"It All Rolls Downstream: Upstream Control of Physical Activity Regulation"

Running Title: Upstream Regulators of Physical Activity

Brianne M. Breidenbach[1]
Liwen Liu[2]
Troy La[1]
Tatiana N. Castro-Padovani[1]
Nathan Keller[1]
Linda S Pescatello[4]
Matthew M. Robinson[5]
Scott A. Kelly[3]
Kevin Gerrish[2]
J. Timothy Lightfoot, FACSM[1]

[1]Department of Health and Kinesiology, Texas A&M University, College Station, TX 77843

[2] Molecular Genomics Core Laboratory, National Institute of Environmental Health Sciences, Raleigh, NC 27709

[3]Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, USA

[4]Department of Kinesiology, College of Agriculture, Health and Natural Resources, University of Connecticut, Storrs, CT, 06269

[5]School of Biological and Population Health Sciences, College of Public Health and Human Sciences, Oregon State University, Corvallis, OR 97331

Corresponding Author: B. Breidenbach-Jenkins, briannebreidenbach@gmail.com

Conflict of Interest Statement: The authors declare no conflicts of interest.

IRB/Ethics Committee: Given that the data in this study have been previously published, no IRB/Ethics Committee review/approval was/is necessary

Abstract

Physical activity is regulated by a variety of genetic molecules. However, the pathways through which those molecules work to regulate activity is largely unknown. The purpose of this study was to gather the known genetic molecules that are associated with activity regulation and define overall upstream regulator pathways through which these molecules work. We conducted a systematic review to gather all available published datasets related to physical activity regulation, standardized the data for genomic location and species, and used this data, in an unbiased manner to create a dataset that was used: (1) to physically map and visualize all identified molecules to homologous chromosome locations and (2) as the dataset for which an Upstream Regulator Analysis (URA) was conducted using Qiagen Ingenuity Pathway Analysis (IPA) software. Our search resulted in 469 genetic molecules (e.g. genomic variant, transcript, protein, micro-RNA) that were split into brain (n=366) and muscle (n=345) sub-groups, which was our attempt to separate differences in central vs peripheral pathways. The brain and muscle data sets had several potential upstream regulators, the top-rated being β-estradiol as a regulator for 19.5% and 21% of the brain and muscle datasets respectively. To our knowledge, β-estradiol's identification as a potential regulator, is the first evidence to link the well-known effects of sex hormones on physical activity with genetic regulation of physical activity. There were a variety of potential upstream regulators for the molecules collected in this review, but interestingly, three of the top five for both brain and muscle are nuclear receptor binding ligands; estradiol (estrogen receptor), dexamethasone (glucocorticoid receptor), and tretinoin (retinoic acid receptor), indicating a potential role of nuclear receptors in the regulation of physical activity. Selective nuclear receptor modulation may be an area of interest in future mechanistic studies of the genetic regulation of physical activity.

Keywords: genomics, proteomics, upstream regulator analysis, daily activity

Introduction

Physical activity has been broadly defined "as purposeful exercise or movement that expends a significant amount of energy"[1], and more specifically defined as any bodily movement produced by the contraction of the skeletal muscles resulting in an increase in caloric requirements over resting energy expenditure[2]. Several published papers have established that daily physical activity levels have a significant genetic control element[3], with observed heritability of physical activity control ranging from 20-92%[4,5] depending upon the model, analysis method, and duration of data collection. Common environmental factors appear to play a small, if negligible role in controlling daily activity, while unique environmental factors play roles similar in magnitude to genetic factors[3] (e.g. factors such as diet and the presence of environmental toxicants).

Overall, there are a wide-variety of genetically controlled physiological mechanisms that can influence physical activity - including various neurotransmitters[1,6–8], muscle energy sources[9], muscle calcium handling mechanisms[10], and/or substrate availability[11,12]. It has been hypothesized[13] and generally accepted[3] that the genetic molecules that regulate physical activity function at the central level (i.e., brain/neural) and/or at the peripheral level (i.e., skeletal muscle). Data supports gene interference, gene silencing, and/or overexpression at either the central and peripheral levels can significantly influence daily activity[10,14–16]. While sex hormones have long been shown to influence daily activity[17–21], it is unclear whether the actions of the sex hormones on physical activity are mediated through genetic mechanisms (e.g., genetic variation in hormone release/receptor activity) or through direct biological mechanisms (e.g., inhibition of a pathway). The literature establishing that genetics are a significant regulator of daily physical activity is extensive; however, a critical gap remains surrounding the integration of these molecules into multi-omics datasets and the identification of common upstream pathways controlling the regulation of daily physical activity.

