle biopsies from the subject's vastus lateralis using NucleoSpin kit. RNA-seq was done on these extracted RNA using HiSeq4000 (Illumina), sequencing up to 80 million bidirectional reads of up to 151 base pairs. FASTP (v0.23.4)[109] was used to trim the adapter sequences and low-quality bases in read sequences. These quality-filtered sequence reads were aligned to the human reference genome GRCh38 primary assembly using STARsolo (v2.7.19a)[110] and the GENCODE v43 annotation file[111] was used as the reference GTF file. The resulting bam files generated feature read counts using the FeatureCounts (v2.0.3).[112] The -waspOutputMode option was used to avoid potential artifacts from allelic mapping bias. The corresponding quality-controlled genotype files were used as input for the wasp-filtering classification of each sample. Genes with an average read count below ten and with zero counts in more that 20% of samples were considered not expressed and filtered.

Skeletal muscle transcriptome profiling and network analysis

We focused on differentially expressed protein-coding and long lincRNA genes in skeletal muscle RNA-seq using DESeq2,[113] comparing gene expression profiles after the intervention to those before. All individuals are male, and all paired samples (pre-intervention and post-intervention for the same sample) are sequenced in the same batch. The primary interest is in detecting differentially

Please cite this article in press as: Wang et al., Impact of polymorphisms on gene expression and splicing in response to exercise and diet-induced weight loss in human skeletal muscle tissues, Cell Genomics (2025), https://doi.org/10.1016/j.xgen.2025.100951

expressed genes between the pre-intervention and post-intervention; considering the pairing samples as covariates, we fit the following linear mixed model for each gene:

$$E(\text{expression}) = \beta_0 + \beta_t * \text{condition} + \text{ID} + \varepsilon$$

Where *E* stands for expressions, t stands for pre/post-intervention.

Genes with log2 |fold change (FC)| > 0.3 and FDR <0.05 were considered differentially expressed genes (DEGs). Pathway enrichment analyses for DE genes were conducted using over-representation analysis (ORA),[114] a computational method determining whether a set of genes defined *a priori* exhibits statistically significant concordant differences between two biological states (pre-intervention and post-intervention). Resulting FDR are adjusted for multiple testing using the qvalue methods controlling at 5%.

To further identify gene co-expression patterns associated with the intervention, we applied weighted gene co-expression network analysis (WGCNA) on the skeletal muscle RNA-seq data. WGCNA[44] was performed to construct gene modules based on correlation networks, allowing us to explore gene sets of highly co-expressed genes. To generate a biologically meaningful input set for co-expression network construction, we applied a custom filtering script that retains only genes meeting two criteria: (1) expression levels (FPKM) above 1 in at least two-thirds of both Pre and Post samples, and (2) a relative difference in mean expression (Pre vs. Post) exceeding 0.2. This stringent approach ensures that only sufficiently expressed and notably varying genes are included, thereby enhancing the reliability of subsequent network analyses. Using the dynamic tree-cutting method (deepSpit = 2, minModuleSize = 30) followed by module merging at a specified threshold (mergeCutHeight = 0.25), we identified three modules in our dataset: the brown module (115 genes), the gray module (171 genes), and the blue module (88 genes). Among the three modules, only the gray module showed a notably higher correlation between the two conditions (r = 0.57, p = 1 × 10-10). To visualize interactions among genes in this module, we used Cytoscape, applying an edge-weight threshold of 0.2 to focus on the most biologically relevant connections. While no single "super hub" gene emerged, a cluster of mitochondrial genes (MT-ND2, MT-ND5, MT-ND6, and MT-ATP8) was conspicuous. Given their well-established roles in muscle metabolism, this mitochondrial gene cluster suggests that alterations in mitochondrial function may underlie the module's correlation with weight loss. We calculated the module eigengenes (MEs) and correlated it with the intervention condition. Network analysis within the significantly correlated module was visualized using Cytoscape.[115]

