NAR Molecular Medicine, 2025, **2**, ugaf010
https://doi.org/10.1093/narmme/ugaf010
Advance access publication date: 28 March 2025
Standard Article

Sex-specific skeletal muscle gene expression responses to exercise reveal novel direct mediators of insulin sensitivity change

Sisi Ma[1,†], Monica J. Hubal[2,†], Matthew C. Morris[3,†], Leanna M. Ross[4], Kim M. Huffman[4], Christopher G. Vann[4], Nadia Moore[4], Elizabeth R. Hauser[4], Akshay Bareja[4], Rong Jiang[5], Eric Kummerfeld[1], Matthew D. Barberio[6], Joseph A. Houmard[7], William C. Bennett[4], Johanna L. Johnson[4], James A. Timmons[8], Gordon Broderick[3], Virginia B. Kraus[4], Constantin F. Aliferis[1], William E. Kraus[4,*]

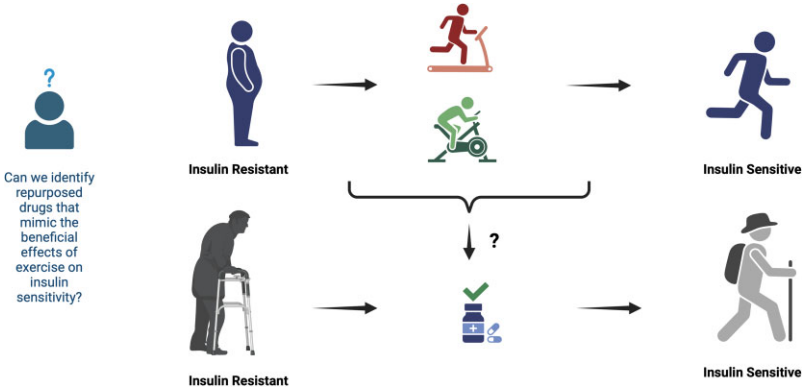
[1]Institute for Health Informatics (IHI), Academic Health Center, University of Minnesota, Minneapolis, MN 55455, United States
[2]Department of Kinesiology, Indiana University Indianapolis, Indianapolis, IN 46202, United States
[3]Center for Clinical Systems Biology, Rochester General Hospital, Rochester, NY 14621, United States
[4]Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC 27701, United States
[5]Department of Head and Neck Surgery & Communication Sciences, Duke University School of Medicine, Durham, NC 27701, United States
[6]Department of Exercise and Nutrition Sciences, George Washington University, Washington DC 20052, United States
[7]Department of Kinesiology, ECU, Greenville, NC 27858, United States
[8]School of Medicine and Dentistry, Queen Mary University of London, EC1M 6BQ, United Kingdom

[*]To whom correspondence should be addressed. Email: william.kraus@duke.edu
[†]Equal contribution.

Abstract

Understanding how exercise improves whole-body insulin sensitivity (Si) involves complex molecular signaling. This study examines skeletal muscle gene expression changes related to Si, considering sex differences, exercise amount, and intensity to identify pharmacologic targets mimicking exercise benefits. Fifty-three participants from STRRIDE (Studies of Targeted Risk Reduction Interventions through Defined Exercise) I and II completed eight months of aerobic training. Gene expression was assessed via Affymetrix and Illumina technologies, and Si was measured using intravenous glucose tolerance tests. A novel discovery protocol integrating literature-derived and data-driven modeling identified causal pathways and direct transcriptional targets. In women, exercise amount primarily influenced transcription factor targets, which were generally inhibitory, while in men, exercise intensity drove activating targets. Common transcription factors included ATF1, CEBPA, BACH2, and STAT1. Si-related transcriptional targets included TACR3 and TMC7 for intensity-driven effects, and GRIN3B and EIF3B for amount-driven effects. Two key pathways mediating Si improvements were identified: estrogen signaling and protein kinase C (PKC) signaling, both converging on the epidermal growth factor receptor (EGFR) and other relevant targets. The molecular pathways underlying Si improvements varied by sex and exercise parameters, highlighting potential skeletal muscle-specific drug targets such as EGFR to replicate the metabolic benefits of exercise.

Graphical abstract



Received: October 20, 2024. **Revised:** March 7, 2025. **Editorial Decision:** March 10, 2025. **Accepted:** March 27, 2025

Introduction

Exercise training provides substantial health benefits; however, too few individuals adopt and maintain it as a lifelong health strategy [1]. Consequently, there is significant interest in developing pharmacologic alternatives able to replicate the health effects of exercise [2]. To facilitate such pharmacologic development, a deeper understanding is needed of the complex, pleiotropic, and sex-specific physiological effects of exercise, which occur across multiple organ systems. Identifying specific molecular mediators and causal pathways that connect specific exercise regimens to specific health outcomes is a crucial step in developing therapeutics that mimic the effects of exercise.

Given the substantial role of whole-body insulin action as a marker and mediator of cardiometabolic risk and dysfunction, skeletal muscle insulin sensitivity (insulin sensitivity index, Si) presents a potentially powerful pharmaceutical target. Extensive literature in both humans and animals indicates that exercise-induced improvements in whole-body insulin sensitivity are closely linked to molecular processes and adaptations in skeletal muscle [3–6]. Based on this rationale, the purpose of this study is to causally model the effects of aerobic exercise training on the skeletal muscle transcriptome and its functional relationship to changes in Si [7].

This study leverages data from the STRRIDE (Studies of Targeted Risk Reduction Interventions through Defined Exercise) series, which examined the effects of varying amounts, intensities, and modes of exercise training on cardiometabolic disease risk factors [8–11]. Conducted between 1998 and 2013, the three STRRIDE studies were designed to explore the temporal effects of eight months of exercise training and subsequent detraining on key clinical cardiometabolic and physiological outcomes. They also investigated the dose-response and mode-specific effects of exercise on these outcomes, with a focus on the molecular mechanisms in skeletal muscle mediating these effects. These studies have produced a robust repository of demographic, clinical, and molecular data from 920 enrollees and 580 completers across the three cohorts.

For this study, we utilized data from a subset of participants who completed aerobic exercise training in STRRIDE I and II. We hypothesized that transcription factor targets would be influenced by biological sex and specific parameters of the exercise training programs (e.g. amount and intensity). To identify regulatory and regulated elements leading to exercise-induced changes in Si, we employed two complementary approaches: integrative molecular physiology and advanced machine learning and causal discovery methods.

Materials and methods

Study cohort

This analysis focused on the aerobic exercise training groups from the STRRIDE I and II studies (NCT00200993 and NCT00275145) [8, 9]. This study was conducted under the oversight of the Duke University IRB; all participants agreed to participate by signing an IRB-approved consent form. Exercise training groups were categorized using a two-digit code based on the exercise program's amount and intensity: low amount (1) of moderate intensity (1), low amount (1) of vigorous intensity (2), and high amount (2) of vigorous intensity (2). These categories were labeled as 1–1, 1–2, or 2–2, allowing us to study the effects of exercise amount while controlling for intensity, and the converse (Supplementary Table S1). The exercise amount was prescribed as kilocalories expended per kilogram of body weight per week (KKW), with low amount defined as 14 KKW and high amount as 23 KKW. Exercise intensity was prescribed relative to participants' baseline peak oxygen consumption ($\dot{V}O2$), assessed through a maximal cardiopulmonary exercise test. Moderate intensity was set at 40%–55% of peak $\dot{V}O2$, and vigorous intensity at 65%–80% [8]. Participants adhered to their assigned exercise protocol for eight months, with a median adherence rate of 91.0% (Interquartile ranges (IQR) 79.8%–99.8%).

Participant characteristics

The STRRIDE I and II studies recruited physically inactive adults (defined as fewer than one self-reported exercise session per week) aged 18–70 years, who had overweight or class I obesity (BMI 25–35 kg/m^2), dyslipidemia, and metabolic syndrome, but without overt coronary artery disease or diabetes [9]. Data for this analysis were drawn from 53 participants who had both pre- and post-training insulin sensitivity index (Si) measured using the FSIVGTT (described below), and skeletal muscle genome-wide gene expression data meeting quality control standards. All women qualifying for STRRIDE I and those selected from STRRIDE II for this analysis were postmenopausal. Of 49 total participants selected for the initial analysis, 26 were women and 10 were on hormone replacement therapy.