The development of multi-omics datasets is most often accomplished through 'omics' data within the same model[22] (e.g., genomic variation, proteomics, transcript expression variation). In the area of "genetic regulation of physical activity", the challenge faced is the lack of 'omics' datasets in the same set of samples. Only one group to our knowledge has produced multiple omics datasets in the same model[10,15,23–25]. Thus, developing an 'omics' dataset of physical activity regulation requires the combining of datasets from various investigators, species, and models. Understanding the systems genetics network(s) involved in physical

activity regulation will help the future synthesis and translation of a general mechanistic understanding of the upstream regulators that control physical activity. Given that the last summary of the known genomic variants associated with physical activity level was published in 2012[26], and that further data regarding transcriptomic, proteomic, and other molecular variants associated with physical activity have been published since then; the purpose of this project was to (i) present an updated summary of the currently known manuscripts and molecules associated with, and mechanistically linked with daily physical activity regulation; (ii) provide a novel visualization of genomic-linked regulators based on chromosome location (mouse and human) and (iii) to identify potential common- upstream regulators of the identified molecules associated with physical activity.

Methods

Overview: We conducted a comprehensive systematic review to gather all available published datasets related to physical activity regulation, standardized the data for genomic location and species, and used this data to create a dataset that was used: (1) to physically map and visualize all identified molecules to chromosome locations and (2) as the dataset for which an Upstream Regulator Analysis (URA) was conducted using Qiagen Ingenuity Pathway Analysis (IPA) software (Qiagen Inc., https://digitalinsights. qiagen.com/IPA, Germantown, MD, USA).

Literature Search: To create our omics dataset, the PubMed/MEDLINE database[27] was utilized to identify manuscripts containing "omic" data (e.g. QTLs, eQTL, miRNA, proteomics) related to physical activity regulation in humans and mice. To build up the literature search in PubMed, we used the following MeSH-terms: Mice, Humans, Motor Activity / genetics, proteome, gene expression profiling, microRNAs, quantitative trait loci, and exercise. Using the same criteria, several additional databases were searched through the University-Library services to locate additional manuscripts (e.g. ERIC, ScienceDirect, Web of Science, Google Scholar). Author expertise and knowledge was also utilized to verify and confirm included manuscripts. Initial searches were completed on April 1, 2020 and subsequent verification of these searches was completed on October 12, 2020, returning a total of 251 potentially qualifying manuscripts. Two authors independently applied eligibility criteria, assessed the quality of the study, and extracted and coded data. We contacted study authors for more information where necessary. Studies were selected based on the following inclusion criteria: (1) original research papers with prospective or retrospective design, (2) physical activity was the primary dependent variable [e.g. wheel running, spontaneous cage activity, locomotion activity, accelerometry data,

self-reported physical activity], (3) contained some variant of 'omics' data (e.g., genomic variation, proteomics, transcript expression variation) as an independent variable that were found to have either been associated with or causative of (i.e., directly affect) physical activity. Manuscripts showing up in multiple searches were counted only once, and review or summary papers were not retained. Studies were excluded when exercise/physical activity was the independent variable (e.g. study was not intended to identify genetic response to exercise). In total, 46 manuscripts qualified for data extraction (Table 1).

Data Extraction and Standardization Methods: From qualifying studies, 464 molecules were identified (Supplemental Table 1), and subdivided into muscle (n=344) and brain (n=361) subsets. Subset assignment was based on whether or not the tissue type was identified in the original manuscript. The majority of genomic variant related manuscripts did not indicate which tissue/sample was utilized for extraction; when this occurred, the molecule(s) were assigned to both datasets (e.g. brain/muscle; Supplemental Table 1). However, when the tissue type was indicated (e.g. soleus muscle vs striatum), the molecules were assigned to either the muscle or brain set. Our intention was to separate differences in central vs peripheral networks, but in reality, the two networks had a large amount of overlapping data points due to lack of disclosure of tissue extraction type for genomic variants.

Integration of omics data from the 46 articles required standardization of genomic location and cross-species comparisons. Genomic locations associated with physical activity from earlier papers [where microsatellites were used as genomic markers] were converted from 'centimorgans' to 'base pairs' utilizing a publicly available converter that was recently decommissioned (http://cgd.jax.org/mousemapconverter/). Genomic locations of the identified variants were updated to the build of the specific species current at the time of data collection for this systematic review (for human GRCh38.13, ref.[28], for mice: GRCm39, ref.[29]). Cross-species comparisons of genetic molecules require translation into a common species to allow these analyses. Given that many of the genetic molecules associated with physical activity (miRNA, transcripts, proteome, and genomic variants) have been identified in mouse models, genomic locations for the molecules in humans (all genomic variants) were converted to the homologous genome location within the mouse chromosome. Ensembl (https:// ensembl.org) was used to convert the human genetic molecules to the mouse homologs[30].

