

## **Sex-Specific Skeletal Muscle Gene Expression Responses to Exercise Reveal Novel Direct Mediators of Insulin Sensitivity Change**

Ma S<sup>\*1</sup>, Morris MC<sup>\*2</sup>, Hubal MJ<sup>3</sup>, Ross LM<sup>4</sup>, Huffman KM<sup>4</sup>, Vann CG<sup>4</sup>, Moore N<sup>4</sup>, Hauser ER<sup>4,5</sup>, Bareja A<sup>4</sup>, Jiang R<sup>5</sup>, Kummerfeld E<sup>1</sup>, Barberio MD<sup>6</sup>, Houmard JA<sup>7</sup>, Bennett WB<sup>4</sup>, Johnson JL<sup>4</sup>, Timmons JA<sup>8</sup>, Broderick G<sup>2</sup>, Kraus VB<sup>4</sup>, Aliferis CF<sup>1</sup>, Kraus WE<sup>4</sup>

\* Equal contribution

<sup>1</sup>Institute for Health Informatics (IHI), Academic Health Center, University of Minnesota, Minneapolis, MN 55455

<sup>2</sup>Center for Clinical Systems Biology, Rochester General Hospital, Rochester, NY 14621

<sup>3</sup>Department of Kinesiology, Indiana University – Indianapolis, Indianapolis IN 46202

<sup>4</sup>Duke Molecular Physiology Institute, Duke University School of Medicine, Durham, NC 27701

<sup>5</sup>Department of Head and Neck Surgery & Communication Sciences, Duke University School of Medicine, Durham, NC 27701

<sup>6</sup>Department of Exercise and Nutrition Sciences, George Washington University, Washington DC 20052

<sup>7</sup>Department of Kinesiology, ECU, Greenville, NC 27858

<sup>8</sup>School of Medicine and Dentistry, Queen Mary University of London, UK

Corresponding author:

William E. Kraus

Duke Molecular Physiology Institute

300 N. Duke St.

Durham, NC 27701

[william.kraus@duke.edu](mailto:william.kraus@duke.edu)

919-681-6733

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

### **ABSTRACT**

#### **BACKGROUND**

Understanding the causal pathways, systems, and mechanisms through which exercise impacts human health is complex. This study explores molecular signaling related to whole-body insulin sensitivity (Si) by examining changes in skeletal muscle gene expression. The analysis considers differences by biological sex, exercise amount, and exercise intensity to identify potential molecular targets for developing pharmacologic agents that replicate the health benefits of exercise.

#### **METHODS**

The study involved 53 participants from the STRRIDE I and II trials who completed eight months of aerobic training. Skeletal muscle gene expression was measured using Affymetrix and Illumina technologies, while pre- and post-training Si was assessed via an intravenous glucose tolerance test. A novel gene discovery protocol, integrating three literature-derived and data-driven modeling strategies, was employed to identify causal pathways and direct causal factors based on differentially expressed transcripts associated with exercise intensity and amount.

#### **RESULTS**

In women, the transcription factor targets identified were primarily influenced by exercise amount and were generally inhibitory. In contrast, in men, these targets were driven by exercise intensity and were generally activating. Transcription factors such as ATF1, CEBPA, BACH2, and STAT1 were commonly activating in both sexes. Specific transcriptional targets related to exercise-induced Si improvements included TACR3 and TMC7 for intensity-driven effects, and GRIN3B and EIF3B for amount-driven effects. Two key signaling pathways mediating aerobic exercise-induced Si improvements were identified: one centered on estrogen signaling and the other on phorbol ester (PKC) signaling, both converging on the epidermal growth factor receptor (EGFR) and other relevant targets.

#### **CONCLUSIONS**

The signaling pathways mediating Si improvements from aerobic exercise differed by sex and were further distinguished by exercise intensity and amount. Transcriptional adaptations in skeletal muscle related to Si improvements appear to be causally linked to estrogen and PKC signaling, with EGFR and other identified targets emerging as potential skeletal muscle-specific drug targets to mimic the beneficial effects of exercise on Si.

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

### INTRODUCTION

Exercise training provides substantial health benefits; however, too few individuals adopt and maintain it as a lifelong health strategy<sup>1</sup>. Consequently, there is significant interest in developing pharmacologic alternatives able to replicate the health effects of exercise<sup>2</sup>. 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<sup>3-6</sup>. 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<sup>7</sup>.

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<sup>8-11</sup>. 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.

### RESULTS

**Transcription Factor Identification.** We used PASTAA to predict transcription factors that regulate a set of genes based on annotated affinities calculated from biophysical interactions.

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

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 ( $p < 0.05$ ) (Supplemental Figure 1).