Muscle sampling and FSIVGTT

Muscle biopsies were taken from the vastus lateralis before the initiation of the exercise training protocol (pre) and 16–24 h after the last training session (post) [12]. Total muscle RNA was prepared as described and used for gene expression analyses [12]. The FSIVGTT was conducted over three hours to assess blood glucose, insulin, and modeled Si at both pre- and post-training time points for all participants [13].

Skeletal muscle genome-wide gene expression

As previously described [12], muscle gene expression data were obtained for 39 participants using the Affymetrix HU U133 Plus 2.0 chip and for 42 participants using the Illumina HT-12 v4 Expression chip. The data were harmonized using standard NCBI gene identifiers, with 28 participants having data available on both platforms. For the Affymetrix analysis, there were five men and five women from each intervention group. A summary of participant demographics is provided in Table 1. Data supporting this analysis are available at the following GEO registrations: U133 Plus Series, Accession: GSE47969, ID: 200047969; Illumina Series, Accession: GSE83352, ID:200083352.

Overall analytic strategy to identify molecular targets of exercise-induced changes in insulin sensitivity

The overall analytic strategy is illustrated in Fig. 1. Initially, we used a previously described approach [14–16] to construct a regulatory circuit model based on our entire observational gene expression dataset, which included harmonized Affymetrix and Illumina data. This model related exercise intensity and amount to changes in muscle transcription factor

Table 1. Effects of exercise training on insulin action in study cohorts

	STRRIDE I and II		Discovery (Affymetrix)		Validation (Illumina)	
	Pre	Post	Pre	Post	Pre	Post
N (m,f)	319 (171 148)		39 (21,18)		42 (19,23)	
Age, y	50 ± 9		52 ± 8		51 ± 10	
Body mass						
Height (cm)	179 ± 9		169 ± 10		168 ± 9.8	
Mass (kg)	88.0 ± 13.2	87.0 ± 13.3	86.6 ± 13.4	85.0 ± 13.4	84.2 ± 10.9	83.8 ± 11.1
BMI (kg/m^2)	30.0 ± 3.0	29.5 ± 3.5	30.4 ± 2.8	29.8 ± 2.7	29.7 ± 2.8	29.5 ± 2.8
Insulin action						
Si (mU/l/min)	3.9 ± 3.4	4.8 ± 4.1*	3.8 ± 2.6	4.9 ± 3.2*	4.5 ± 3.0	6.6 ± 5.7*
Change in Si (%)	46 ± 84*		44 ± 62*		61 ± 88*	
Fasting insulin (U/l)	9.3 ± 5.8	8.2 ± 4.8*	8.8 ± 6.1	7.5 ± 4.2	8.6 ± 6.7	7.5 ± 4.2
Fasting glucose (mg/dl)	94.3 ± 10.3	94.3 ± 9.5	95.0 ± 10.4	95.8 ± 9.7	92.6 ± 11.6	94.2 ± 9.3

Data are means ± SD. *$P < 0.05$ compared to Pre.

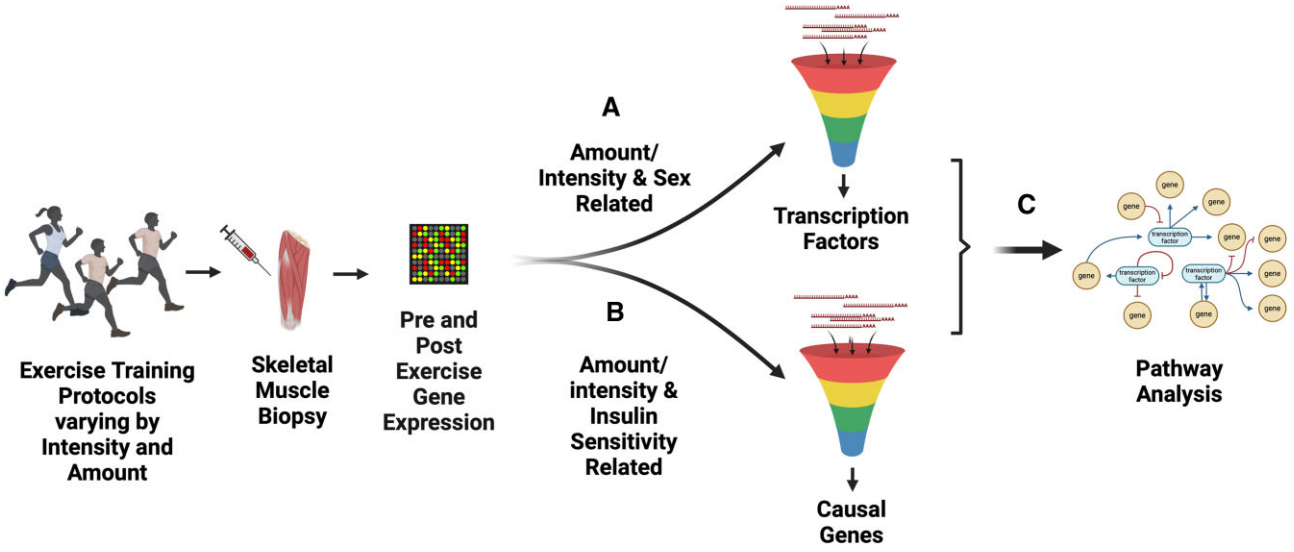


Figure 1. Conceptual design. A candidate list of gene transcription targets was developed through three approaches: (**A**) identification of a set of gene transcription factors modified by exercise depending on amount and intensity of exercise and biological sex constrained by prior knowledge. (**B**) Identification of a set of genes causally related to and constrained by their relation to change in insulin sensitivity as defined by the insulin sensitivity index. (**C**) Identification of the intersection of these two gene sets were used to identify gene expression networks and gene regulatory nodes causally connecting exercise of different amounts and intensities to changes in insulin sensitivity. Created in BioRender. Kraus, W. (2024) BioRender.com/t18t722.

expression, resulting in a partially-directed causal network curated with existing knowledge from the literature using the Elsevier Pathway Studio database [17] (Fig. 1, step A). In parallel, we applied causal discovery Markov boundary induction methods [18, 19] to estimate the skeletal muscle genes directly responsible for changes in Si, accounting for exercise intensity and amount, as well as the direct effects of exercise (Fig. 1, step B). The gene list generated from these analyses—constrained by the transcription factors identified in the first analysis and the causal network linking skeletal muscle gene expression to Si—was used to create a gene expression network through annotated pathway analysis software (Fig. 1, step C).

Sex-specific transcription factor targets of exercise amount and intensity (Fig. 2). While it is known that the expression of genes regulating insulin sensitivity differs between men and women [6], the sex-specific effects of exercise intensity and amount have not been fully explored. To address this gap, we analyzed transcription factor gene expression by sex. Log2 difference scores (fold-change) were calculated for each gene from pre- and post-exercise timepoints. A two-way analysis of variance (ANOVA) identified probes significantly influenced ($P < 0.05$) by the exercise protocol, the participant's sex, or the interaction of these variables. We then performed a multidimensional scaling (cluster analysis) to visualize the pre- to post-exercise changes in transcription factors by biological sex.