Quantification of mRNA alternative splicing

We used the LeafCutter[116] pipeline to identify alternatively excised introns by pooling all junction reads to quantify alternative splicing events. The GENCODE v43 annotation file was used as the reference annotation file. We used LeafCutter to quantify intron usage levels with the following steps: (1) We used RegTools '- *junctions extract'* to analyze CIGAR strings in each bam file to quantify the usage of each intron with the options –Min_anchor_length 8, –Min_intron_length 50, –Max_intron_length 500000, –Strand_specificity FR; (2) Intron clusters were generated using leafcutter_cluster_regtools.py script from LeafCutter with the options –min_clu_reads 50, –max_intron_len 500000; (3) prepare_phenotype_table.py script from LeafCutter was used to normalize the counts and the resulting per-chromosome files were merged and converted to BED format. Introns without any read counts in at least 40% of the samples or too little variation (standard deviation <0.005) were removed in this procedure. This resulted in 101,492 alternatively excised introns from 29,923 clusters. Each cluster comprises on average 3.4 introns.

Differential alternative splicing analysis

We performed differential splicing with LeafCutter,[116] using age and five genotype PCs as covariates. Default parameters were used (-i 5 -g 3; a minimum of five samples per intron and a minimum of three samples per group). The *p*-value were adjusted using the Benjamin-Hochberg procedure. We call the intervention differential splicing with FDR <0.05. Sashimi plots to show the junction usage difference of interested position were generated by ggsashimi.[117]

Expression quantitative trait loci (eQTLs) with RASQUAL

Prior to eQTL analysis, we first filtered the expression matrices for both pre- and post-intervention conditions to ensure reliable and efficient RASQUAL implementation. Specifically, genes with average read count below 10 and with zero counts in more than 20% of samples were considered not expressed and removed, as low-expressed genes could lead to model convergence issues in RASQUAL. We determined age, genomic ancestry and hidden confounders of phenotype calculated by PEER.[118] We performed covariate selection by empirically maximizing the power to detect eQTL. We randomly selected three chromosomes to perform covariate selection for computation feasibility and avoid overfitting. We added age, genotype PCs (maximum of ten) and hidden confounders (maximum of ten) sequentially. We tested the top three genotype PCs because they explained most of the variability in the genotyping data. After multiple hypothesis correction, the number of eAssociations (defined as an SNP-Gene pair that passed hierarchical multiple hypothesis testing by TreeQTL) increased for Pre and Post conditions as the number of covariates increased. Therefore, we decided to use age, top 2 genotype PCs and top 2 hidden confounders of phenotype both for Pre and Post conditions, respectively. Mathematically, the model is the following:

$$E(\text{expression}) = \beta_0 + \beta_g * \text{genotype} + \beta_a * \text{Age} + \sum_{i=1}^{2} \beta_g * \text{PC} + \sum_{i=1}^{n} \beta_g * \text{HF}$$

Where *E* stands for expressions, g stands for genotypes, PC stands for genotype principal components, and HF stands for hidden factors. And $n = 2$ both for Pre and Post conditions.

We mapped eQTL using RASQUAL,[51] which integrated total read count with allele-specific expression (ASE) to boost power for eQTL mapping. To obtain GC-corrected library size, we first calculated GC content using GENCODE v43 by taking the average GC content of the exons of all genes. Next, we calculated GC-corrected library sizes based on read count output from FeatureCounts. We determined whether the *p*-values were inflated by visualizing their distribution. The *p*-values showed a uniform distribution with a spike near 0, indicating a strong signal of significant associations. The upward trend in the QQ plots demonstrates clear enrichment of significant eQTLs. Furthermore, eQTLs with low *p*-values were predominantly enriched around transcription start sites (TSS), with those located closer to the TSS showing higher effect sizes. We obtained gene-level and association-level FDR using a hierarchical hypothesis correction procedure implemented in TreeQTL. TreeQTL uses a hierarchical FDR correction procedure, which performs FDR correction first on the gene level and then on the association level (gene by SNP). We used FDR <0.05 on both gene and association levels. To determine whether the *p*-values were inflated, we visualized the distribution of all *p*-values. The spike around zero and upward trend in the QQ plots show clear enrichment for significant eQTL. As expected, eQTLs with low *p*-values were enriched around transcription start sites. We further performed multi-tissue eQTL calling using the RASQUAL *p*-values and the multi-tissue version of TreeQTL. We set the gene as the first level, the intervention as the second level, and the gene-intervention-SNP as the third level and used the default FDR <0.05 cutoff for all three levels.