**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 6,078 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 (Figures 2, 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) (Figure 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 (Figure 1; Step C).** We identified four Markov boundaries (sets of candidate direct causes for Si change), with each set containing four genes. RASGRP1 and the Affymetrix probe-set 236423\_at (originally annotated to refseq NM\_207337, 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  $0.14 \pm 0.13$ , 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.

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

**Direct Effects of Exercise Amount and Intensity on Muscle Gene Expression (Figure 1, 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 (Table 5). 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 24 transcription factors targeted by exercise amount or intensity in either men or women. 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 (Figure 3).

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 sensitivity 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 the regulatory pathways among the transcription factors and their downstream targets revealed in the three previous analytic approaches. The expanded gene sets were used to generate the Venn diagram in Figure 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, and VEGF. The beta-estradiol and protein kinase C (PKC) signaling networks were identified as key connectors among transcription factors and causal modeling of Si and exercise-responsive genes (mapped in biological space in Figure 6). The strength of the phorbol ester- and estradiol-directed nodes was evident by the number of connections they had to other factors, with a notable common connection to the epidermal growth factor receptor (EGFR).

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

### 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<sup>12</sup>. Second, when considering the cumulative effects across exercise conditions, we identified two major interacting pathways: one directed by phorbol ester and 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 (Figure 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<sup>13</sup>. 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 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.

**Interaction of Estrogen and Phorbol Ester 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<sup>14-16</sup>. 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<sup>17</sup>. 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<sup>17</sup>. 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.

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

Similarly, extensive research has established the role of phorbol ester (PKC) signaling in exercise-induced improvements in insulin sensitivity. Insulin signaling through PKC to AMP-kinase and mTOR is central to glucose transport in skeletal muscle<sup>18,19</sup>. Phorbol esters affect various cellular processes, including the repression of insulin sensitivity in skeletal muscle, potentially by acting as analogs of diacylglycerol and activating PKC regulatory pathways<sup>19-22</sup>. Multiple PKC isoforms are dysregulated in metabolic disease, with each subtype often having divergent roles<sup>23</sup>. Recently, we reported that pharmacological PKC activation strongly induced a transcriptional signature of insulin resistance in vitro<sup>24</sup>. We also observed that loss of PKC- $\alpha$  expression mimicked insulin resistance in vitro, while overexpression had the opposite effect. Furthermore, PKC- $\delta$  activates mTORC1, and the PKC- $\delta$  inhibitor Ruboxistaurin reverses insulin resistance in preclinical models<sup>25,26</sup>. Thus, the key gene regulatory mechanisms identified in our present modeling are strongly supported by these mechanistic studies.

**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 Figure 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 epithelial growth factor receptor (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 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<sup>24</sup>. This association was later supported by extensive analyses of preclinical genetic models<sup>27</sup>, and more recently, EGFR has been recognized as a therapeutic link between insulin resistance and hypertrophic cardiomyopathy<sup>28</sup>. EGFR, a receptor tyrosine kinase, undergoes ligand-binding-mediated dimerization similar to insulin receptors<sup>29</sup>. Amphiregulin, the endogenous EGFR ligand, represents a direct link between obesity and inflammation<sup>30</sup>. Therapeutically, inhibiting EGFR enhances autophagy and insulin action<sup>31-33</sup>. 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<sup>24</sup>. 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



## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

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 phorbol ester 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<sup>14-17</sup>, 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<sup>34-36</sup>. 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.<sup>24</sup>, 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<sup>37</sup>, liver, and pancreas<sup>10,38,39</sup>, 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. 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.

## METHODS



## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

**Study Cohort.** This analysis focused on the aerobic exercise training groups from the STRRIDE I and II studies (NCT00200993 and NCT00275145)<sup>8,9</sup>. 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 vice versa. 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 ( $\dot{V}O_2$ ), assessed through a maximal cardiopulmonary exercise test. Moderate intensity was set at 40-55% of peak  $\dot{V}O_2$ , and vigorous intensity at 65-80%<sup>8</sup>. Participants adhered to their assigned exercise protocol for eight months, with a median adherence rate of 91.0% (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 to 70 years, who had overweight or class I obesity (BMI 25-35 kg/m<sup>2</sup>), dyslipidemia, and metabolic syndrome, but without overt coronary artery disease or diabetes<sup>9</sup>. 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.

**Muscle Sampling and FSIVGTT.** Muscle biopsies were taken from the vastus lateralis before the initiation of the exercise training protocol (pre) and 16 to 24 hours after the last training session (post)<sup>12</sup>. Total muscle RNA was prepared as described and used for gene expression analyses<sup>12</sup>. 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<sup>40</sup>.

**Skeletal Muscle Genome-Wide Gene Expression.** As previously described<sup>12</sup>, 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.