The list of significantly modified genes was submitted to PASTAA (Predict Associated Transcription factors from Annotated Affinities), an online tool that identifies transcription factors regulating differentially expressed genes [20, 21]. A *P*-value of < 0.05 was used as the threshold for association. To calculate an "activity score" for each identified transcription factor, we multiplied the expression value of each significantly expressed gene at pre- and post-intervention timepoints by the affinity score of the associated transcription factor. These activity scores were then categorized using a gamma distribution-based expectation-maximization clustering algorithm [22] implemented in MatLab (Mathworks, Nat-

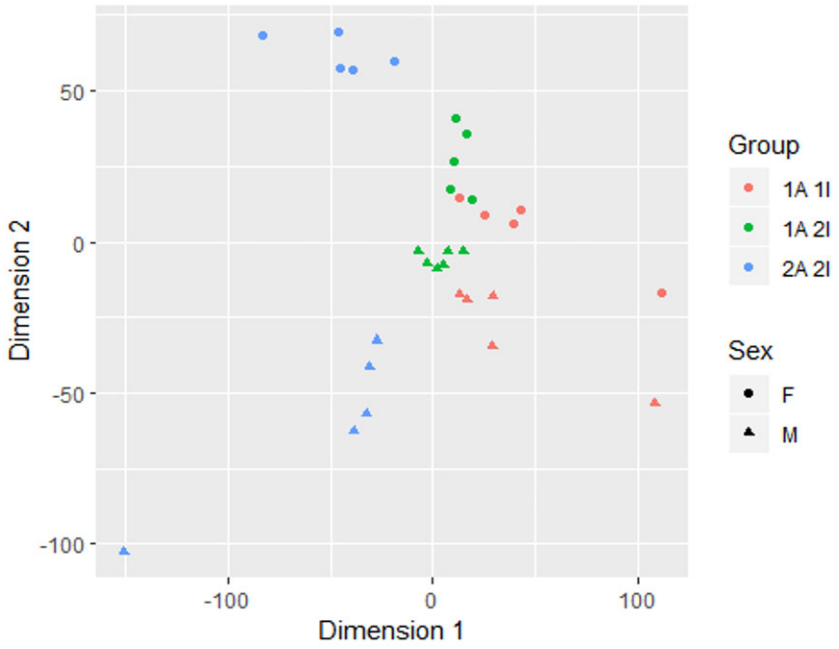


Figure 2. Multidimensional scaling (cluster analysis) of pre- to post-exercise change of transcriptional factors with significant variation according to participant sex (M, W) and/or exercise parameter (amount or intensity). The three colors—red, green, and blue—are assigned to the different exercise groups, where "1A" and "2A" represent different amounts controlling for intensity, and "1I" and "2I" represent different intensities controlling for amount. The women are represented by filled circles and the men by filled triangles. Each point represents one participant. Overall transcriptomic responses in subjects correlated with both exercise protocol and biological sex. Dimension 1 separates by exercise groups characterized by intensity and amount; Dimension 2 separates by biological sex.

ick, MA). Clinical measurements of blood glucose, insulin, and Si at the pre- and post-exercise timepoints were similarly categorized. The median values for each variable within a sex group at both timepoints were used to define response trajectories by biological sex. Additional details are provided in Supplementary Material S1.

Simulating regulatory pathway dynamics

The role of biological sex as an independent variable influencing exercise responses is not well understood, so we created separate sex-specific models. The regulatory logic of each model was adjusted independently to match the data from male and female participants, aiming to minimize hypothetical connections and use the simplest regulatory logic while maintaining a close fit to the input data, with <5% departure (Manhattan distance) between the reference input and predicted output trajectories. Since only two time-separated measurements were available, the number of intermediate discrete state transitions best representing network evolution over the eight-month period was not predetermined (see Supplementary Methods).

Data-driven identification of candidate genes mediating exercise amount and intensity effects on insulin sensitivity (step B)

To identify candidate direct causal genes responsive to exercise and those mediating exercise-induced changes in Si, we applied methods to identify multiple Markov boundaries [23] with exercise-induced changes in gene expression and Si as the target variables (see Supplementary Methods). Assuming causal sufficiency, variables present in all identified Markov boundaries were deemed true direct causal factors (true direct causes of Si or true direct effects of exercise on muscle gene expression). Variables identified in some—but not all—Markov boundaries may or may not have been causal, but the complete set included all true direct causal factors among the measured variables. The predictive performance of the linear regression models built with these Markov boundary variables for Si change was assessed via cross-validation.

Pathway analysis (step C)

Genes identified through transcription factor analysis (exercise-responsive genes, Fig. 1) and causal gene analysis (exercise-responsive mediators of Si, Fig. 1) were further explored for enrichment in key biological pathways using Ingenuity Pathway Analysis software (Qiagen; Winter 2021 release). The genes in each of the three sets—transcription factors (Table 2), causal Si genes (Table 3), and causal exercise-responsive genes (Table 4)—were linked using the Ingenuity Pathway Analysis (IPA) Knowledge database. Since the number of candidate genes was small due to a conservative identification strategy, we expanded each set by adding one additional linking node (a gene connected to two or more elements of the gene set) to establish the molecular networks involved. The expanded gene sets were displayed as a Venn diagram illustrating the overlap among subsets. The intersection of all three sets were developed into a pathway analysis.

Results

Results are presented according to the analytic plan presented in Fig. 1.

Table 2. Summary of Fig. 2: Regulatory relationships related to exercise amount or intensity consistent in men and women

Source	Target	Polarity
Amount	ATF1	Inhibiting
Amount	CEBPA	Inhibiting
Intensity	BACH2	Activating

Table 3. Markov boundaries for direct effects of skeletal muscle transcripts on changes in insulin sensitivity, numbers in the table shows the coefficients in the linear regression models with insulin sensitivity change as the dependent variable. By nature of the model, exercise intensity, amount and biological sex are internally controlled

Probe ID	Gene symbol	Model 1	Model 2	Model 3	Model 4
205 590_at	RASGRP1	− 0.688	− 0.839	− 0.647	− 0.749
213 442_x_at	SPDEF	− 2.054			
236 423_at	–	1.488	0.869	1.668	1.097
239 427_at	SLAMF1	− 0.719			
202 094_at	BIRC5		− 0.770		− 0.817
234 329_at	CLIC5		− 2.537		
201 268_at	NME1/NME2			2.969	
214 339_s_at	MAP4K1			− 2.967	
243 897_at	–				− 1.482

Transcription factor identification

We used PASTAA to predict transcription factors that regulate a set of genes based on annotated affinities calculated from biophysical interactions. Among the muscle genes that varied significantly by sex and/or exercise protocol parameters (intensity or amount), the PASTAA query identified 30 transcription factors predicted to be associated with the selected genes ($P < 0.05$).

Sex-specificity of exercise amount and intensity transcription factor targets

To examine the sex-specific effects of exercise amount and intensity, we stratified the analyses by sex and controlled for sex in subsequent analyses. A total of 6078 probes showed significant variation in expression across groups of men and women (uncorrected $P < 0.05$). The cluster patterns resulting from multidimensional scaling of these probes indicated that exercise intensity, amount, and biological sex all significantly influenced the transcriptomic response (Figs 2 and 3). The greatest divergence in co-expression profiles between men and women occurred where the effects of exercise amount and intensity were most pronounced.

Network structure for transcription factor regulation of insulin sensitivity

A query of the Pathway Studio database for documented regulatory interactions among the 30 identified transcription factors and three clinical measures (Si, glucose, and insulin) yielded 58 network edges, supported by 367 peer-reviewed references (minimum 2, median 3.5 per edge) (Fig. 4). We then integrated regulatory actions from exercise amount and intensity into the molecular network. Both amount and intensity were potentially connected to any of the 30 transcription factors in the network, with undetermined effects (either positive or negative polarity). Amount and intensity were represented as ternary state nodes, reflecting the levels applied in the study (absent = 0, low = 1, or high = 2).

Candidate direct causes of Si change (step C)

We identified four Markov boundaries (sets of direct-cause candidates for Si change), with each set containing four genes. RASGRP1 and the Affymetrix probe-set 236423_at (originally annotated to refseq NM_20 337, now mapping to ENSG00000276900, a novel transcript antisense to AMN1) were present in all four Markov boundaries, indicating they were estimated to be direct causes of exercise-induced changes in Si. Since these two genes were identified at pre-training, this suggests a predisposition for these genes to influence training responses, possibly through genetic effects. Other variables appeared in some of the Markov boundaries and were estimated to be putative local causal factors. The cross-validation R-squared value was 0.14 ± 0.13, indicating the predictive performance of the linear regression models built with these Markov boundary variables for Si change.

Table 3 lists the variables in the four Markov boundaries and the coefficients of the linear regression models using these variables as independent predictors of Si change (post- minus pre-exercise). The coefficients indicate the expected change in Si for a one-unit change in the variable, holding the other variables in the model constant. In all models, RASGRP1 was negatively related to Si change, while probe-set 236423_at (potentially an AMN1 antisense) was positively related to Si change.

