geneXX: An online tool for the exploration of transcript changes in skeletal muscle associated with exercise

Saskia Reibe[1,2] , Marit Hjorth[1,2] , Mark A. Febbraio[1,2] & Martin Whitham[1,2]*

[1]Cellular and Molecular Metabolism Laboratory, Diabetes & Metabolism Division, Garvan Institute of Medical Research, Sydney, New South Wales, Australia

[2]St Vincent's Clinical School, University of New South Wales, Sydney, New South Wales, Australia

* Corresponding author, Garvan Institute of Medical Research, 384 Victoria Street, Darlinghurst, New South Wales 2010, Australia, ph: +61292958214, email: m.whitham@garvan.org.au

Running head: GeneXX database of exercise regulated genes

Keywords: Exercise, Meta-analysis, Transcriptomics, Skeletal muscle

Author contributions statement

Conceptualization, S.R, M.W; Methodology, S.R; Software, S.R; Formal Analysis, S.R; Investigation, S.R; Data curation S.R; Writing - Original draft, S.R, M.W, M.H; Writing - review and editing, MW, M.A.F; Visualization, S.R; Supervision, M.W, M.A.F; Funding Acquisition, M.A.F

Abstract

Exercise stimulates a wide array of biological processes, but the mechanisms involved are incompletely understood. Many previous studies have adopted transcriptomic analyses of skeletal muscle to address particular research questions, a process that ultimately results in the collection of large amounts of publicly available data that has not been fully integrated or interrogated. To maximize the use of these available transcriptomic exercise data sets, we have downloaded and re-analyzed them and formulated the data into a searchable online tool, geneXX. GeneXX is highly intuitive, free, and provides immediate information regarding the response of a transcript of interest to exercise in skeletal muscle. To demonstrate it's utility, we carried out a meta-analysis on the included data sets and show transcript changes in skeletal muscle that persist regardless of sex, exercise mode and duration, some of which have had minimal attention in the context of exercise. We also demonstrate how geneXX can be used to formulate novel hypotheses on the complex effects of exercise, using preliminary data already generated. This resource represents a valuable tool for researchers with interests in human skeletal muscle adaptation to exercise.

Introduction

Physical exercise is typically referred to as an essential aspect of general health and well-being and the capacity to perform exercise is the single most powerful predictor of mortality (22, 31). It is associated with a reduced risk of a number of chronic conditions including cardiovascular disease, cancer, type 2 diabetes, depression, sarcopenia and osteoporosis (8, 9, 15, 21, 39, 45, 50, 54). The diverse effects of exercise on human function attracts, therefore, the interest of researchers in a wide variety of disciplines. Despite this, the complex underlying mechanisms mediating the effects of exercise are incompletely understood. Some effects have been attributed to increased energy consumption and cardiorespiratory fitness, reduced adiposity and circulating lipids, metabolic and immunological adaptations and the maintenance of skeletal muscle mass. Pertinently, since skeletal muscle undergoes significant metabolic perturbations during repeated contraction, regular physical activity induces extensive molecular adaptations, often leading to improved muscle function (13). Further, there is mounting evidence that during exercise, skeletal muscle can participate in tissue cross-talk via proteins classically secreted or enclosed within extracellular vesicles (51, 52)

There are several challenges to fully understanding signaling pathways altered in skeletal muscle in response to repeated activity. In particular, conservative estimates suggest there are over 10,000 gene products abundantly expressed in skeletal muscle across several orders of magnitude (10). In an attempt to make sense of this level of complexity, transcriptomic approaches such as Microarrays and RNA-seq have some clear advantages. For example, one can derive accurate quantitative information of a large number of transcripts in an unbiased manner, inferring information that might not have been gleaned from specific hypothesis driven experiments. To identify molecular pathways or genes involved in exercise adaptation, several research groups have analyzed the skeletal muscle transcriptome of human subjects undergoing an acute exercise bout with or without the implementation of a chronic exercise training intervention. This has given rise to large amounts of data that has been used to analyze exercise related gene expression patterns (32), to compare these patterns with gene

expression in ageing (29) or muscle disease (46), or to identify exercise responsive myokines (34). However, one disadvantage of using transcriptomic assessments of skeletal muscle in many research contexts is that it is often unpractical or unfeasible to collect data on a large number of participants. This can limit statistical power, which is problematic given the known heterogeneity of human responses to exercise.

Meta-analyses of data sets are useful in order to strengthen the significance of gene expression profiles or effects that might go unacknowledged in single experiments, identify biomarkers, or simply help formulate new research hypotheses (1, 5, 37, 40). Mindful of the considerable data that is often under utilized in these contexts, we identified and downloaded 10 publically available transcriptomic data sets from skeletal muscle of healthy human subjects undergoing acute endurance or resistance exercise. We then re-analyzed the data and made them available in an intuitive to use and searchable online tool, geneXX. We discuss here, how data extracted can be analyzed in a way to create novel hypotheses on aspects of gene responses to exercise on data already generated.

Methods

Data collection and review

Data sets were selected and downloaded from Gene Expression Omnibus (GEO NCBI) database (12) and are summarized in Table 1. Criteria for inclusion were that the data were collected on healthy participants completing an acute bout of endurance or resistance exercise, before or at the end of a chronic training regime, as defined by the conductors of each study. Included are both cross-sectional (exercise versus sedentary) or within subjects (pre versus post exercise) comparisons analyzing biopsies of the *vastus lateralis* or *biceps brachii* (GSE24235 only) and in both male and female participants of all ages. Transcriptome data sets were either derived using Microarray (Illumina, Affymetrix or Agilent) or RNA-seq technology (Illumina HiSeq 2000). Data sets were excluded when they were processed before 2010 to keep all data on a similar level of microarray and RNA-seq technology. Further, data

sets were also excluded when after normalizing the data, quality control criteria were not met (e.g. comparing median expression of samples using boxplots). Finally, data sets were excluded when there were no differentially expressed genes in our analysis pipeline (most often due to low number of study participants).

Individual data set analysis

Following download of the raw data of each set, independent analysis was performed in R (Version 3.4.0) (44) using packages from the Bioconductor consortium (16). RMA (Robust Multichip Average) expression measure of microarray data was computed using the rma function (18), which applies quantile normalization. Data were then analyzed via limma (38) using linear models for the assessment of differential expression, expressed as log2-fold change. For RNA-seq data (GSE60590), fastq files were downloaded from the European Nucleotide Archive (ENA), developed and maintained at the EMBL-EBI. Quality control and subsequent trimming was carried out using FastQC (Version 0.11.5) and Trim Galore! (Version 0.4.0), respectively. Alignment was performed with STAR aligner (Version 2.5.1) (11) using Homo_sapiens.GRCh38.dna.primary_assembly.fa as reference genome. Transcript quantification was performed using RSEM(Version 1.2.26) (23) and only exon alignments were included. Counts were further analyzed with DESeq2 (26) for differential gene expression between groups (exercised *versus* sedentary) by applying a generalized linear model for which counts are modeled using a negative binomial distribution with fitted mean and a gene-specific dispersion parameter. In all instances, differences in gene expression were corrected for multiple hypotheses testing via Benjamini Hochberg correction (3). Each independent analysis generated a result table, detailing gene ID, official gene symbol, type of exercise (resistance or endurance), log fold change/difference between exercise and rest (logFC), whether the bout was completed with or without a preceding training regime ("untrained" or "post training"), adjusted p value (q value), sex of participants and the time in minutes of tissue sampling after completion of the exercise bout (final one in instances when a training regime was implemented).