**Overall Analytic Strategy to Identify Molecular Targets of Exercise-Induced Changes in Insulin Sensitivity.** The overall analytic strategy is illustrated in Figure 1. Initially, we used a previously described approach<sup>41-43</sup> 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

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

factor expression, resulting in a partially-directed causal network curated with existing knowledge from the literature using the Elsevier Pathway Studio database<sup>44</sup> (Figure 1, Step A). In parallel, we applied causal discovery Markov boundary induction methods<sup>45,46</sup> 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 (Figure 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 (Figure 1, Step C).

Sex-Specific Transcription Factor Targets of Exercise Amount and Intensity (Figure 2). While it is known that the expression of genes regulating insulin sensitivity differs between men and women<sup>6</sup>, 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<sup>47,48</sup>. 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<sup>49</sup> implemented in MatLab (Mathworks, Natick, 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 Supplement 1.

**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 less than a 5% departure (Manhattan distance) between the reference input and predicted output trajectories. Since only two time-separated measurements were available, the number of

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

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 (Figure 1, Step B).** To identify candidate direct causal genes responsive to exercise and those mediating exercise-induced changes in  $S_i$ , we applied methods to identify multiple Markov boundaries<sup>50</sup> with exercise-induced changes in gene expression and  $S_i$  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  $S_i$  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  $S_i$  change was assessed via cross-validation.

**Pathway Analysis (Figure 1, Step C).** Genes identified through transcription factor analysis (exercise-responsive genes, Figure 1) and causal gene analysis (exercise-responsive mediators of  $S_i$ , Figure 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  $S_i$  genes (Table 3), and causal exercise-responsive genes (Table 4)—were linked using the IPA Knowledge database. Since direct relationships among these genes were rare, we expanded each set by adding one additional linking node (a gene connected to two or more elements of the gene set) to enhance interconnectedness. The expanded gene sets are depicted in Figure 5 as a Venn diagram illustrating the overlap among subsets. The intersection of all three sets is shown in the pathway analysis (Figure 6).

**Acknowledgements.** This study is supported by HL153497. STRRIDE I and II were supported by R01HL57354. LMR is supported by 23CDA1051777. We would like to acknowledge Eric P. Hoffman with whom both WEK and MJH worked to develop the STRRIDE gene expression work.

**Conflicts of Interest.** The authors of this manuscript do not report any conflicts. The results of the present study do not constitute endorsement by ACSM. The results of the study are presented clearly, honestly, and without fabrication, falsification, or inappropriate data manipulation.

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

### REFERENCES

1. Powell, K. E. et al. The Scientific Foundation for the Physical Activity Guidelines for Americans, 2nd Edition. *J Phys Act Health*, 1-11 (2018).
2. Williams, R. S. & Kraus, W. E. Exercise and health: can biotechnology confer similar benefits? *PLoS medicine* 2, e68 (2005).
3. Jensen, T. E. & Richter, E. A. Regulation of glucose and glycogen metabolism during and after exercise. *J Physiol* 590, 1069-1076 (2012).
4. Wojtaszewski, J. F. & Richter, E. A. Effects of acute exercise and training on insulin action and sensitivity: focus on molecular mechanisms in muscle. *Essays Biochem* 42, 31-46 (2006).
5. Kraus, W. E. & Slentz, C. A. Exercise training, lipid regulation, and insulin action: a tangled web of cause and effect. *Obesity (Silver Spring)* 17 Suppl 3, S21-26 (2009).
6. Timmons, J. A. et al. A coding and non-coding transcriptomic perspective on the genomics of human metabolic disease. *Nucleic Acids Res* 46, 7772-7792 (2018).
7. Ross, L. M. 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* 12, 626142 (2021).
8. Kraus, W. E. et al. Studies of a targeted risk reduction intervention through defined exercise (STRRIDE). *Medicine and science in sports and exercise* 33, 1774-1784 (2001).
9. Bateman, L. A. 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). *The American journal of cardiology* 108, 838-844 (2011).
10. Slentz, C. A. 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. Endocrinology and metabolism* 301, E1033-1039 (2011).
11. Slentz, C. A. 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* 59, 2088-2098 (2016).
12. Barberio, M. D. et al. Pyruvate Dehydrogenase Phosphatase Regulatory Gene Expression Correlates with Exercise Training Insulin Sensitivity Changes. *Medicine and science in sports and exercise* 48, 2387-2397 (2016).
13. Duscha, B. D. 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* 128, 2788-2793 (2005).
14. Moreau, K. L., Donato, A. J., Seals, D. R., DeSouza, C. A. & Tanaka, H. Regular exercise, hormone replacement therapy and the age-related decline in carotid arterial compliance in healthy women. *Cardiovasc Res* 57, 861-868 (2003).