Direct effects of exercise amount and intensity on muscle gene expression (step B)

We identified one Markov boundary for intensity (Table 4), indicating that the genes within this boundary are direct effects of exercise intensity. For exercise amount, we identified two Markov boundaries. The Affymetrix probe-set 1557214_at (originally annotated as Hs.380602 and now as long intergenic noncoding RNA, LINC02382) was present in both Markov boundaries, suggesting it is likely a direct effect of exercise amount. GRIN3B and EIF3B each appeared in one of the two Markov boundaries, making them candidates for direct effects of exercise amount.

Reconciling literature-based pathway discovery and data-driven causal discovery of concordant transcription targets

In an independent analysis guided by literature-based regulatory logic models, we found that genes associated with significant changes in insulin sensitivity were regulated by the 24 transcription factors targeted by exercise amount or intensity in either men or women identified above (Fig. 4). This model identifies regulatory circuits among transcription factor genes, it does not control for biological sex, or exercise characteristics unless so described in the literature. The literature is heterogeneous in how it controls for these characteristics. To make maximum use of known relationships, we included the union of regulatory transcription factors in both men and women as identified in Fig. 3. Among these, three transcription factors (E2F1, EGR1, and YY1) were also regulators of candidate genes directly causing changes in insulin sensitivity, but not CEBPA. Furthermore, 12 of these transcription factors were predicted to be direct targets of exercise amount and/or intensity in either the men- or women-specific regulatory models (Fig. 4).

Transcription factors such as BHLHE40, E2F1, EGR1, NKX2-5, PDX1, SP1, SP3, SREBF1, USF1, YY1, and ZEB1 were not direct mediators of exercise effects on insulin sen-

Table 4. Expression of Markov boundary members in different exercise intensity (4A) and amount (4B) using the entire harmonized dataset

(A) Intensity		Control		Moderate		Vigorous			
Probe ID	Gene symbol	*n*	Mean (sd)	*n*	Mean (sd)	*n*	Mean (sd)	*F*	*P*
208183_at	TACR3	16	0.72 ± 0.82	10	0.36 ± 0.55	23	−0.28 ± 0.59	11.19	0.0001
1561030_at	TMC7	16	−2.72 ± 4.12	10	0.81 ± 3.01	23	1.99 ± 4.87	5.72	0.006
1556422_at	–	16	0.09 ± 0.33	10	0.27 ± 0.43	23	−0.29 ± 0.42	8.5	0.0007
(B) Amount		Control		Low		High			
1557214_at	–	16	−0.21 ± 0.67	23	0.06 ± 0.61	10	0.84 ± 0.54	9.13	0.0005
233892_at	GRIN3B	16	0.22 ± 0.33	23	0.03 ± 0.40	10	0.25 ± 0.25	5.67	0.0062
237457_at	EIF3B	16	−2.41 ± 4.07	23	−0.18 ± 3.71	10	2.44 ± 3.79	4.94	0.0114

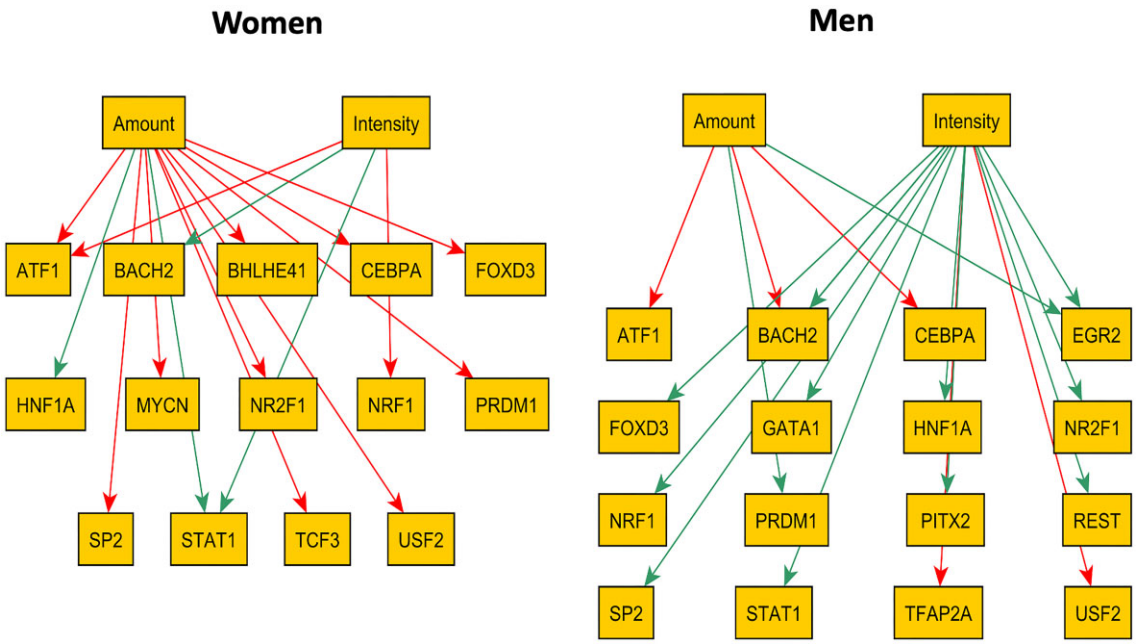


Figure 3. Exercise amount and intensity transcription factor gene targets; predicted direct targets of exercise amount and intensity in models for women (left panel) and men (right panel). Green edges are activating, and red edges are repressing. In women, exercise amount had substantially more targets (12) than intensity (4). In men, this pattern was reversed (5 targets for amount and 13 for intensity). The predicted effects on these targets differed as well: putative edges predominantly had a negative (inhibitory) polarity in the woman model and were concentrated on amount (10 out of 12 for amount, 2 out of 4 for intensity); in the man model putative edges predominantly had a positive (activating) polarity and were concentrated on intensity (2 out of 5 for amount, 11 out of 13 for intensity).

sitivity in the circuit model and were therefore excluded from the first-order (direct) effects models for both men and women. The only direct regulator of insulin sensitivity among the identified exercise transcriptional targets was CEBPA, which was predicted to be downregulated by exercise amount in both men and women. For the first-order effects models, all other regulators of insulin sensitivity were removed from the set of potential exercise targets. Table 2 lists the transcription factors commonly regulated by both sexes, even if the direction of regulation by exercise training was opposite.

Enriched biological pathways of exercise-responsive and insulin sensitivity-related target genes

To elucidate relationships among the genes identified using transcription factor-centric and causal modeling approaches, we explored biological pathways that include the transcription factors and their downstream targets revealed in the three previous analytic approaches. We expanded each gene set from those directly identified to include an additional element of connection—for example, the intensity-causal genes PLAUR and GRAMD4 are connected via interleukin-6 (IL-6). It is important to note that these additional elements may not be single genes, instead representing elements that can interact with one or more genes. As an example, an added element may be a drug or chemical known to regulate the original genes in the set.

The expanded gene sets were used to generate the Venn diagram in Fig. 5, which lists the 13 genes represented in all three expanded sets. These genes include CD3, CREBBP, EGFR, ESR1, ID2, MYC, RARA, TGFB1, TNF, TP53, VEGF, β-estradiol, and tetradecanoylphorbol acetate (mapped in biological space in Fig. 6). The latter two of these elements in the Ingenuity database represent elements with strong predictions as being upstream regulators of the other genes. While β-estradiol itself is strongly related to exercise responses, due to its wide downstream effects, the inclusion of tetrade-