Shiny web app: geneXX

To enable visualization of all data embedded in the new results tables, on a single gene basis, a Shiny web app (https://shiny.rstudio.com/) was created. Data are expressed as log fold change with a positive value indicating greater transcript abundance consequential of exercise ($q \leq 0.05$). Rows with missing values for official gene symbols and duplicated entries for genes within the same result table were excluded, keeping the entry with the lowest q value where necessary. Data are plotted via ggplot2 (14). GeneXX also displays the frequency of publication hits, herein referred to as Pubmed score, for each selected gene by querying the name paired with "AND exercise" in the pubmed search engine, as well as retrieving the NCBI summary for the gene using the R package rentrez (https://www.rdocumentation.org/packages/rentrez/versions/1.1.0.).

Meta analysis

Individual adjusted p values from each result table were combined by applying the combine.test function from the survcomp package (41) using the Z-transform test (53). Combined p values were again corrected for multiple testing by applying Benjamini Hochberg correction (3) ($q \leq 0.05$). To infer additional significance for exercise regulated genes, the frequency at which the transcript is measured as significantly different across all data sets was measured. To account for the heterogeneous nature of the responses captured within the data sets, an arbitrary, conservative cut-off of 8/19 independent significant occurrences was used to determine overall significance. Significant genes were also cross-referenced against Pubmed score using geneXX (Supplementary Table 1).

Gene enrichment analyses

Enrichment analyses of significant exercise responsive genes was carried out in cluster-profiler (55) and ToppCluster (20). The enrich function of clusterProfiler assesses the extent to which a number of significant genes that associate with a gene ontology (GO) is greater

than expected. ToppCluster was used to interrogate the DisGeNET database (33) basing computation of enrichments on the hypergeometric distribution test. P values were adjusted via Benjamini Hochberg correction ($q \leq 0.05$). The network of resulting terms was manually curated to reduce repetitive terms and manually arranged in Cytoscape (Version 3.5.1) (42).

Results and Discussion

geneXX: An exercise gene exploration tool

To facilitate exploration of skeletal muscle gene responses to exercise, we have developed a new web-based resource: geneXX. The online tool can be accessed freely at http://garvan.org.au/genexx and can provide a wealth of immediate information on the response to exercise of a specific gene. In Figure 1, we showcase the individual functions of the tool. Users can enter any human gene of interest, (e.g. PPARGC1A) and immediately observe log fold change values, adjusted p values (q value) and the time point post exercise at which the transcript was measured, with color and shape coded symbols to indicate statistical significance and sex of participants, respectively. Also included are Pubmed scores and a short summary about the gene of interest from the NCBI gene site. The main feature of geneXX is that it provides an accessible and instant insight into the response of a particular gene of interest to exercise in human skeletal muscle. This is made possible completely independently of a search of the literature and significantly, allows analysis of genes not necessarily reported by the authors of the original publications. We offer this resource for anyone interested in skeletal muscle responses to exercise and encourage exploration of these gene changes from any conceivable research discipline. This new tool may be particularly useful, for example, when attempting to infer some translational significance in humans from genes initially identified in non-human models. For example, Mansueto *et al* recently identified TFEB as a predominant regulator of mitochondrial biogenesis and glucose homeostasis, exclusively in mice (27). Similarly, a selection of newly characterized exercise-responsive gene products have been described in transgenic mouse models, such as METRNL in the case of PGC-1α 4 transgenic mice (35) and FGF21 examined in models of muscle specific Akt1 overexpression (19). As shown in Figure 2, geneXX provides a rapid first enquiry into how these genes have responded to acute exercise bouts in humans, stratified, where possible, by exercise type, training status and sex of the participants and time point after the exercise bout. Accession numbers of the original data for each analysis is also shown for further, independent enquiry. Importantly, the data visualized are by no means a complete

characterization of each gene, but rather a convenient resource to summarize data that has already been collected, but perhaps not reported in the literature. Importantly, new data sets on skeletal muscle responses to exercise will be added to the geneXX database as and when available, facilitating a first point of enquiry for any gene of interest in the context of exercise.

Meta-analysis of geneXX data highlights exercise responsive genes with little previous connection to exercise

Gene responses to exercise can be highly variable and heterogeneous. By combining several data sets involving a wide range of exercise types, durations and different participants one can attempt to identify genes that robustly respond to exercise in a hypothesis free manner. In total we included 19 result tables comparing exercise with rest (Table 1) in the meta-analysis, and 106 genes were persistently shown to be regulated by exercise (Figure 3, Supplementary Table 1). By way of proof of principle for this approach, gene ontology enrichment analysis of these genes identified several biological processes known to be stimulated by exercise, such as muscle tissue development, metabolic control and kinase signaling (Figure 4). Furthermore, genes well characterized in the context of skeletal muscle responses to exercise, such as PPARGC1A (32) or VEGFA (43) are featured in the top 15 responsive genes (Table 2) and scored highly in the Pubmed score. Interestingly, a selection of genes, such as IFRD1, SDC4, HEY1 and VGLL2 have had minimal attention in the published literature in the context of exercise, despite the high frequency in exercise transcriptomic data sets in which they are shown to be differentially regulated (Table 2). We envisage this information as being one example of how geneXX is particularly useful in the initiation of hypothesis driven research. For example, IFRD1 encodes the protein interferon related development regulator 1 and has been shown to be involved in the process of muscle cell differentiation. Gain- and loss-of function experiments show that up-regulation of the mouse homologue of IFRD1, PC4, significantly amplified myogenesis in adult muscle after injury (30). Taken together, our meta-analysis results may indicate that IFRD1 might therefore play a role in muscle

hypertrophy after exercise, which, to our knowledge, has not been reported as such before. Similarly, SDC4, which encodes the protein syndecan-4, has had little attention in the context of exercise. Interestingly, the syndecan heparan sulphate proteoglycans, through their interaction with regulators of endosomal sorting, are thought to play a role in the complex biogenesis of exosomes (2). Since we recently demonstrated that exercise stimulates the release into circulation of small vesicles such as exosomes, some of which were liberated from the exercising limb (52), whether SDC4 might contribute to this process in skeletal muscle warrants further investigation. By compiling published data sets and re-analyzing them with a rigorous meta-analysis we are, therefore, able to create novel hypotheses on adaptation to exercise in the absence of our own preliminary data collection.

Exploration of exercise responsive genes potentially mediating resistance to disease. Regular exercise is known to affect both the prevalence and treatment of non-communicable diseases. Since we identified genes robustly responding to exercise, we carried out enrichment analyses on these genes against the DisGeNET database of conditions and phenotypes related to medical genetics (Figure 5, Supplementary Table 2). While these analyses by no means provide proof of causation, we are able to identify a selection of genes responding to exercise that play a role in the etiology of specific phenotypes. For example, we observed significant enrichment in exercise responsive genes that are also associated with type 2 diabetes, listed and highlighted in Supplementary Table 2. In reviewing these genes using geneXX, we observed several genes, CLIP1, PCNT, RRAD, SIK2 and USP2 that present with a Pubmed score of 0. Therefore, we identify genes that are annotated in the context of diabetes and are exercise responsive, but are not widely reported as such. Using this approach, one can therefore create a first line of enquiry regarding the effects of exercise on the etiology of a disease phenotype on data already collected.