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

15. Moreau, K. L., Stauffer, B. L., Kohrt, W. M. & Seals, D. R. Essential role of estrogen for improvements in vascular endothelial function with endurance exercise in postmenopausal women. *J Clin Endocrinol Metab* 98, 4507-4515 (2013).
16. Seals, D. R., Nagy, E. E. & Moreau, K. L. Aerobic exercise training and vascular function with ageing in healthy men and women. *J Physiol* 597, 4901-4914 (2019).
17. Huffman, K. M. et al. Impact of hormone replacement therapy on exercise training-induced improvements in insulin action in sedentary overweight adults. *Metabolism: clinical and experimental* 57, 888-895 (2008).
18. Richter, E. A., Derave, W. & Wojtaszewski, J. F. Glucose, exercise and insulin: emerging concepts. *J Physiol* 535, 313-322 (2001).
19. Griffin, M. E. 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* 48, 1270-1274 (1999).
20. Itani, S. I., Ruderman, N. B., Schmieder, F. & Boden, G. Lipid-induced insulin resistance in human muscle is associated with changes in diacylglycerol, protein kinase C, and I $\kappa$ B $\alpha$ . *Diabetes* 51, 2005-2011 (2002).
21. Li, M., Vienberg, S. G., Bezy, O., O'Neill, B. T. & Kahn, C. R. Role of PKC $\delta$  in Insulin Sensitivity and Skeletal Muscle Metabolism. *Diabetes* 64, 4023-4032 (2015).
22. Schmitz-Peiffer, C. Protein kinase C and lipid-induced insulin resistance in skeletal muscle. *Ann N Y Acad Sci* 967, 146-157 (2002).
23. Roden, M. & Shulman, G. I. The integrative biology of type 2 diabetes. *Nature* 576, 51-60 (2019).
24. Timmons, J. A. et al. A human-based multi-gene signature enables quantitative drug repurposing for metabolic disease. *Elife* 11 (2022).
25. Guo, W. H. et al. Crosstalk between PKC and MAPK pathway activation in cardiac fibroblasts in a rat model of atrial fibrillation. *Biotechnol Lett* 42, 1219-1227 (2020).
26. Naruse, K. et al. Activation of vascular protein kinase C-beta inhibits Akt-dependent endothelial nitric oxide synthase function in obesity-associated insulin resistance. *Diabetes* 55, 691-698 (2006).
27. Masson, S. W. C. et al. Leveraging genetic diversity to identify small molecules that reverse mouse skeletal muscle insulin resistance. *Elife* 12 (2023).
28. Algül, S. et al. EGFR/IGF1R Signaling Modulates Relaxation in Hypertrophic Cardiomyopathy. *Circ Res* 133, 387-399 (2023).
29. Ferguson, K. M., Hu, C. & Lemmon, M. A. Insulin and epidermal growth factor receptor family members share parallel activation mechanisms. *Protein Sci* 29, 1331-1344 (2020).
30. Skurski, J. et al. Loss of iRhom2 accelerates fat gain and insulin resistance in diet-induced obesity despite reduced adipose tissue inflammation. *Metabolism: clinical and experimental* 106, 154194 (2020).
31. Bowman, C. J., Ayer, D. E. & Dynlacht, B. D. Foxk proteins repress the initiation of starvation-induced atrophy and autophagy programs. *Nat Cell Biol* 16, 1202-1214 (2014).