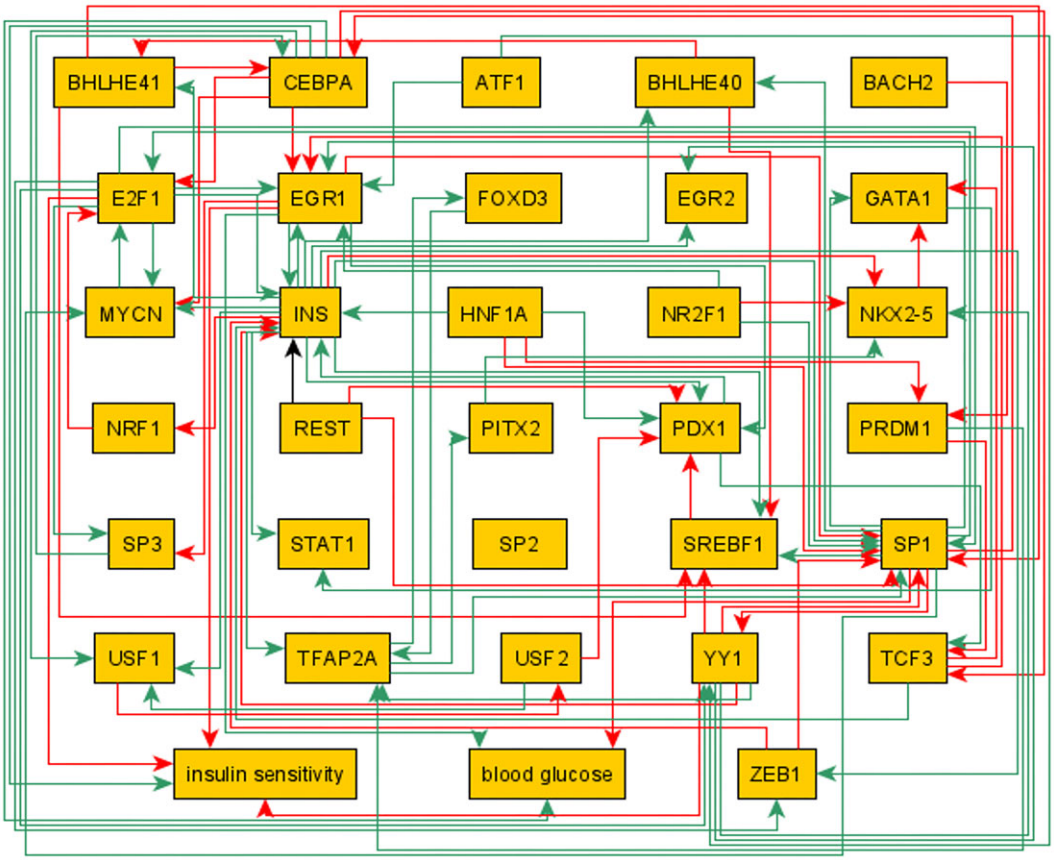


Figure 4. Regulatory circuit model. A transcription factor network linking 30 transcription factors through 58 documented regulatory interactions (edges) extracted from 367 full-text, peer-reviewed journal publications. Fasting insulin, fasting glucose, and Si were modeled as outcomes. In this regulatory circuit diagram, green edges represent positive regulation of connecting two parameters, with the arrow pointing in the direction of the causal relationship. Red arrows are similar but indicate negative regulation. As this is a model based upon the literature, this analysis modeling regulatory circuits among transcription factor genes, it does not control for biological sex, or exercise characteristics unless so described in the literature.

canoylphorbol acetate was interesting. As a potential upstream regulator of the 12 other elements of interest, tetradecanoylphorbol acetate (which is labeled a chemical/drug) is predicted to influence 10 (CREBBP, EGFR, ESR1, ID2, MYC, RARA, TGFB1, TNF, TP53, and VEGFA; *P*-value: 9.99E-14). We noticed a significant overlap between this list and that for protein kinase C (PKC) (labeled as gene group) signaling (EGFR, ID2, RARA, TGFB1, TNF, and VEGFA; *P*-value: 9.45E-11). Given the connection of PKC signaling to insulin sensitivity and exercise, we chose to further discuss the implications of PKC involvement. PKC and estradiol-directed interconnection was evident by the number of connections they had to other factors, with a notable common connection to the epidermal growth factor receptor (EGFR).

Discussion

Using an innovative and newly developed analytic approach, we integrated physiological and organ-level molecular data from a prospective, randomized clinical study of exercise training with literature-driven pathway and causal modeling to create a novel and significant model. First, we identified sex-specific transcriptional pathways, responsive to exercise intensity and amount, that are causally related to exercise-induced improvements in insulin sensitivity. This model marks a significant advancement over our previous work with this dataset [12]. Second, when considering the cumulative effects across exercise conditions, we identified two major interacting pathways: one directed by PKC signaling and the other by estrogen receptor signaling. Notably, these pathways, along with two others, converged on EGFR through direct connections, highlighting EGFR as a potential pharmacologic target for enhancing Si. Additionally, other downstream factors in these pathways were identified (Fig. 6).

Sexual dimorphism in the transcriptional mediators of exercise-induced insulin sensitivity (Si) responses

While physiological differences in exercise responsiveness between men and women have been recognized for some time, the significant influence of sex on our transcription factor regulatory models was unexpected. As previously reported, both exercise amount and intensity are associated with variations in training efficacy [24]. In this study, transcriptomic analysis of muscle biopsies collected before and after an eight-month exercise training program allowed us to identify key transcription factors linked to changes in transcriptional regulation. The predicted activity of these transcription factors at pre- and post-exercise timepoints constrained logic models designed to identify the minimal set of targets directly affected by exercise

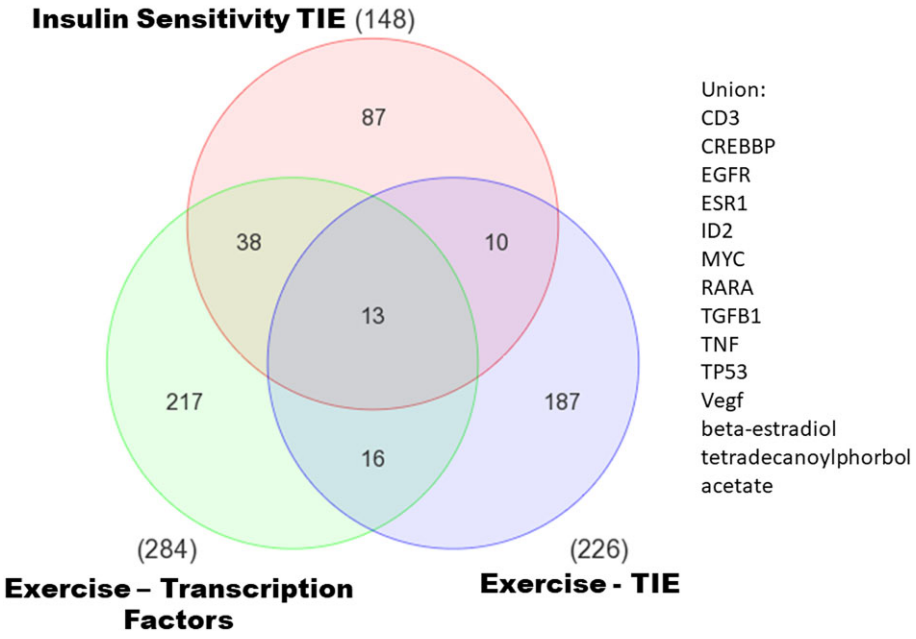


Figure 5. Relationships among exercise- and insulin sensitivity-related gene sets. Gene transcriptome candidates derived from Fig. 1, steps A, B, and C are represented as discs in as Exercise-Transcription Factors, Exercise-TIE, and Insulin Sensitivity TIE, respectively. Thirteen transcripts or those involved in a synthesis pathway (i.e. for β-estradiol or tetradecanoylphorbol acetate) in common were identified. The designations are listed to the right. "TIE" refers to the methods uses to derive the gene sets for insulin sensitivity-related and exercise-related genes as illustrated in Fig. 1. TIE is an acronym for Target Information Equivalence; it is a family of algorithms for identifying all Markov boundaries of a variable of interest in a distribution [50]. See Supplementary Methods.

amount and intensity. Separate models for men and women yielded distinct sets of predicted transcriptional targets, suggesting that the adaptive and/or refractory effects of exercise exhibit striking sexual dimorphism.

Estrogen and PKC signaling

The role of estrogen in exercise-induced health improvements is well established. Estrogen replacement is considered crucial for aerobic exercise-induced enhancements in vascular endothelial function in postmenopausal women [25–27]. Additionally, exercise-induced improvements in insulin sensitivity are linked to increased skeletal muscle capillarity, with sex-specific differences observed in postmenopausal women compared to age-matched men [28]. Estrogen's role as a major driver of exercise-induced improvements in insulin sensitivity aligns with our previous findings, where we observed an interaction between estrogen supplementation in postmenopausal women and exercise intensity on insulin sensitivity changes [28]. In men, the differential effects of exercise intensity on insulin sensitivity were less pronounced than in women. This study confirms those findings and provides a potential mechanistic explanation.