Caveats to the geneXX analysis

We introduce here geneXX, a web tool offering a first enquiry exploration of gene responses

to exercise in skeletal muscle. So far this tool includes 19 different comparisons and stratified by exercise type, gender, training status and time point of biopsy after exercise. Clearly, the data presented by our new tool is reliant on relevant previous data and robust and efficient data collection methods therein. We chose to perform a meta-analysis of the transcriptomic data based on combining p values rather than directly merging the raw data as the latter is usually restricted to selecting studies from the same array platform and even then precautions have to be taken (as discussed in (47)). The most common method for combining p values in meta-analyses is to apply Fisher's method. Since it has been demonstrated that the Z method is superior in terms of power and precision (53), we chose the latter for our meta-analysis. Significantly, as RNA-seq analysis technologies become more accessible, merging and normalizing the raw count data would be an attractive approach, offering greater statistical power. As more RNA-seq data sets on skeletal muscle responses to exercise become available, it is possible these comparisons can be added to geneXX. We encourage researchers with all appropriate transcriptome datasets to make them available for inclusion, when suitable.

The meta-analysis as we performed it, is unquestionably affected by the heterogeneity of the data. For example, the utilized studies frequently vary in their timing of biopsy, and a consistent temporal analysis of gene expression responses to exercise would improve confidence in the data. However, while studies that try to identify time-course dependent changes in gene expression (32) might provide a greater transcriptomic coverage, our approach of combining all data, inclusive of different time points, exercise types and participants allows us to pick up robust changes in gene expression instead of false positives. Significantly, many of these gene changes persist across many variables, yet have had little attention in the context of skeletal muscle adaptation to exercise.

Batch effects are unavoidable when working, as here, with different data sets generated on different high-throughput technologies. Performing a meta-analysis will not entirely

circumvent these (4) but will result in features with lower statistical power or with small effect size being lost. However, we reiterate that by performing a conservative analysis we aimed to retrieve a shorter list of genes affected by exercise that contains little false positives rather identify many genes affected at the expense of a high false positive rate.

In conclusion, we have downloaded and independently analyzed a series of transcriptomic data sets measuring mRNA changes in skeletal muscle with exercise. In combining these analyses, we introduce geneXX, a freely available web tool, which allows one to visualize these data, stratified by exercise type, training status, sex and time point post exercise and to gauge the extent of previous focused analyses on any gene of interest. We also show how these data can be analyzed to formulate novel hypotheses on the involvement of under appreciated genes in skeletal muscle responses to exercise. GeneXX will be frequently updated and we encourage its use in exercise related research disciplines.

Acknowledgements

This project was supported by a Project Grant (APP1122227) and a Senior Principal Research Fellowship (APP1116936) from National Health & Medical Research Council (NHMRC) awarded to M.A.F.

References

1. **Afroz S, Giddaluru J, Abbas MM, and Khan N**. Transcriptome meta-analysis reveals a dysregulation in extra cellular matrix and cell junction associated gene signatures during Dengue virus infection. *Sci Rep* 6: 33752, 2016.
2. **Baietti MF, Zhang Z, Mortier E, Melchior A, Degeest G, Geeraerts A, Ivarsson Y, Depoortere F, Coomans C, Vermeiren E, Zimmermann P, and David G**. Syndecan-syntenin-ALIX regulates the biogenesis of exosomes. *Nat Cell Biol* 14: 677-685, 2012.
3. **Benjamini Y, and Hochberg Y**. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J Royal Stat Soc* 57: 289-300, 1995.
4. **Bin Goh WW, Wang W, and Wong L**. Why Batch Effects Matter in Omics Data, and How to Avoid Them. *Trends Biotechnol* 35: 498-507, 2017.
5. **Caldas J, and Vinga S**. Global Meta-Analysis of Transcriptomics Studies. *PLoS One* 9: e89318-89311, 2014.
6. **Catoire M, Alex S, Paraskevopulos N, Mattijssen F, Evers-van Gogh I, Schaart G, Jeppesen J, Kneppers A, Mensink M, Voshol PJ, Olivecrona G, Tan NS, Hesselink MKC, Berbee JF, Rensen PCN, Kalkhoven E, Schrauwen P, and Kersten S**. Fatty acid-inducible ANGPTL4 governs lipid metabolic response to exercise. *Proc Natl Acad Sci USA* 111: E1043-E1052, 2014.
7. **Catoire M, Mensink M, Boekschoten MV, Hangelbroek R, Müller M, Schrauwen P, and Kersten S**. Pronounced Effects of Acute Endurance Exercise on Gene Expression in Resting and Exercising Human Skeletal Muscle. *PLoS One* 7: e51066-51010, 2012.
8. **Coughlin K**. A global perspective on cancer prevention. *Br J Nurs* 16: 1252-1252, 2007.
9. **Cruz-Jentoft AJ, Landi F, Schneider SM, Zúñiga C, Arai H, Boirie Y, Chen L-K, Fielding RA, Martin FC, Michel J-P, Sieber C, Stout JR, Studenski SA, Vellas B, Woo J, Zamboni M, and Cederholm T**. Prevalence of and interventions for sarcopenia in ageing adults: a systematic review. Report of the International Sarcopenia Initiative (EWGSOP and IWGS). *Age Ageing* 43: 748-759, 2014.
10. **Deshmukh AS, Murgia M, Nagaraj N, Treebak JT, Cox J, and Mann M**. Deep proteomics of mouse skeletal muscle enables quantitation of protein isoforms, metabolic pathways, and transcription factors. *Mol Cell Proteomics* 14: 841-853, 2015.
11. **Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR**. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15-21, 2013.
12. **Edgar R, Domrachev M, and Lash AE**. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30: 207-210, 2002.
13. **Egan B, and Zierath JR**. Exercise Metabolism and the Molecular Regulation of Skeletal Muscle Adaptation. *Cell Metab* 17: 162-184, 2013.
14. **Ginestet C**. ggplot2: Elegant Graphics for Data Analysis. *J Royal Stat Soc* 174: 245-245, 2011.
15. **Howe TE, Shea B, Dawson LJ, Downie F, Murray A, Ross C, Harbour RT, Caldwell LM, and Creed G**. Exercise for preventing and treating osteoporosis in postmenopausal women. *Cochrane Database Syst Rev* 32: CD000333, 2011.