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

32. Jahng, J. W. S. et al. Iron overload inhibits late stage autophagic flux leading to insulin resistance. *EMBO Rep* 20, e47911 (2019).
33. Klaeger, S. et al. The target landscape of clinical kinase drugs. *Science* 358 (2017).
34. Brass, L. M. Hormone replacement therapy and stroke: clinical trials review. *Stroke* 35, 2644-2647 (2004).
35. Harman, S. M., Naftolin, F., Brinton, E. A. & Judelson, D. R. Is the estrogen controversy over? Deconstructing the Women's Health Initiative study: a critical evaluation of the evidence. *Ann N Y Acad Sci* 1052, 43-56 (2005).
36. Simon, J. A. et al. Postmenopausal hormone therapy and risk of stroke: The Heart and Estrogen/progestin Replacement Study (HERS). *Circulation* 103, 638-642 (2001).
37. Nakhuda, A. et al. Biomarkers of browning of white adipose tissue and their regulation during exercise- and diet-induced weight loss. *Am J Clin Nutr* 104, 557-565 (2016).
38. Slentz, C. A., Houmard, J. A. & Kraus, W. E. Exercise, abdominal obesity, skeletal muscle, and metabolic risk: evidence for a dose response. *Obesity (Silver Spring)* 17 Suppl 3, S27-33 (2009).
39. Slentz, C. A. et al. Effects of exercise training intensity on pancreatic beta-cell function. *Diabetes Care* 32, 1807-1811 (2009).
40. Houmard, J. A. et al. Effect of the volume and intensity of exercise training on insulin sensitivity. *J Appl Physiol* (1985) 96, 101-106 (2004).
41. Sedghamiz, H., Morris, M., Craddock, T. J. A., Whitley, D. & Broderick, G. Bio-ModelChecker: Using Bounded Constraint Satisfaction to Seamlessly Integrate Observed Behavior With Prior Knowledge of Biological Networks. *Front Bioeng Biotech* 7 (2019).
42. Sedghamiz, H. et al. Computation of Robust Minimal Intervention Sets in Multi-Valued Biological Regulatory Networks. *Frontiers in Physiology* 10 (2019).
43. Vashishtha, S., Broderick, G., Craddock, T. J. A., Fletcher, M. A. & Klimas, N. G. Inferring Broad Regulatory Biology from Time Course Data: Have We Reached an Upper Bound under Constraints Typical of In Vivo Studies? *Plos One* 10 (2015).
44. Nikitin, A., Egorov, S., Daraselia, N. & Mazo, I. Pathway studio--the analysis and navigation of molecular networks. *Bioinformatics* 19, 2155-2157 (2003).
45. Pearl, J. *Causality: models, reasoning and inference*. Vol. 29 (Springer, 2000).
46. Aliferis, C. F., Statnikov, A., Tsamardinos, I., Mani, S. & Koutsoukos, X. D. Local Causal and Markov Blanket Induction for Causal Discovery and Feature Selection for Classification Part I: Algorithms and Empirical Evaluation. *Journal of Machine Learning Research* 11, 171-234 (2010).
47. Roider, H. G., Kanhere, A., Manke, T. & Vingron, M. Predicting transcription factor affinities to DNA from a biophysical model. *Bioinformatics* 23, 134-141 (2007).
48. Roider, H. G., Manke, T., O'Keefe, S., Vingron, M. & Haas, S. A. PASTAA: identifying transcription factors associated with sets of co-regulated genes. *Bioinformatics* 25, 435-442 (2009).



## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

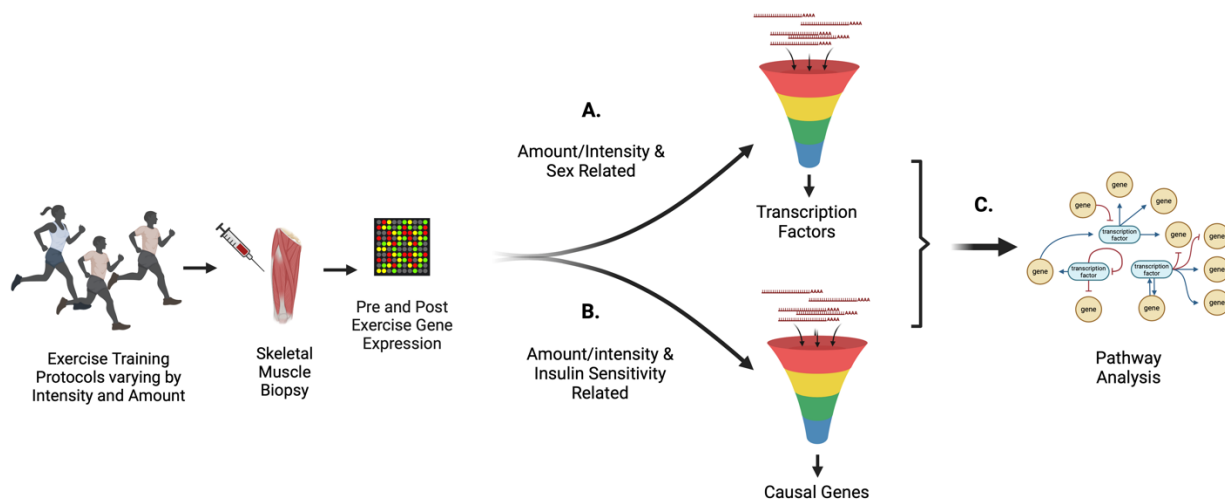
49. Efroni, S., Schaefer, C. F. & Buetow, K. H. Identification of key processes underlying cancer phenotypes using biologic pathway analysis. *PLoS One* 2, e425 (2007).
50. Statnikov, A., Lytkin, N. I., Lemeire, J. & Aliferis, C. F. Algorithms for discovery of multiple Markov boundaries. *J Mach Learn Res* 14, 499–566 (2013).
51. Sopariwala, D., Nguyen, H. & Narkar, V. Estrogen-related Receptor Signaling in Skeletal Muscle Fitness. *Int J Sports Med* 44, 609-617 (2023).
52. Novichkova, S., Egorov, S. & Daraselia, N. MedScan, a natural language processing engine for MEDLINE abstracts. *Bioinformatics* 19, 1699-1706 (2003).
53. Sedghamiz, H., Morris, M., Craddock, T. J. A., Whitley, D. & Broderick, G. Bio-ModelChecker: Using Bounded Constraint Satisfaction to Seamlessly Integrate Observed Behavior With Prior Knowledge of Biological Networks. *Front Bioeng Biotechnol* 7, 48 (2019).
54. Chuffed, a lazy clause generation solver (Department of Computing and Information Systems University of Melbourne, Australia, 2016).



## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

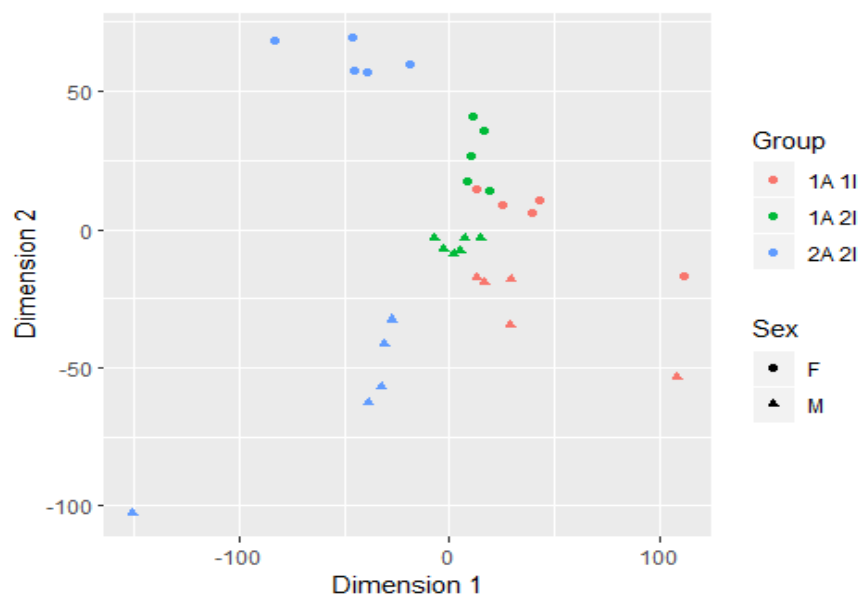
### FIGURES:

Figure 1.



**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.

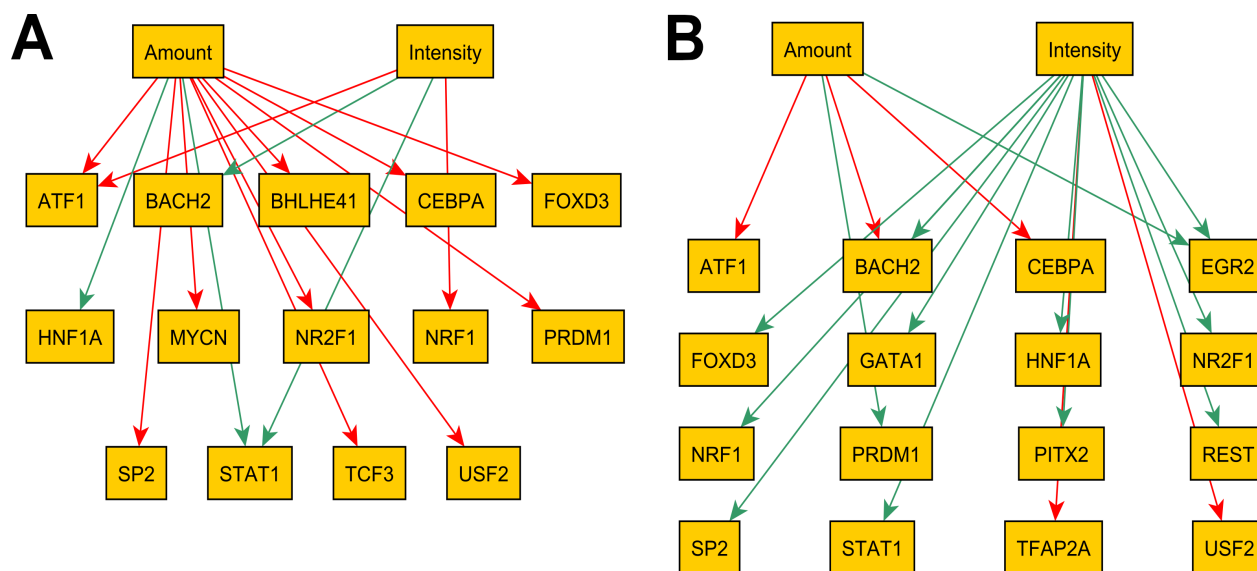
Figure 2.



## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

**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.

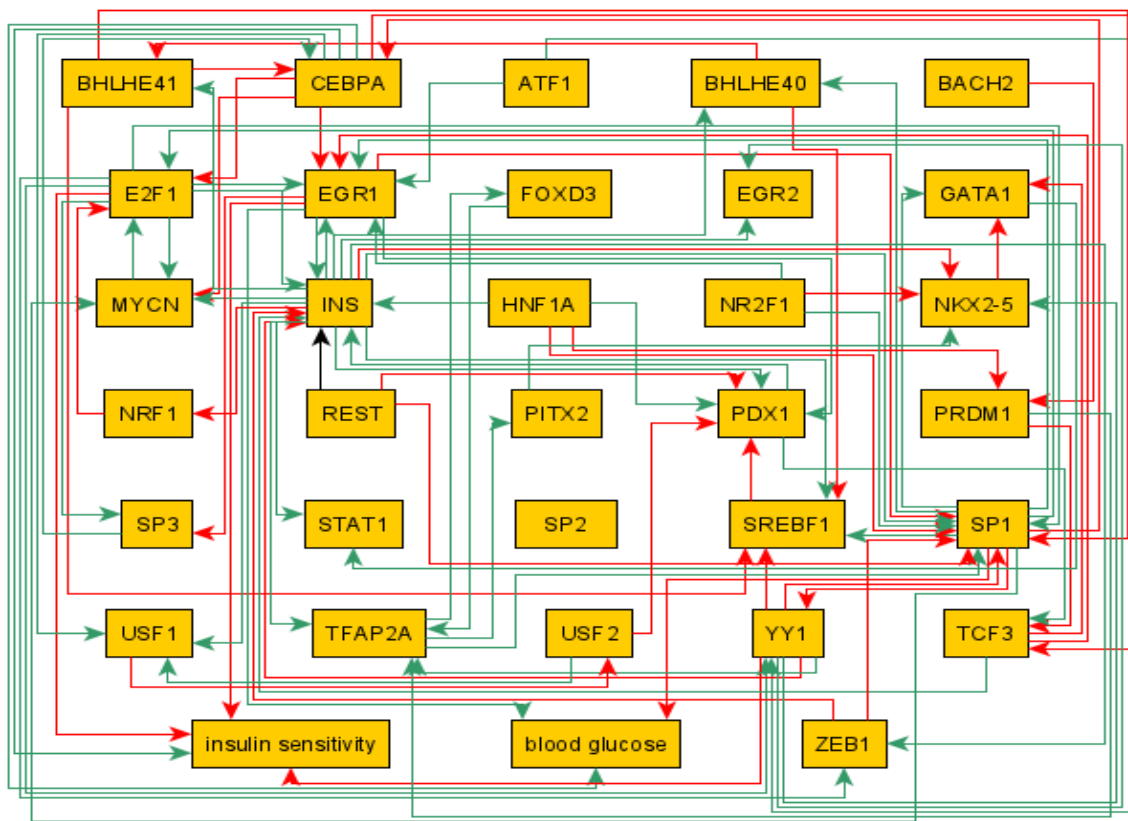
**Figure 3.**



**Figure 3. Exercise amount and intensity transcription factor gene targets; predicted direct targets of exercise amount and intensity in models for women (A) and men (B).** 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).