In the current study, PKC signaling was predicted to be an upstream regulator of insulin sensitivity determined 16–24 h after the final exercise training bout. In support of a PKC/exercise training/insulin action relationship, atypical PKC (aPKC) activity was elevated at rest in endurance-trained vs sedentary individuals [29]. aPKC is involved in insulin-mediated GLUT4 translocation [30] In both lean insulin-sensitive and obese insulin-resistant Zucker rats, aPKC isoform ζ mRNA and protein content increases with endurance training and are positively related to glucose tolerance [31]. A potential pharmacologic target was identified through phorbol esters, which act as analogs for diacylglycerol (DAG) and can activate both the conventional and novel PKC isoforms linked with the repression of insulin sensitivity [32–35]. Recently, we reported that pharmacological PKC activation strongly induces a transcriptional signature of insulin resistance *in vitro* [36]. Also, PKC-α expression mimicks insulin resistance *in vitro*, while overexpression has the opposite effect. Furthermore, PKC-δ activates mTORC1, and the PKC-δ inhibitor Ruboxistaurin reverses insulin resistance in preclinical models [37, 38]. Thus, the key gene regulatory mechanisms identified in our present modeling are strongly supported by these mechanistic studies. It is important to note that our mRNA data was not transcript-specific and PKC elements themselves were not identified as being changed transcriptionally; however, downstream transcripts influenced by PKC were identified.

Identification of drug target candidates for sex- and intensity-specific effects of exercise training on insulin sensitivity

A major goal of our work is to identify molecular targets for pharmacologic agents that could mediate the health benefits of exercise. The network model presented in this study, leading to Fig. 6, integrates longitudinal human clinical data, molecular data from skeletal muscle, and a novel analytic strategy—using causal modeling in a randomized prospective study—to identify potential targets. We noted that both phorbol ester and estrogen signaling converge on the EGFR as a potential target. In fact, using a distinct modeling approach with similar objectives and different data, we also identified EGFR as a potential mediator of muscle insulin

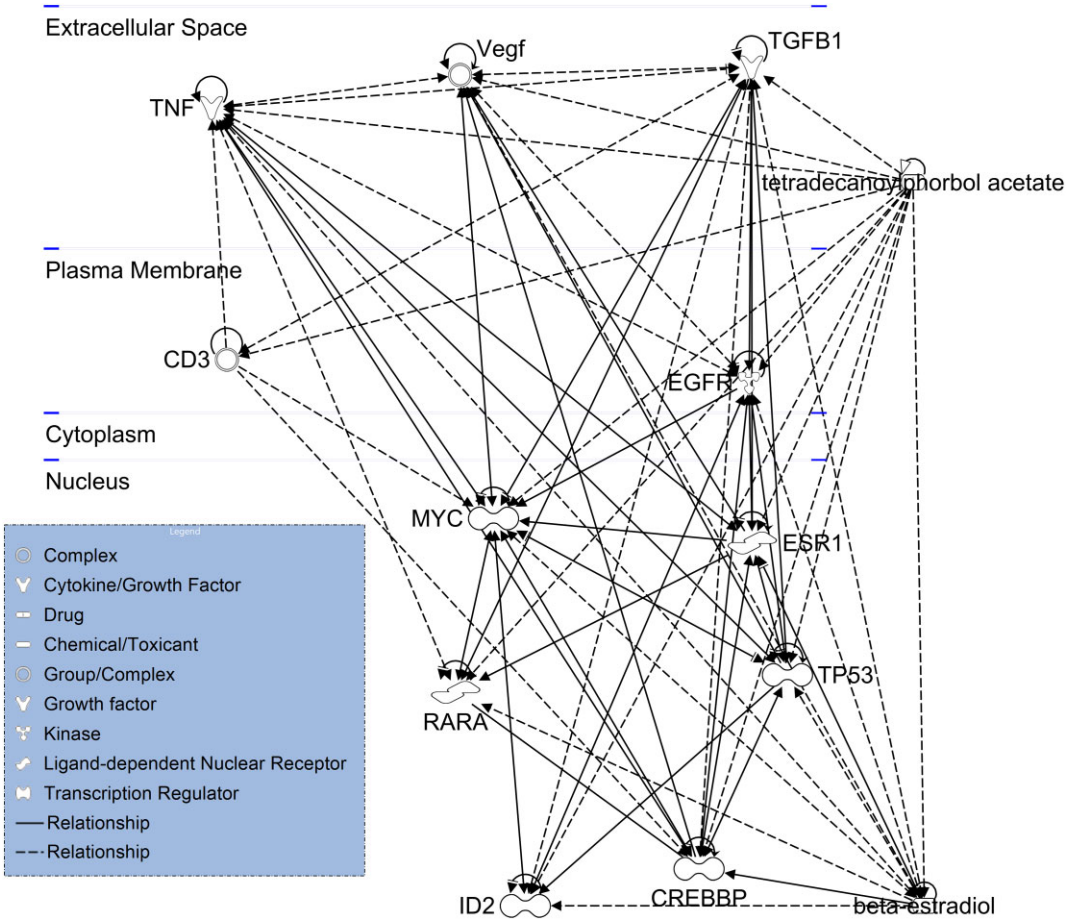


Figure 6. Biological relationships among exercise- and insulin sensitivity-related genes and factors identified in the analysis presented in Fig. 5. In the diagram, items at the top are localized to the extracellular space, followed by the plasma membrane, cytoplasm, and nucleus (transcription factors) as one moves down the diagram and as indicated on the left. Solid arrows indicate direct molecular interactions, and dotted arrows indirect effects. Master control nodes related to phorbol ester (PKC) and estradiol signaling are seen to the far right to denote that they are not isolated to a particular subcellular compartment. β-Estradiol, although a circulating hormone, has direct gene regulatory functions through the estrogen receptor, which are necessary for maintaining skeletal muscle fitness.

sensitivity. By integrating signatures for fasting and exercise-responsive insulin action with a drug repurposing database, we identified numerous EGFR tyrosine kinase inhibitors as pharmacologic agents that mimic insulin-related pathways in skeletal muscle *in vitro* [36]. This association was later supported by extensive analyses of preclinical genetic models [39], and more recently, EGFR has been recognized as a therapeutic link between insulin resistance and hypertrophic cardiomyopathy [40]. EGFR, a receptor tyrosine kinase, undergoes ligand-binding-mediated dimerization similar to insulin receptors [41]. Amphiregulin, the endogenous EGFR ligand, represents a direct link between obesity and inflammation [42]. Therapeutically, inhibiting EGFR enhances autophagy and insulin action [43–45]. However, using EGFR kinase domain inhibitors to treat metabolic diseases may be challenging due to the similarity of kinase domains across the kinase proteome. Therefore, alternative protein targets upstream or downstream of EGFR that regulate its activation may represent more optimal drug targets [36]. Furthermore, the findings from our present analysis, which account for sex-specific responses and exercise exposure characteristics (intensity and amount), provide a framework for understanding exercise dose-response relationships at a molecular level.

Implications of this work

The translational goal of these studies was to identify putative transcription factor targets of exercise amount and intensity in skeletal muscles, with a focus on sex-specific responses and their relationship to systemic measures of insulin, blood glucose, and insulin sensitivity. The identification of these specific targets offers promising candidates for potential intervention strategies aimed at improving insulin sensitivity in humans. However, pinpointing master regulators like estrogen and PKC signaling is only the first step, as these cannot be used solely for therapies targeting Si.

For instance, while estrogen has shown beneficial effects on cardiometabolic risk factors, insulin sensitivity, and capillarity in women [25–28], clinical trials of estrogen supplementation in postmenopausal women did not prevent cardiovascular events or reduce mortality. In fact, some randomized trials have reported an increased risk of cerebrovascular events (stroke) with estrogen use [46–48]. These findings highlight the potential off-target effects of systemic drug therapies intended for organ-specific outcomes, suggesting that a more practical and effective approach may involve striving for tissue-targeted effects.