16. **Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, and Morgan M**. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods* 12: 115-121, 2015.
17. **Hyldahl RD, Nelson B, Xin L, Welling T, Groscost L, Hubal MJ, Chipkin S, Clarkson PM, and Parcell AC**. Extracellular matrix remodeling and its contribution to protective adaptation following lengthening contractions in human muscle. *FASEB J* 29: 2894-2904, 2015.
18. **Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, and Speed TP**. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4: 249-264, 2003.
19. **Izumiya Y, Bina HA, Ouchi N, Akasaki Y, Kharitonenkov A, and Walsh K**. FGF21 is an Akt-regulated myokine. *FEBS Lett* 582: 3805-3810, 2008.
20. **Kaimal V, Bardes EE, Tabar SC, Jegga AG, and Aronow BJ**. ToppCluster: a multiple gene list feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. *Nucleic Acids Res* 38: W96-W102, 2010.
21. **Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM, and Group DPPR**. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 346: 393-403, 2002.
22. **Korpelainen R, Lämsä J, Kaikkonen KM, Korpelainen J, Laukkanen J, Palatsi I, Takala TE, Ikäheimo TM, and Hautala AJ**. Exercise capacity and mortality - a follow-up study of 3033 subjects referred to clinical exercise testing. *Ann Med* 48: 1-8, 2016.
23. **Li B, and Dewey CN**. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12: 323, 2011.
24. **Lindholm MEE, Marabita F, Gomez-Cabrero D, Rundqvist H, Ekström TJ, Tegnér J, and Sundberg CJ**. An integrative analysis reveals coordinated reprogramming of the epigenome and the transcriptome in human skeletal muscle after training. *Epigenetics* 9: 1557-1569, 2014.
25. **Liu D, Sartor MA, Nader GA, Gutmann L, Treutelaar MK, Pistilli EE, IglayReger HB, Burant CF, Hoffman EP, and Gordon PM**. Skeletal muscle gene expression in response to resistance exercise: sex specific regulation. *BMC Genomics* 11: 659-614, 2010.
26. **Love MI, Huber W, and Anders S**. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15: 550, 2014.
27. **Mansueto G, Armani A, Viscomi C, D'Orsi L, Cegli R, Polishchuk EV, Lamperti C, Meo I, Romanello V, Marchet S, Saha PK, Zong H, Blaauw B, Solagna F, Tezze C, Grumati P, Bonaldo P, Pessin JE, Zeviani M, Sandri M, and Ballabio A**. Transcription Factor EB Controls Metabolic Flexibility during Exercise. *Cell Metab* 25: 182-196, 2017.
28. **McLean CS, Mielke C, Cordova JM, Langlais PR, Bowen B, Miranda D, Coletta DK, and Mandarino LJ**. Gene and MicroRNA Expression Responses to Exercise; Relationship with Insulin Sensitivity. *PLoS One* 10: e0127089-0127014, 2015.
29. **Melov S, Tarnopolsky MA, Beckman K, Felkey K, and Hubbard A**. Resistance exercise reverses aging in human skeletal muscle. *PLoS One* 2: e465, 2007.

30. **Micheli L, Leonardi L, Conti F, Maresca G, Colazingari S, Mattei E, Lira SA, Farioli-Vecchioli S, Caruso M, and Tirone F**. PC4/Tis7/IFRD1Stimulates Skeletal Muscle Regeneration and Is Involved in Myoblast Differentiation as a Regulator of MyoD and NF-κB. *J Biol Chem* 286: 5691-5707, 2011.
31. **Myers J, Prakash M, Froelicher V, Do D, Partington S, and Atwood JE**. Exercise capacity and mortality among men referred for exercise testing. *N Engl J Med* 346: 793-801, 2002.
32. **Neubauer O, Sabapathy S, Lazarus R, Jowett JBM, Desbrow B, Peake JM, Cameron-Smith D, Haseler LJ, Wagner KH, and Bulmer AC**. Transcriptome analysis of neutrophils after endurance exercise reveals novel signaling mechanisms in the immune response to physiological stress. *J Appl Physiol* 114: 1677-1688, 2013.
33. **Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, and Furlong LI**. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res* 45: D833-D839, 2017.
34. **Pourteymour S, Eckardt K, Holen T, Langleite T, Lee S, Jensen J, Birkeland KI, Drevon CA, and Hjorth M**. Global mRNA sequencing of human skeletal muscle: Search for novel exercise-regulated myokines. *Mol Met* 6: 352-365, 2017.
35. **Rao RR, Long JZ, White JP, Svensson KJ, Lou J, Lokurkar I, Jedrychowski MP, Ruas JL, Wrann CD, Lo JC, Camera DM, Lachey J, Gygi S, Seehra J, Hawley JA, and Spiegelman BM**. Meteorin-like is a hormone that regulates immune-adipose interactions to increase beige fat thermogenesis. *Cell* 157: 1279-1291, 2014.
36. **Raue U, Trappe TA, Estrem ST, Qian H-R, Helvering LM, Smith RC, and Trappe S**. Transcriptome signature of resistance exercise adaptations: mixed muscle and fiber type specific profiles in young and old adults. *J Appl Physiol* 112: 1625-1636, 2012.
37. **Riddell N, and Crewther SG**. Novel evidence for complement system activation in chick myopia and hyperopia models: a meta- analysis of transcriptome datasets. *Sci Rep* 7: 1-14, 2017.
38. **Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, and Smyth GK**. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 43: e47-e47, 2015.
39. **Rosenbaum S, Tiedemann A, Sherrington C, Curtis J, and Ward PB**. Physical activity interventions for people with mental illness: a systematic review and meta-analysis. *J Clin Psychiatry* 75: 964-974, 2014.
40. **Sanders AR, Drigalenko EI, Duan J, Moy W, Freda J, ring HHHGo, and Gejman PV**. Transcriptome sequencing study implicates immune-related genes differentially expressed in schizophrenia: new data and a meta-analysis. *Transl Psychiatry* 7: e1093-1010, 2017.
41. **Schröder MS, Culhane ACC, Quackenbush J, and Haibe-Kains B**. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. *Bioinformatics* 27: 3206-3208, 2011.
42. **Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, and Ideker T**. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13: 2498-2504, 2003.
43. **Silvennoinen M, Ahtiainen JP, Hulmi JJ, Pekkala S, Taipale RS, Nindl BC, Laine T, Häkkinen K, Selänne H, Kyröläinen H, and Kainulainen H**. PGC-1

isoforms and their target genes are expressed differently in human skeletal muscle following resistance and endurance exercise. *Physiol Rep* 3: e12563, 2015.
44. **Team RC**. R: A language and environment for statistical computing. In: *R Foundation for Statistical Computing*2017.
45. **Thompson PD**. Exercise and physical activity in the prevention and treatment of atherosclerotic cardiovascular disease. *Arterioscler Thromb Vasc Biol* 23: 1319-1321, 2003.
46. **Timmons JA, Larsson O, Jansson E, Fischer H, Gustafsson T, Greenhaff PL, Ridden J, Rachman J, Peyrard-Janvid M, Wahlestedt C, and Sundberg CJ**. Human muscle gene expression responses to endurance training provide a novel perspective on Duchenne muscular dystrophy. *FASEB J* 19: 750-760, 2005.
47. **Tseng GC, Ghosh D, and Feingold E**. Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res* 40: 3785-3799, 2012.
48. **Turan N, Kalko S, Stincone A, Clarke K, Sabah A, Howlett K, Curnow SJ, Rodriguez DA, Cascante M, O'Neill L, Egginton S, Roca J, and Falciani F**. A Systems Biology Approach Identifies Molecular Networks Defining Skeletal Muscle Abnormalities in Chronic Obstructive Pulmonary Disease. *PLoS Comput Biol* 7: e1002129-1002116, 2011.
49. **Vissing K, and Schjerling P**. Simplified data access on human skeletal muscle transcriptome responses to differentiated exercise. *Sci Data* 1: 140041-140049, 2014.
50. **Wadden TA, Webb VL, Moran CH, and Bailer BA**. Lifestyle modification for obesity: new developments in diet, physical activity, and behavior therapy. *Circulation* 125: 1157-1170, 2012.
51. **Whitham M, and Febbraio MA**. The ever-expanding myokinome: discovery challenges and therapeutic implications. *Nat Rev Drug Discov* 15: 719-729, 2016.
52. **Whitham M, Parker BL, Friedrichsen M, Hingst JR, Hjorth M, Hughes WE, Egan CL, Cron L, Watt KI, Kuchel RP, Jayasooriah N, Estevez E, Petzold T, Suter CM, Gregorevic P, Kiens B, Richter EA, James DE, Wojtaszewski J, and Febbraio MA**. Extracellular Vesicles Provide a Means for Tissue Crosstalk during Exercise. *Cell Metab* 27: 237-251, 2018.
53. **Whitlock MC**. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J Evol Biol* 18: 1368-1373, 2005.
54. **Wilkins LW**. *ACSM Guidelines for Exercise Testing and Prescription, Sixth Edition*. 2001, p. 55-55.
55. **Yu G, Wang L-G, Han Y, and He Q-Y**. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omics-a Journal of Integrative Biology* 16: 284-287, 2012.