Splicing quantitative trait loci (sQTLs) with QTLtools

We performed sQTL mapping with QTLtools[70] using the intron excision ratios and a *cis*-window of 1Mb up- and downstream of the intron junction. The linear model used age, PCs derived from splicing ratios and 2 genotype PCs as covariates. Mathematically, the model is the following:

$$E(\text{intron}) = \beta_0 + \beta_g * \text{genotype} + \beta_a * \text{Age} + \sum_{i=1}^{2} \beta_g * \text{PC} + \sum_{i=1}^{n} \beta_g * \text{HF}$$

Where *s* stands for the ratio between reads overlapping each intron and the total number of reads over-lapping the intron cluster (to obtain cluster-level *p*-values, we used a conservative approach to correct for family-wise error rate with the Bonferroni procedure across introns within each cluster), g stands for genotypes, PC stands for genotype principal components, and HF stands for hidden factors. And $n = 1$ both for Pre and Post conditions.

To identify top nominal sQTL for an intron cluster, QTLtools was run using the permutation pass mode (1,000) permutations, and beta-approximated permutation *p*-values were adjusted for multiple test correlation per phenotype using Storey's q-value in the qvalue R package.[119] The significant threshold was set at FDR <0.05. We obtained cluster-based nominal *p*-value thresholds and then conducted the conditional analysis.

Independent sQTL mapping

To identify multiple independent regulatory signals per cluster, we performed conditional analysis using forward stepwise regression followed by backward selection for a given splicing ratio phenotype described previously. A scan for sQTLs using QTLtools[70] was used to correct previously discovered sIntron-SNP pairs, and the covariates used in standard sQTL mapping. If the adjusted *p*-value for the lead variant was not significant at the cluster-level threshold, the forward stage was finished, and the procedure proceeded to the backward stage. If the adjusted *p*-value was significant, the lead variant was included in the list of sQTLs as an independent signal, and the forward step proceeded to the next iteration. Testing each sIntron-SNP pair was done in the backward stage. Only significant sIntron-SNP pairs in the backward stage were kept as independent sQTLs. And if the sIntron-SNP pairs were not significant in the test of the backward stage, they would be dropped. As a quality control, we determined whether the *p*-values were inflated by visualizing their distribution. The *p*-values showed a uniform distribution with a spike near 0. The upward trend in the QQ plots shows clear enrichment for significant sQTL. Further, sQTLs with low *p*-values were enriched around splicing donor and acceptor sites, and intronic sQTL SNPs were enriched at intron boundaries.

e/sQTL replication

We downloaded the entire set of skeletal muscle SNP-gene and SNP-Intron association tests for 361 GTEx skeletal muscle samples from GTEx portal.[68] During replication, we selected *cis*-eQTLs and *cis*-sQTLs from each cell type and queried their *p*-values in above GTEx datasets. This replication was quantified using the $\pi 1$ statistic implemented in the qvalue R package.[119]