Similarly, the systemic administration of PKC-signaling activators or inhibitors to improve insulin sensitivity could lead to unintended consequences. Drug targets should be more proximal, molecular, and tissue-specific, focusing directly on the site of the desired effect—in this case, Si. EGFR emerged as one such target. As identified by Timmons *et al.* [36], existing drugs can modify canonical signaling pathways involving PKC, EGFR, and mTOR, some of which could be repurposed to enhance insulin sensitivity in muscle or amplify the benefits of exercise. The regulatory networks identified in this study should be further explored to uncover other pathways amenable to pharmacologic targeting. Achieving this will require studies with larger sample sizes and broader molecular measures, such as epigenetics and proteomics, as well as investigating signaling in other organ systems like adipose tissue [49], liver, and pancreas [10, 50, 51], to fully assess the pleiotropic health effects of regular exercise.

Conclusions

By integrating physiological and organ-level molecular data with literature-driven pathway and causal modeling, this study identified potential molecular targets for developing pharmacologic agents that mimic the health effects of exercise training on insulin sensitivity. Aerobic exercise-induced signaling pathways that mediate Si vary by sex and are influenced by exercise intensity and amount. The analyses provide evidence that transcriptional adaptations in skeletal muscle related to insulin sensitivity improvements are causally linked to estrogen and PKC signaling, while also accounting for sex- and exposure-related differences in exercise. Future work should validate these findings in other similar datasets and test the activities of these pathways in preclinical and cell culture models. By focusing on additional health benefits of regular exercise training, such as improvements in cardiorespiratory fitness, lipid metabolism, pancreatic function, and body composition, future studies using similar methods and larger sample sizes are likely to expand these findings and identify additional molecular targets. This work will critically inform the development of new therapies for the numerous health conditions that exercise effectively addresses.

Acknowledgements

We would like to acknowledge Eric P. Hoffman with whom both W.E.K. and M.J.H. worked to develop the STRRIDE gene expression work. We also thank Milton E. Campbell and Melissa Hurdle for technical assistance with sample handling. Graphical Abstract created in BioRender. Kraus, W. (2025) https://BioRender.com/z64j080.

Author contributions: conceptualization and funding (S.M., M.J.H., G.B., C.F.A. and W.E.K); methodology and investigation (S.M., M.J.H., M.C.M., J.A.H., J.L.J., G.B., C.F.A. and W.E.K); data curation (S.M., M.J.H., M.C.M., L.M.R., K.M.H., E.R.H., A.B., R.J., J.A.H., W.B.B., J.L.J., J.A.H., B.G., and W.E.K); original draft (W.E.K.); review and editing (S.M., M.J.H., M.C.M., L.M.R., K.M.H., M.D.B., J.A.H., J.A.T., G.B., V.B.K., C.F.A., and W.E.K.).

Supplementary data

Supplementary data is available at NAR Molecular Medicine online.

Conflict of interest

Constantin Aliferis is co-inventor in several patents on machine learning modeling methods but receives no income from them. The remaining authors declare no competing interests. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

Funding

The study was supported by funding from NIH/NNLBI R01 HL153497 (to W.E.K., S.M., C.F.A., G.B., and M.J.H.), R01 HL057354 (to W.E.K. and J.A.H.), and NIH/NCATS UL1TR002494 (to S.M. and C.F.A.). L.M.R. is supported by 23CDA1051777.

Data availability

Data supporting this analysis are available at the following GEO registrations: U133 Plus Series, Accession: GSE47969, ID: 200047969; Illumina Series, Accession: GSE83352, ID: 200083352.

References

1. Powell KE, King AC, Buchner DM *et al.* The scientific foundation for the Physical Activity Guidelines for Americans, 2nd Edition. *J Phys Act Health* 2018;**2018**:1–11. https://doi.org/10.1123/jpah.2018-0618
2. Williams RS, Kraus WE. Exercise and health: can biotechnology confer similar benefits? *PLoS Med* 2005; **2**:e68. https://doi.org/10.1371/journal.pmed.0020068
3. Jensen TE, Richter EA. Regulation of glucose and glycogen metabolism during and after exercise. *J Physiol* 2012; **590**:1069–76. https://doi.org/10.1113/jphysiol.2011.224972
4. Wojtaszewski JF, Richter EA. Effects of acute exercise and training on insulin action and sensitivity: focus on molecular mechanisms in muscle. *Essays Biochem* 2006; **42**:31–46.
5. Kraus WE, Slentz CA. Exercise training, lipid regulation, and insulin action: a tangled web of cause and effect. *obes* 2009; **17(Suppl 3)**:S21–26. https://doi.org/10.1038/oby.2009.384
6. Timmons JA, Atherton PJ, Larsson O *et al.* A coding and non-coding transcriptomic perspective on the genomics of human metabolic disease. *Nucleic Acids Res* 2018; **46**:7772–92. https://doi.org/10.1093/nar/gky570
7. Ross LM, Slentz CA, Zidek AM *et al.* Effects of amount, intensity, and mode of exercise training on insulin resistance and type 2 diabetes risk in the STRRIDE randomized trials. *Front Physiol* 2021; **12**:626142. https://doi.org/10.3389/fphys.2021.626142
8. Kraus WE, Torgan CE, Duscha BD *et al.* Studies of a targeted risk reduction intervention through defined exercise (STRRIDE). *Med Sci Sports Exercise* 2001; **33**:1774–84. https://doi.org/10.1097/00005768-200110000-00025
9. Bateman LA, Slentz CA, Willis LH *et al.* Comparison of aerobic versus resistance exercise training effects on metabolic syndrome (from the Studies of a Targeted Risk Reduction Intervention Through Defined Exercise - STRRIDE-AT/RT). *Am J Cardiol* 2011; **108**:838–44. https://doi.org/10.1016/j.amjcard.2011.04.037
10. Slentz CA *et al.* Effects of aerobic vs. resistance training on visceral and liver fat stores, liver enzymes, and insulin resistance by HOMA in overweight adults from STRRIDE AT/RT. American journal of physiology. *Endocrin Metab* 2011; **301**:E1033–1039.
11. Slentz CA, Bateman LA, Willis LH *et al.* Effects of exercise training alone vs a combined exercise and nutritional lifestyle intervention on glucose homeostasis in prediabetic individuals: a randomised