Figure Legends

Figure 1. geneXX displaying logFC ratios for PPARGC1A expression comparing exercise versus sedentary over all included exercise data-sets. (1) Gene name is entered into the tool and the panel (2) indicates the number of data sets that were analyzed for the selected gene and in how many data sets that it was observed as significantly different. Also viewable is the pubmed score (3) and an NCBI summary of the gene. Data on the gene response to exercise is visualized in the graph panels (4), displaying the log fold change value (y-axis) from each result table, stratified for exercise type, whether the bout was preceded by a training regime and by time point post exercise (x-axis)**.** The data points are shape coded for sex and color-coded for significance $q \leq 0.05$ (dark purple) and $q > 0.05$ (light purple). Below the graph panels, (5) a summary table of the visualized data is shown (limited, here, to 3 data sets for brevity), including accession numbers and methods details from which the data were derived. This table can be downloaded in CSV format for further, independent analysis or visualization, if required. Missing data points from the number of data sets sampled indicates that expression values failed quality control for the analysis.

Figure 2. Using geneXX to investigate individual genes of interest. Skeletal muscle gene expression levels of (a) TFEB, (b) METRNL, and (c) FGF21) after exercise in humans. The graph displays the log fold change value (y-axis) from each result table (19 in total), stratified for exercise type and involvement of training regime, and for the time point the muscle tissue was been taken following exercise (x-axis). The data points are shape coded for sex and color coded for $q \leq 0.05$ (dark purple) and $q > 0.05$ (light purple).

Figure 3. Volcano plot for genes affected by exercise following meta-analysis. Green dots indicate genes that were significant in the initial, individual, comparison of exercise *versus* rest. In red, 106 genes that were significant after performing a meta-analysis over all result tables using the Z-method.

Figure 4: Enriched Gene Ontology (GO) Biological Pathway terms for exercise responsive genes. The dot size represents the number of genes from the result list after meta-analysis that could be mapped to each particular pathway. Also represented are these data expressed as a ratio versus the number of genes in each GO term (x axis). The color of the dot indicates the significance level ($q \leq 0.05$). All terms presented are significantly enriched in the exercise responsive gene dataset

Figure 5. Enrichment of disease states within genes affected by exercise. Significant enrichment of disease annotations when analyzing 106 exercise responsive genes against the DisGeNET database ($q \leq 0.05$)

Table 1. Included data sets with corresponding GEO accession number and experimental conditions. The separation of accession numbers in a, b, c has been made after dividing the data according to experimental conditions. "Time point (min)" refers to the time point of muscle biopsy after the exercise bout (last bout in cases of an included exercise regime). N is the number of study participants in each group of the comparison exercise vs rest.

Table 2. Top 15 exercise responsive genes identified by the meta-analysis, including combined p value (Z-method), q value, significant counts over all result tables and Pubmed score.

geneXX
ene expression after exercise

Thanks for using geneXX. Please be patient while data and figure are loading up, especially after changing the input gene, it might take a couple of seconds to display all the information.

Enter here gene of interest (human official gene symbol)

PPARGC1A

1

Number of analysed datasets in each category

exercise	bout	N
endurance	post-training	6
endurance	untrained	1
resistance	post-training	7
resistance	untrained	4

2

Number of significant ocurrences

exercise	bout	N
endurance	post-training	3
resistance	post-training	6
resistance	untrained	4

Number of publications in context with exercise

498

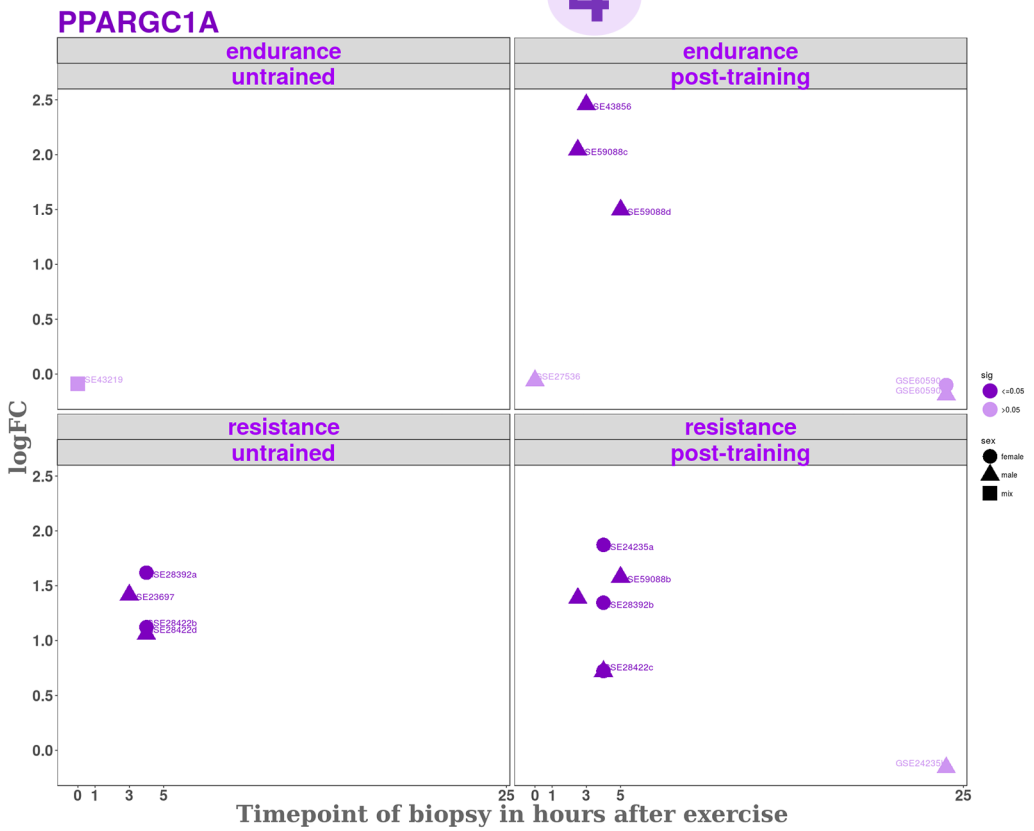
3

NCBI summary for selcted gene

The protein encoded by this gene is a transcriptional coactivator that regulates the genes involved in energy metabolism. This protein interacts with PPARgamma, which permits the interaction of this protein with multiple transcription factors. This protein can interact with, and regulate the activities of, cAMP response element binding protein (CREB) and nuclear respiratory factors (NRFs). It provides a direct link between external physiological stimuli and the regulation of mitochondrial biogenesis, and is a major factor that regulates muscle fiber type determination. This protein may be also involved in controlling blood pressure, regulating cellular cholesterol homoeostasis, and the development of obesity. [provided by RefSeq, Jul 2008]

Effect of exercise on selected gene expressed as logFC

4



5

Summary table for selected gene

Download

dataset	GENE_SYMBOL	logFC	AveExpr	adj.P.Val	time	sex	sig	Type_of_exercise	Muscle_sampled	Comparison	Platform
GSE23697	PPARGC1A	1.42	11.27	0.00	180.00	male	<=0.05	Unilateral resistance (eccentric)	Vastus Lateralis	Exercised vs non-exercised leg	Agilent
GSE24235a	PPARGC1A	1.87	7.34	0.02	240.00	female	<=0.05	Unilateral resistance	Biceps Brachii	Exercised vs non-exercised leg	Affymetrix
GSE24235b	PPARGC1A	-0.16	2.95	0.92	1440.00	male	>0.05	Resistance	Biceps Brachii	Exercised vs non-exercised leg	Affymetrix

a) geneXX
gene expression after exercise

Thanks for using geneXX. Please be patient while data and figure are loading up, especially after changing the input gene, it might take a couple of seconds to display all the information.