Following extraction and standardization, and when possible, the molecules were: (1) mapped to their chromosome location using the online program MG2C[31] (mg2c.iask.in) with the molecules only found in

humans (i.e., not homologous) mapped to the human chromosome and (2) uploaded and linked to molecules in the IPA database for Upstream Regulator Analysis (URA) (Supplemental Table 3,4). Kramer et. al (2014) describes the algorithms, tools, and visualizations in IPA utilized for the URA analysis[32]. Briefly, URA determines likely upstream regulators that are connected to dataset genes through a set of direct or indirect relationships[32]. Our dataset did not include expression-fold change, expression ratios, or expression log ratios. Consequently, the URA is unable to predict activation status (activated or inhibited) but may be a useful tool to investigate mechanistic targets in future studies. In the results, the p-value of overlap indicated the most significant upstream regulators that influence target molecules from the dataset.

Results

The resulting papers from our systematic literature review (Table 1) netted a total of 469 genetic molecules (Supplemental table 1) that have either been associated with or causative of (i.e., directly affect) physical activity. Genomic variants made up the majority of molecules (n=225, 48%) while other types of genetic molecules were less prominent in the literature (proteomic, n=61, 13%; transcript, n=139, 30%; and miRNA, n=44, 9%).

First Author / Year	Type of genetic molecule reported	Citation in Bibliography	First Author / Year	Type of genetic molecule reported	Citation in Bibliography
Bruneau 2018	Genomic variant	33	Keeney 2012	Proteomic	34
Caetano-Anolles, K, et al. 2016	Transcriptomic	35	Klimentidis 2018	Genomic variant	36
Cai 2006	Genomic variant	37	Knab 2009	Proteomic	1
Comuzzie 2012	Genomic variant	38	Kostrzewa 2014	Genomic variant	39
Dawes 2014	Transcriptomic	23	Kumar 2010	Transcriptomic	40
Dawes 2015	miRNA	24	Leamy 2008	Genomic variant	41
De Moor 2007	Genomic variant	42	Leamy 2010	Genomic variant	43
De Moor 2009	Genomic variant	44	Leamy 2011	Genomic variant	45
Doherty 2018	Genomic variant	46	Lightfoot 2008	Genomic variant	25
Dubreucq, 2012	Proteomic	16	Lightfoot 2010	Genomic variant	47
Eisener-Dorman 2010	Transcriptomic	48	Lin 2018	Genomic variant	49
Ferguson 2014	Proteomic	10	Maestu 2013	Genomic variant	50
Ferguson 2015	Proteomic	15	Nehrenberg 2010	Genomic variant	51
Furuse 2002	Transcriptomic	52	Pistilli 2011	Proteomic / Transcriptomic / Genomic Variant	53
Gielen 2014	Genomic variant	54	Saul 2017	Transcriptomic	55
Hartmann 2008	Transcriptomic / miRNA	56	Simonen 2003	Genomic variant	57
Hillis 2020	Genomic variant	58	Steffan 2002	Genomic variant	59
Jean 2012	Proteomic	60	Tsao 2001	Proteomic	9
Kas 2009	Transcriptomic / Genomic Variant	61	Umemori 2009	Transcriptomic	62
Kelly 2010	Genomic variant	63	Wilkinson 2013	Genomic variant	64
Kelly 2012	Transcriptomic	13	Williams 2014	Transcriptomic	65
Kelly 2014	Transcriptomic	66	Yang 2009	Transcriptomic	67
Keeney 2008	Proteomic	68	Yang 2012	Genomic variant	69

Table 1. Citations reporting genetic molecules associated with/causative of physical activity.

Location Mapping: Of these molecules, 385 were mappable to the mouse chromosome map (Fig. 1) and an additional 74 were mappable only to the human chromosome map (Fig. 2). Ten of the molecules previously found (3 transcripts and 7 genomic variants) were unmappable due to outdated markers and locations that could not be standardized. Of the 116 human genomic variants associated with physical activity (Supplemental Table 1), only 40 had homologs on the mouse genome (indicated in red on Figs. 1 and 2) and two were not mappable due to outdated location markers that could not be standardized.