controlled trial. *Diabetologia* 2016; **59**:2088–98. https://doi.org/10.1007/s00125-016-4051-z
12. Barberio MD, Huffman KM, Giri M *et al.* Pyruvate dehydrogenase phosphatase regulatory gene expression correlates with exercise training insulin sensitivity changes. *Med Sci Sports Exercise* 2016; **48**:2387–97. https://doi.org/10.1249/MSS.0000000000001041
13. Duscha BD, Slentz CA, Johnson JL *et al.* Effects of exercise training amount and intensity on peak oxygen consumption in middle-age men and women at risk for cardiovascular disease. *Chest* 2005; **128**:2788–93. https://doi.org/10.1378/chest.128.4.2788
14. Moreau K. Regular exercise, hormone replacement therapy and the age-related decline in carotid arterial compliance in healthy women. *Cardiovasc Res* 2003; **57**:861–8. https://doi.org/10.1016/S0008-6363(02)00777-0
15. Moreau KL, Stauffer BL, Kohrt WM *et al.* Essential role of estrogen for improvements in vascular endothelial function with endurance exercise in postmenopausal women. *J Clin Endocrinol Metab* 2013; **98**:4507–15. https://doi.org/10.1210/jc.2013-2183
16. Seals DR, Nagy EE, Moreau KL. Aerobic exercise training and vascular function with ageing in healthy men and women. *J Physiol* 2019; **597**:4901–14. https://doi.org/10.1113/JP277764
17. Huffman KM, Slentz CA, Johnson JL *et al.* Impact of hormone replacement therapy on exercise training-induced improvements in insulin action in sedentary overweight adults. *Metabolism* 2008; **57**:888–95. https://doi.org/10.1016/j.metabol.2008.01.034
18. Richter EA, Derave W, Wojtaszewski JF. Glucose, exercise and insulin: emerging concepts. *J Physiol* 2001; **535**:313–22. https://doi.org/10.1111/j.1469-7793.2001.t01-2-00313.x
19. Griffin ME, Marcucci MJ, Cline GW *et al.* Free fatty acid-induced insulin resistance is associated with activation of protein kinase C theta and alterations in the insulin signaling cascade. *Diabetes* 1999; **48**:1270–4. https://doi.org/10.2337/diabetes.48.6.1270
20. Itani SI, Ruderman NB, Schmieder F *et al.* Lipid-induced insulin resistance in human muscle is associated with changes in diacylglycerol, protein kinase C, and IkappaB-alpha. *Diabetes* 2002; **51**:2005–11. https://doi.org/10.2337/diabetes.51.7.2005
21. Li M, Vienberg SG, Bezy O *et al.* Role of pkcδ in insulin sensitivity and skeletal muscle metabolism. *Diabetes* 2015; **64**:4023–32. https://doi.org/10.2337/db14-1891
22. Schmitz-Peiffer C. Protein kinase C and lipid-induced insulin resistance in skeletal muscle. *Ann NY Acad Sci* 2002; **967**:146–57. https://doi.org/10.1111/j.1749-6632.2002.tb04272.x
23. Roden M, Shulman GI. The integrative biology of type 2 diabetes. *Nature* 2019; **576**:51–60. https://doi.org/10.1038/s41586-019-1797-8
24. Timmons JA, Anighoro A, Brogan RJ *et al.* A human-based multi-gene signature enables quantitative drug repurposing for metabolic disease. *eLife* 2022; **11**:e68832. https://doi.org/10.7554/eLife.68832
25. Guo W-H, Wang X, Shang M-S *et al.* Crosstalk between PKC and MAPK pathway activation in cardiac fibroblasts in a rat model of atrial fibrillation. *Biotechnol Lett* 2020; **42**:1219–27. https://doi.org/10.1007/s10529-020-02843-y
26. Naruse K, Rask-Madsen C, Takahara N *et al.* Activation of vascular protein kinase C-beta inhibits akt-dependent endothelial nitric oxide synthase function in obesity-associated insulin resistance. *Diabetes* 2006; **55**:691–8. https://doi.org/10.2337/diabetes.55.03.06.db05-0771
27. Masson SW, Madsen S, Cooke KC *et al.* Leveraging genetic diversity to identify small molecules that reverse mouse skeletal muscle insulin resistance. *eLife* 2023; **12**:RP86961. https://doi.org/10.7554/eLife.86961.3
28. Algül S, Schuldt M, Manders E *et al.* EGFR/IGF1R signaling modulates relaxation in hypertrophic cardiomyopathy. *Circ Res* 2023; **133**:387–99. https://doi.org/10.1161/CIRCRESAHA.122.322133
29. Ferguson KM, Hu C, Lemmon MA. Insulin and epidermal growth factor receptor family members share parallel activation mechanisms. *Protein Sci* 2020; **29**:1331–44. https://doi.org/10.1002/pro.3871
30. Skurski J, Penniman CM, Geesala R *et al.* Loss of iRhom2 accelerates fat gain and insulin resistance in diet-induced obesity despite reduced adipose tissue inflammation. *Metabolism* 2020; **106**:154194. https://doi.org/10.1016/j.metabol.2020.154194
31. Bowman CJ, Ayer DE, Dynlacht BD. Foxk proteins repress the initiation of starvation-induced atrophy and autophagy programs. *Nat Cell Biol* 2014; **16**:1202–14. https://doi.org/10.1038/ncb3062
32. Jahng JWS, Alsaadi RM, Palanivel R *et al.* Iron overload inhibits late stage autophagic flux leading to insulin resistance. *EMBO Rep* 2019; **20**:e47911. https://doi.org/10.15252/embr.201947911
33. Klaeger S, Heinzlmeir S, Wilhelm M *et al.* The target landscape of clinical kinase drugs. *Science* 2017;**358**:eaan4368. https://doi.org/10.1126/science.aan4368
34. Brass LM. Hormone replacement therapy and stroke: clinical trials review. *Stroke* 2004; **35**:2644–7. https://doi.org/10.1161/01.STR.0000143218.20061.ac
35. Harman SM, Naftolin F, Brinton EA *et al.* Is the estrogen controversy over? Deconstructing the Women's Health Initiative study: a critical evaluation of the evidence. *Ann NY Acad Sci* 2005; **1052**:43–56. https://doi.org/10.1196/annals.1347.004
36. Simon JA, Hsia J, Cauley JA *et al.* Postmenopausal hormone therapy and risk of stroke: the heart and estrogen-progestin replacement study (HERS). *Circulation* 2001; **103**:638–42. https://doi.org/10.1161/01.CIR.103.5.638
37. Nakhuda A, Josse AR, Gburcik V *et al.* Biomarkers of browning of white adipose tissue and their regulation during exercise- and diet-induced weight loss. *Am J Clin Nutr* 2016; **104**:557–65. https://doi.org/10.3945/ajcn.116.132563
38. Slentz CA, Houmard JA, Kraus WE. Exercise, abdominal obesity, skeletal muscle, and metabolic risk: evidence for a dose response. *obes* 2009;**17**(**Suppl 3**),S27–33. https://doi.org/10.1038/oby.2009.385
39. Slentz CA, Tanner CJ, Bateman LA *et al.* Effects of exercise training intensity on pancreatic beta-cell function. *Diabetes Care* 2009; **32**:1807–11. https://doi.org/10.2337/dc09-0032
40. Houmard JA, Tanner CJ, Slentz CA *et al.* Effect of the volume and intensity of exercise training on insulin sensitivity. *J Appl Physiol* 2004; **96**:101–6. https://doi.org/10.1152/japplphysiol.00707.2003
41. Sedghamiz H, Morris M, Craddock TJA *et al.* Bio-ModelChecker: using bounded constraint satisfaction to seamlessly integrate observed behavior with prior knowledge of biological networks. *Front Bioeng Biotechnol* 2019;**7**:48. https://doi.org/10.3389/fbioe.2019.00048
42. Sedghamiz H, Morris M, Whitley D *et al.* Computation of robust minimal intervention sets in multi-valued biological regulatory networks. *Front Physiol* 2019; **10**:241. https://doi.org/10.3389/fphys.2019.00241
43. Vashishtha S, Broderick G, Craddock TJA *et al.* Inferring broad regulatory biology from time course data: have we reached an upper bound under constraints typical of in vivo studies? *PLoS One* 2015;**10**:e0127364. https://doi.org/10.1371/journal.pone.0127364
44. Nikitin A, Egorov S, Daraselia N *et al.* Pathway studio–the analysis and navigation of molecular networks. *Bioinformatics* 2003; **19**:2155–7. https://doi.org/10.1093/bioinformatics/btg290
45. Pearl J. *Causality: Models, Reasoning and Inference*. Vol. **29**, Cham, SZ: Springer, 2000.
46. Aliferis CF, Statnikov A, Tsamardinos I *et al.* Local causal and Markov blanket induction for causal discovery and feature selection for classification part I: algorithms and empirical evaluation. *J Mach Learn Res* 2010; **11**:171–234.
47. Roider HG, Kanhere A, Manke T *et al.* Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics* 2007; **23**:134–41. https://doi.org/10.1093/bioinformatics/btl565
48. Roider HG, Manke T, O'Keeffe S *et al.* PASTAA: identifying transcription factors associated with sets of co-regulated genes.

Bioinformatics 2009; **25**:435–42. https://doi.org/10.1093/bioinformatics/btn627
49. Efroni S, Schaefer CF, Buetow KH. Identification of key processes underlying cancer phenotypes using biologic pathway analysis. *PLoS One* 2007; **2**:e425. https://doi.org/10.1371/journal.pone.0000425
50. Statnikov A, Lytkin NI, Lemeire J *et al.* Algorithms for discovery of multiple Markov boundaries. *J Mach Learn Res* 2013; **14**:499−566.
51. Sopariwala D, Nguyen H, Narkar V. Estrogen-related receptor signaling in skeletal muscle fitness. *Int J Sports Med* 2023; **44**:609–17.

Received: October 20, 2024. **Revised:** March 7, 2025. **Editorial Decision:** March 10, 2025. **Accepted:** March 27, 2025