Enter here gene of interest (human official gene symbol)

TFEB

Number of analysed datasets in each category

exercise	bout	N
endurance	post-training	6
endurance	untrained	1
resistance	post-training	7
resistance	untrained	4

Number of significant ocurrences

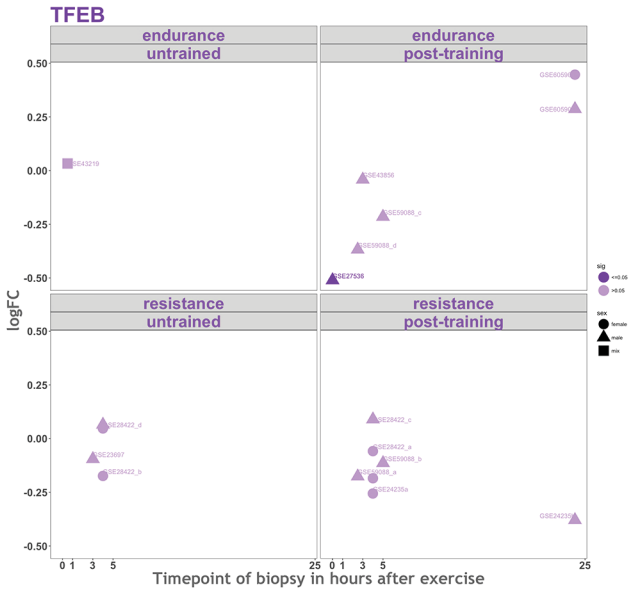
exercise	bout	N
endurance	post-training	1

Number of publications in context with exercise

10

NCBI summary for selcted gene

Effect of exercise on selected gene expressed as logFC



b) geneXX
gene expression after exercise

Thanks for using geneXX. Please be patient while data and figure are loading up, especially after changing the input gene, it might take a couple of seconds to display all the information.

Enter here gene of interest (human official gene symbol)

METRNL

Number of analysed datasets in each category

exercise	bout	N
endurance	post-training	6
endurance	untrained	2
resistance	post-training	7
resistance	untrained	4

Number of significant ocurrences

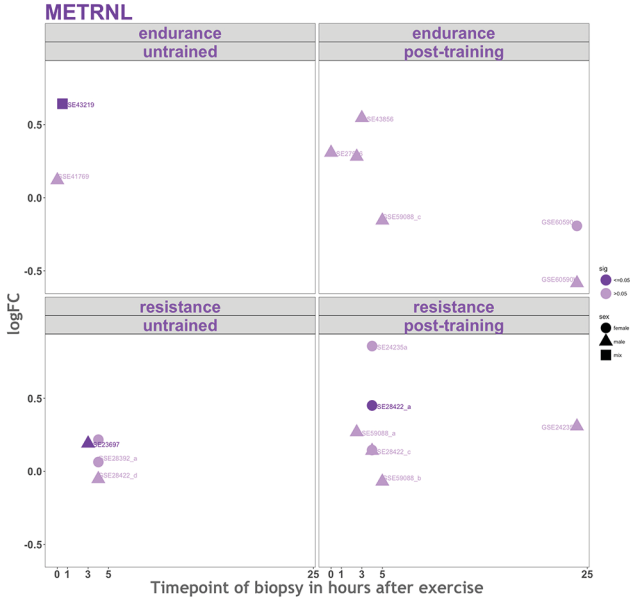
exercise	bout	N
endurance	untrained	1
resistance	post-training	1
resistance	untrained	1

Number of publications in context with exercise

1

NCBI summary for selcted gene

Effect of exercise on selected gene expressed as logFC



c) geneXX
gene expression after exercise

Thanks for using geneXX. Please be patient while data and figure are loading up, especially after changing the input gene, it might take a couple of seconds to display all the information.

Enter here gene of interest (human official gene symbol)

FGF21

Number of analysed datasets in each category

exercise	bout	N
endurance	post-training	3
endurance	untrained	1
resistance	post-training	5
resistance	untrained	1

Number of significant ocurrences

logFC	adj.P.Val	GENE_SYMBOL	time	exercise	sex	bout	da

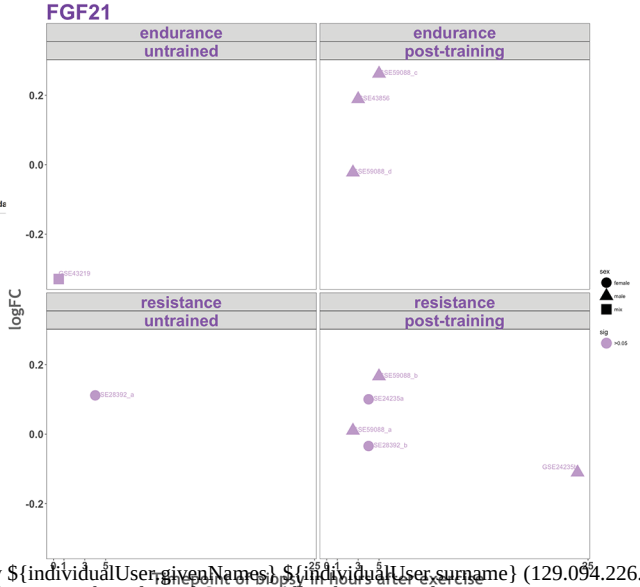
Number of publications in context with exercise