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

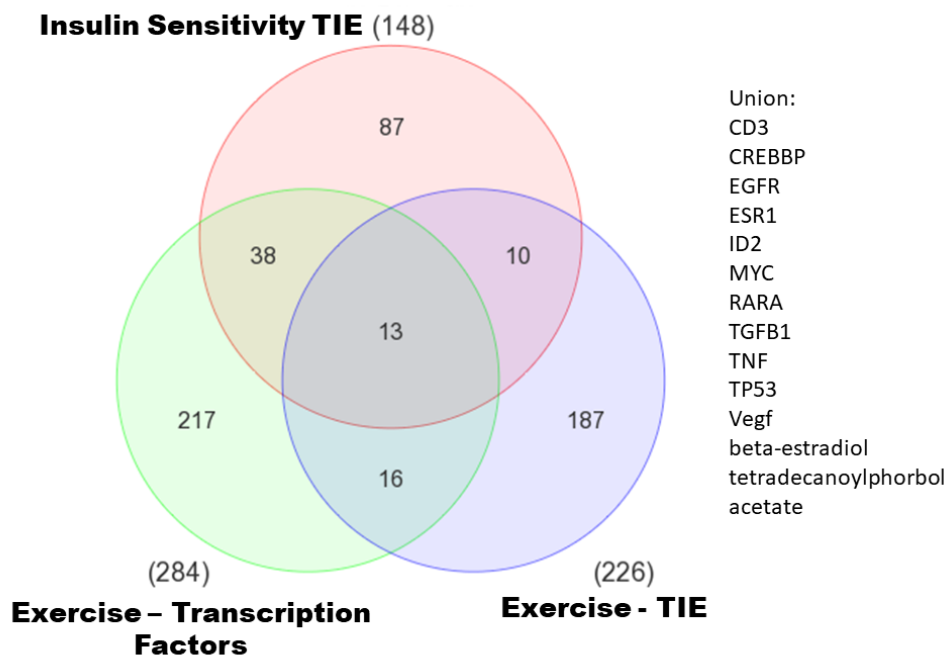
**Figure 4:**



**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 modelled 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.

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

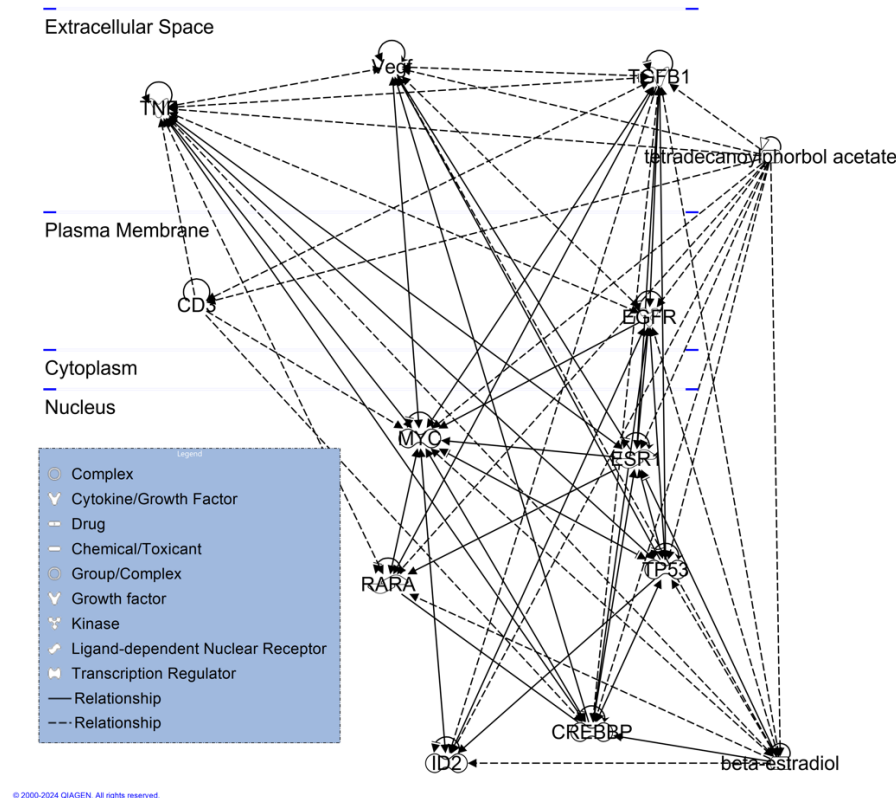
Figure 5.



**Figure 5. Relationships among exercise- and insulin sensitivity-related gene sets.** Gene transcriptome candidates derived from Figure 1, steps A, B, and C are represented as discs in green, purple, and red respectively. Thirteen transcripts or those involved in a synthesis pathway (i.e., for beta-estradiol or tetradecanoylphorbol acetate) in common were identified. The designations are listed to the right.

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

**Figure 6.**



**Figure 6.** Biological relationships among exercise- and insulin sensitivity-related genes and factors identified in the analysis presented in Figure 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. Beta-estradiol, although a circulating hormone, has direct gene regulatory functions through the estrogen receptor, which are necessary for maintaining skeletal muscle fitness.<sup>51</sup>

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

### TABLES

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

|                              | STRIDE I & 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 <sup>2</sup> )     | 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               |               |             |                        |             |                       |             |
| S <sub>i</sub> (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 S <sub>i</sub> (%) | 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

**Table 2.** Summary of Figure 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 |
|-------------|-------------|---------|---------|---------|---------|
| 205590_at   | RASGRP1     | -0.688  | -0.839  | -0.647  | -0.749  |
| 213442_x_at | SPDEF       | -2.054  |         |         |         |
| 236423_at   | --          | 1.488   | 0.869   | 1.668   | 1.097   |
| 239427_at   | SLAMF1      | -0.719  |         |         |         |
| 202094_at   | BIRC5       |         | -0.770  |         | -0.817  |
| 234329_at   | CLIC5       |         | -2.537  |         |         |
| 201268_at   | NME1/NME2   |         |         | 2.969   |         |
| 214339_s_at | MAP4K1      |         |         | -2.967  |         |
| 243897_at   | --          |         |         |         | -1.482  |

## Kraus, Sex-specific Skeletal Muscle Gene Expression Exercise

**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 |              |       |        |
|--------------|--------|---------|--------------|----------|-------------|----------|--------------|-------|--------|
| Gene         |        |         |              |          |             |          |              |       |        |
| Probe ID     | 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 |             |      |        |
|------------|--------|---------|--------------|-----|--------------|------|-------------|------|--------|
|            |        |         |              |     |              |      |             |      |        |
| Probe ID   | Symbol | n       | mean (sd)    | n   | mean (sd)    | n    | mean (sd)   | F    | p      |
| 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 |