Identification of gene-by-lifestyle interactions e/sQTLs

We systematically classified the significant e/sQTLs based on their detection patterns across pre- and post-intervention conditions into three categories: Pre-biased, Post-biased, and Shared QTLs. Pre-biased QTLs were exclusively significant in the pre-intervention condition, while Post-biased QTLs showed significance only after intervention. QTLs that demonstrated significance in both conditions (affecting the same eGene or sIntron) were categorized as Shared e/sQTLs. While context-biased QTLs (Pre-biased

and Post-biased) inherently represent gene-by-lifestyle interactions, Shared QTLs may also harbor significant gene-by-lifestyle effects despite showing significance in both conditions, as their effect sizes and significance levels could differ substantially between conditions. To further identify gene-by-lifestyle interaction QTLs for significant e/sQTLs observed in both pre- and post-intervention conditions, we employed a linear regression model:

$$Y = \beta_0 + \beta_1 G + \beta_2 L + \beta_3 (G \times L) + \beta_i PC_i + \varepsilon$$

Y represents either gene expression (for eQTLs) or normalized junction usage ratio (for sQTLs), G denotes genotype, L represents lifestyle intervention status, and (G×L) is their interaction term.

We included relevant covariates, such as phenotype PCs, to account for potential confounding factors. The model was fitted for each QTL, and those with an FDR <0.05 for the interaction term (G×L) were classified as significant, indicating a gene-by-lifestyle interaction effect on the molecular phenotype of interest.

Asian-specific e/sQTLs

To identify Asian-specific e/sQTLs, we leveraged the eQTL and sQTL results obtained from SAMS2, which contains East Asian-derived muscle samples. We cross-referenced these significant e/sQTLs with those in the GTEx database,[67,68] which primarily comprises European populations. By comparing the minor allele frequencies of East Asian-specific e/sQTL SNPs with those in European populations using data from the 1000 Genomes Project,[69] we identified variants common (MAF >0.05) in East Asians but rare (MAF <0.05) in Europeans. These e/sQTLs were thus classified as Asian-specific e/sQTLs, as their elevated frequency in East Asians suggests they are likely to drive regulatory differences that are unique to this population.

GWAS summary statistics

We organized publicly available Asian-specific GWAS summary statistics of 12 phenotypes covering a broad range of categories[4,79,80] (BMI, body mass index; BW, Body Weight; HbA1c, glycated hemoglobin; Glucose; TC, Total Cholesterol; TG, triglycerides; HDL, High-Density Lipoprotein; LDL, Low-Density Lipoprotein; T2D, type-2 diabetes; CAD, coronary artery disease).

Colocalization between GWAS and e/sQTLs

We performed colocalization analysis using the organized GWAS summary statistics to determine whether the risk variants at GWAS loci will likely be the same regulatory variants for eQTL or sQTL. To assess colocalization between GWAS and eQTL or sQTL signals, lead GWAS variants with p-values below 5×10^{-7} were selected. The HLA region (chr6:25-35Mb) was excluded from the colocalization analysis. For each GWAS loci identified above, colocalization was tested only if it occupied a QTL with a nominal p-value below 1×10^{-3}. All GWAS variants were sorted by p-values in ascending order to identify independent GWAS loci. Starting from the variant with the most significant p-value, the remaining variants within a 500kb window on either side of the selected lead variant were removed. The same procedure was then applied to the next most significant variant among the remaining variants until the last variant in the list was reached. For each locus, we then tested colocalization between GWAS and QTL signals using the 'coloc.abf' function of the COLOC R package.[81] The MAF of the SAMS2 was used to estimate sdY. PP4 > 0.75 was set as the threshold for significant colocalization. The colocalization results were visualized using LocusCompare.[120]

Fine-mapping

Fine-mapping analysis was conducted to identify potential causal variants. The analysis was performed using the R package SuSiE.[82] All SNPs within a 200kb window centered around the lead SNP of targets QTL were selected as candidate causal SNPs in this analysis. Function 'susie_rss()' was used to calculate the posterior inclusion probabilities (PIP) for these SNPs using QTL summary statistics. The LD matrix calculated by LDlink[121] based on SAMS2 donors was used as the correlation matrix in this process.