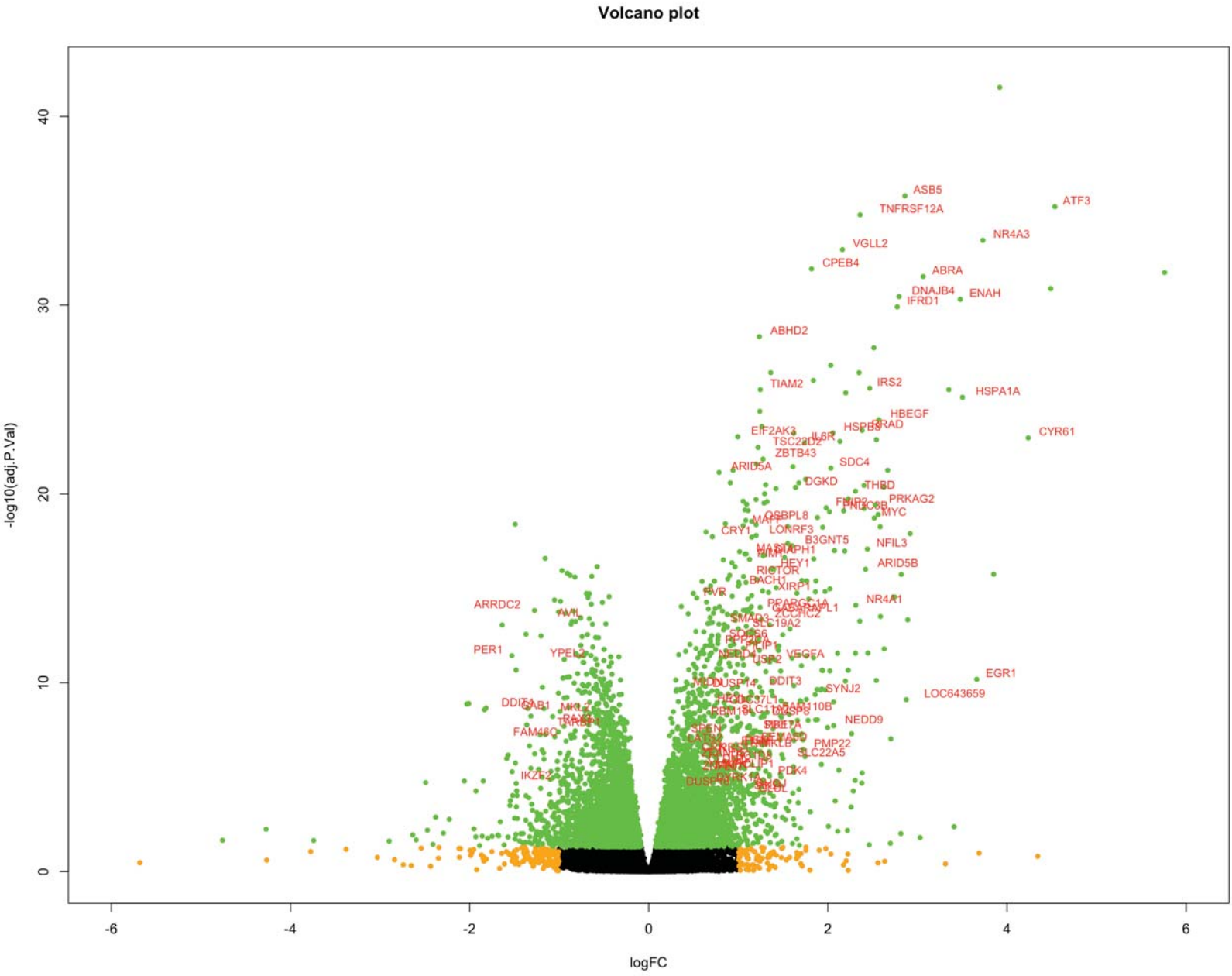
69

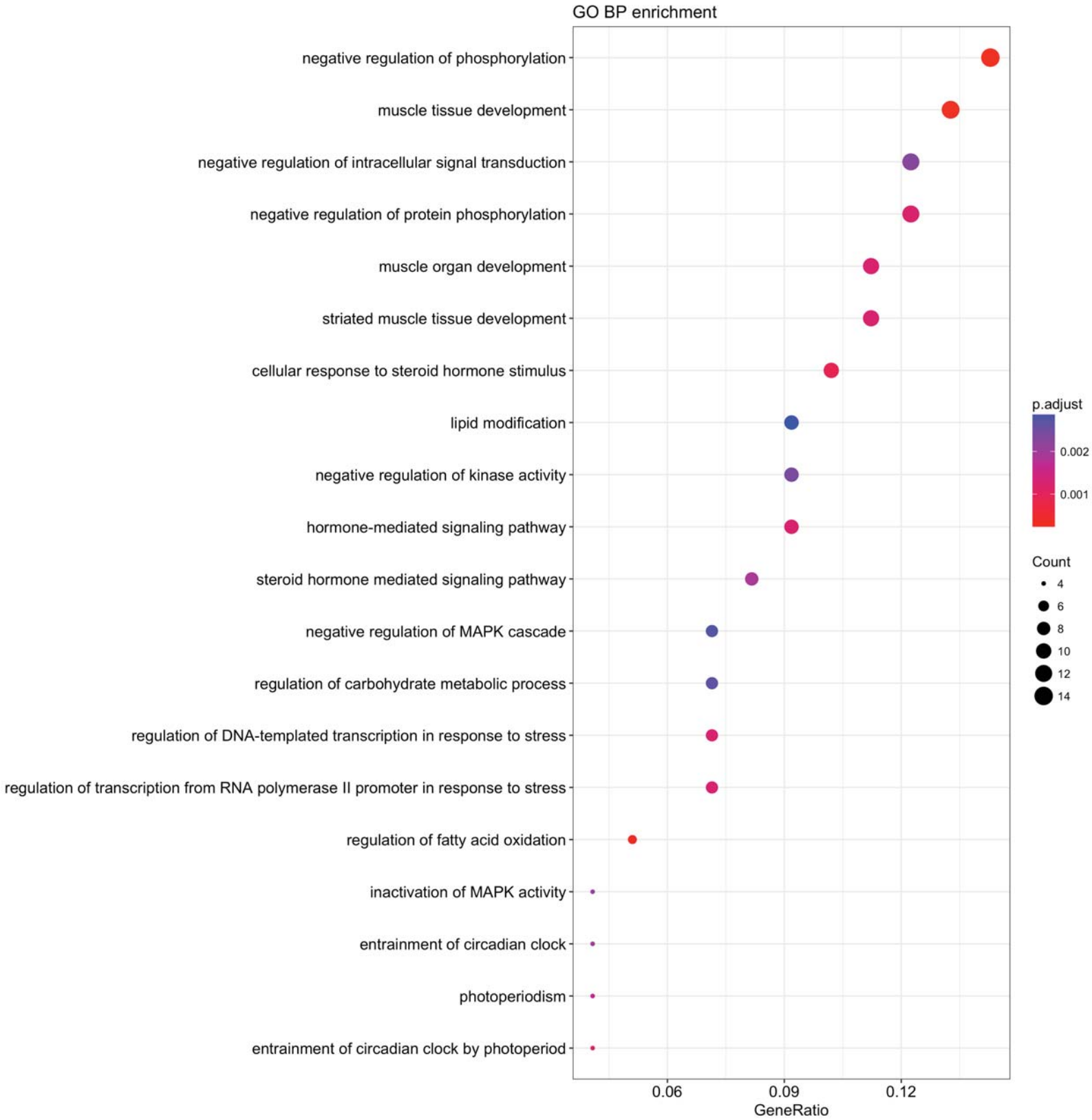
NCBI summary for selcted gene

Theis gene encodes a member of the fibroblast growth factor (FGF) family. FGF family members possess broad mitogenic and cell survival activities and are involved in a variety of biological processes. This protein is a secreted endocrine factor that functions as a major metabolic regulator. The encoded protein stimulates the uptake of glucose in adipose tissue. [provided by RefSeq, Mar 2016]

Effect of exercise on selected gene expressed as logFC







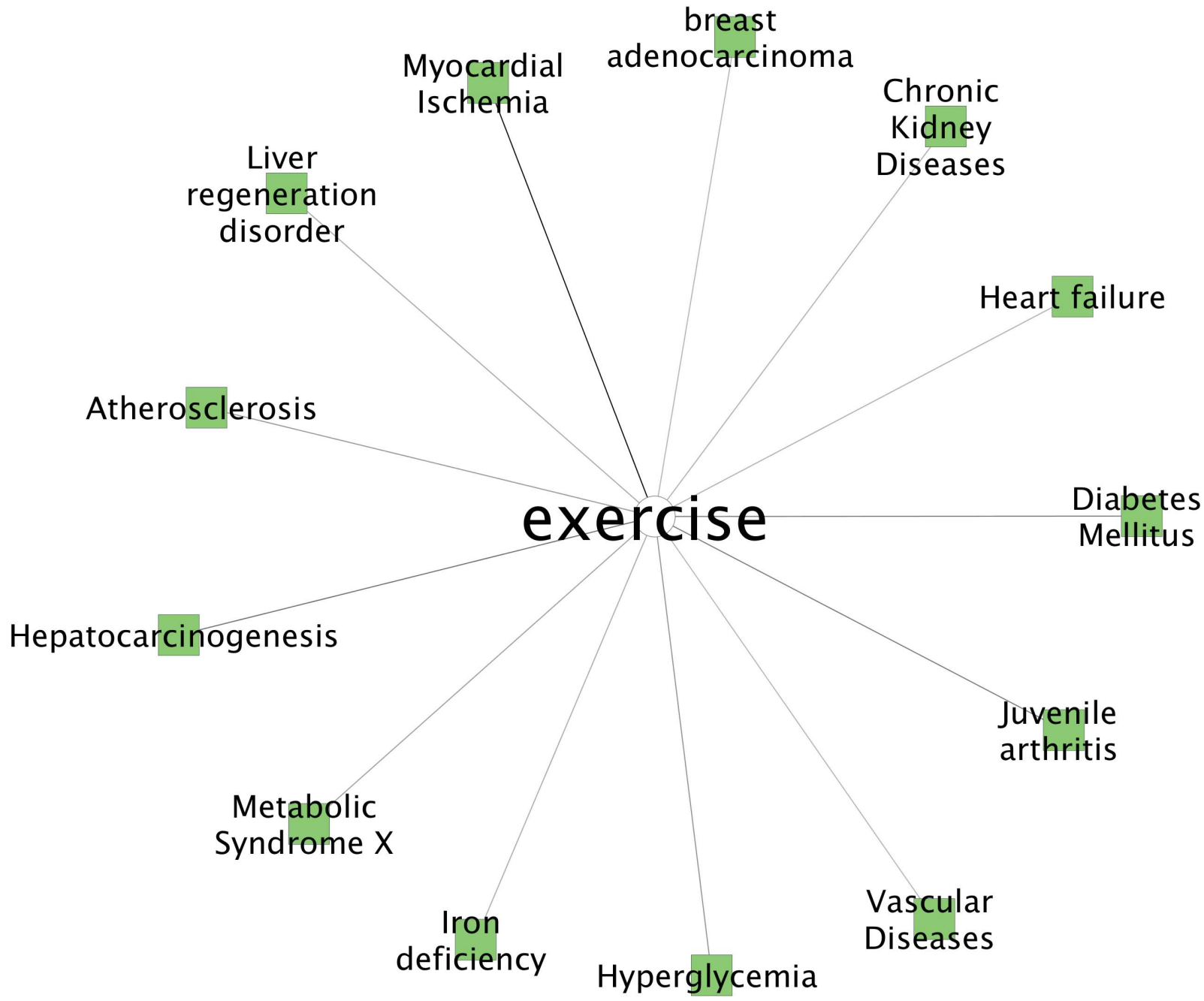


Table 1

GEO#	Type of exercise	Muscle sampled	Training regime*	Sex	Age	Comparison	Time point (min)	N	Platform	Ref
GSE23697	Unilateral resistance (eccentric)	*Vastus Lateralis*	untrained	male	20.9 ± 0.5	Exercised vs. non-exercised leg	180	35	Agilent	(17)
GSE24235a	Unilateral resistance	*Biceps Brachii*	post-training	female	22.7± 0.8	Exercised vs. non-exercised leg	240	4	Affymetrix	(25)
GSE24235b	Resistance	*Biceps Brachii*	post-training	male	24.7±0. 8	Exercised vs. non-exercised leg	1440	3	Affymetrix	(25)
GSE27536	No acute bout	*Vastus Lateralis*	post-training	male	65	Post vs. pre 8 weeks training (at rest)	-	12	Affymetrix	(48)
GSE28392a	Resistance	*Vastus Lateralis*	untrained	female	Young (23 ± 2) or old (85 ± 1)	Post vs pre exercise	240	14	Affymetrix	(36)
GSE28392b	Resistance	*Vastus Lateralis*	post-training	female	Young (23 ± 2) or old (85 ± 1	Post vs pre exercise	240	14	Affymetrix	(36)
GSE28422a	Resistance	*Vastus Lateralis*	post-training	female	Young (24 ± 1) or old (84 ± 1)	Post vs pre exercise	240	14	Affymetrix	(36)
GSE28422b	Resistance	*Vastus Lateralis*	untrained	female	Young (24 ± 1) or old (84 ± 1)	Post vs pre exercise	240	14	Affymetrix	(36)

GSE28422c	Resistance	*Vastus Lateralis*	post-training	male	Young (24 ± 1) or old (84 ± 1)	Post vs pre exercise	240	14	Affymetrix	(36)
GSE28422d	Resistance	*Vastus Lateralis*	untrained	male	Young (24 ± 1) or old (84 ± 1)	Post vs pre exercise	240	14	Affymetrix	(36)
GSE41769	Unilateral endurance	*Vastus Lateralis*	untrained	male	52±5	Post vs Pre exercised leg	0	9	Affymetrix	(6),(5)
GSE43219	endurance	*Vastus Lateralis*	untrained	both	33±2	Post vs pre exercise	30	14	Agilent	(28)
GSE43856	endurance	*Vastus Lateralis*	post-training	male	25±4	Post vs pre exercise	180	8	Illumina	(32)
GSE59088a	resistance	*Vastus Lateralis*	post-training	male	23±1	Post acute exercise vs pre 10 wk exercise training	150	6	Affymetrix	(49)
GSE59088b	resistance	*Vastus Lateralis*	post-training	male	23±1	Post acute exercise vs pre 10 wk exercise training	300	6	Affymetrix	(49)
GSE59088c	endurance	*Vastus Lateralis*	post-training	male	23±1	Post acute exercise vs pre 10 wk exercise training	300	6	Affymetrix	(49)
GSE59088d	endurance	*Vastus Lateralis*	post-training	male	23±1	Post acute exercise vs pre 10 wk exercise training	150	6	Affymetrix	(49)
GSE60590a	Unilateral endurance	*Vastus Lateralis*	post-training	female	26±1	Exercised vs. non-exercised	1440	11	RNA-seq	(24)

						leg				
GSE60590b	Unilateral endurance	*Vastus Lateralis*	post-training	male	27.5±1	Exercised vs. non-exercised leg	1440	12	RNA-seq	(24)

*Post-training: An acute exercise bout followed a period of exercise training.

Table 2

Gene symbol	comb.p.value	adj.p.value	count	Pubmed
NR4A3	4.15E-152	2.93E-148	16	13
NR4A1	3.17E-41	2.24E-37	14	12
GABARAPL1	1.87E-43	1.32E-39	14	9
MAFF	8.08E-42	5.69E-38	14	**3**
VGLL2	9.88E-53	6.96E-49	14	**2**
PPARGC1A	3.16E-48	2.22E-44	13	487
VEGFA	1.15E-33	8.13E-30	13	113
ABRA	8.37E-55	5.90E-51	13	9
PMP22	1.80E-13	1.27E-09	13	6
PRKAG2	4.73E-50	3.33E-46	13	6
SDC4	2.33E-48	1.64E-44	13	**0**
SLC22A5	4.67E-13	3.29E-09	12	6
IFRD1	4.50E-40	3.17E-36	12	**3**
HEY1	1.67E-19	1.18E-15	12	**1**
IRS2	1.55E-28	1.09E-24	11	29