Figure 1. Chromosome Map for Omics Data Associated With/Causal for Physical Activity.
Color indicates the type of "omic" data; with green- genomic, blue – transcriptomic, purple – proteomic, yellow – miRNA, red – homologous human genomic variant.
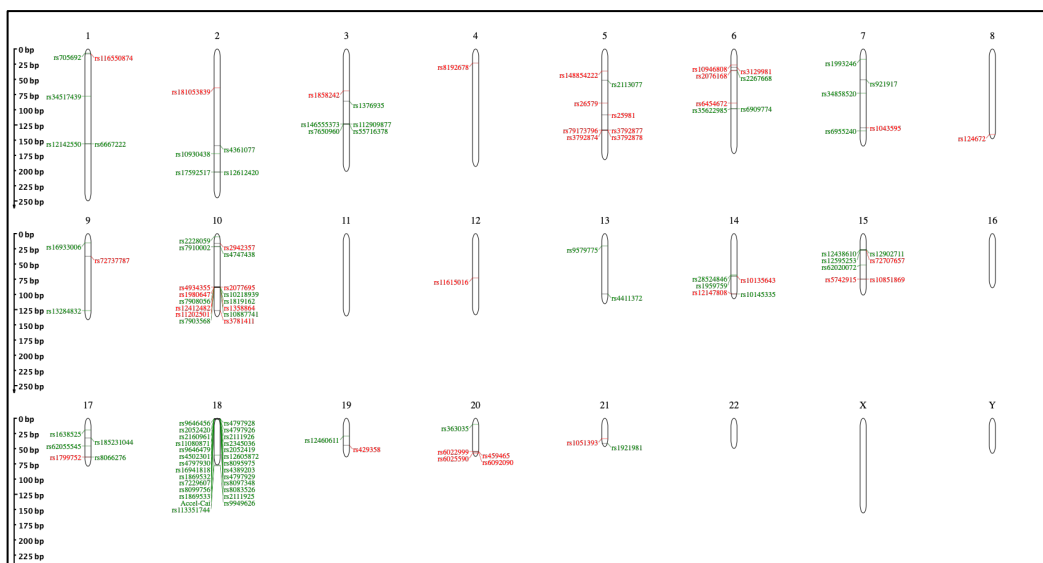

Figure 2. Map of Genetic Molecules Associated with Physical Activity - Human Genome

Potential Upstream Regulators

As detailed in the methods, subsets of data were created for brain and muscle. Not all the molecules from our brain or muscle datasets corresponded to molecules in the IPA database. Of the 366 molecules in the brain and 345 molecules in the muscle dataset, 226 (61.75%) and 195 molecules (56.52%) corresponded to molecules in the IPA database for brain and muscle respectively (Supplemental Table 2). The upstream regulator analysis (Table 2; Suppl Tables 3 and 4) revealed that β-estradiol was the highest predicted regulator for both data sets, with 44 molecules in the brain dataset affected (p=0.000000048) and 41 molecules in the muscle dataset (p=0.00000000179). This finding indicates that 19.5% of the brain data and 21% of the muscle data are potentially regulated upstream by β-estradiol. The other predicted regulators in the top five for both sets (Table 2) were also similar (e.g. TP53 ranked 5th in brain and 2nd in muscle), varying primarily in the number of molecules affected.   Also unique to these data were the potential for exogenous chemical drugs – lipopolysaccharide (2nd brain, 30 molecules / 3rd muscle, 32 molecules) and dexamethasone (3rd brain, 27 molecules / 5th muscle, 27 molecules) - as potential regulators of physical activity. The amount of overlap between the two datasets (muscle vs brain subsets) was not surprising due to the overlap of data points discussed previously in the methods.

Table 2. Top 5 Upstream Regulators for Brain and Muscle Subsets

Brain Upstream Regulator	Molecule Type	p-value of overlap	# molecules affected	IPA Target Molecules in Dataset
beta-estradiol	chemical - endogenous mammalian	4.8E-08	44	ACE, ADORA2A, AFF4, AHR, AKAP10, ANXA2, APBA1, APOE, BAZ1A, CHRM1, CHRM3, CKB, CNR1, CUBN, CYP24A1, CYP4F8, DNAJC1, DRD1, DRD2, EGR3, ENO1, FAM107A, FOXG1, GABRG3, GPR88, GSTM1, HSPA9, HTR4, ICAM5, IL15, IQGAP2, NPW, NTS, PAPSS2, PPARGC1A, PRKCB, SLC38A2, SNAP25, SST, TH, THRA, TLN2, TPH2, VDR
lipopolysaccharide	chemical drug	0.00349	30	ACE,ADORA2A,AHR,ANXA2,APOE,CASR,CUBN,DDN,EGR3,ENO1, FOXG1, GPR3, GPR34, IFNAR2, IL15, LGALS1, NTS, PAPSS2, PLA2G7, PML, PPARD, PPARGC1A, PRKCB, RORA, RXRG, SLC38A2, SLCO2A1, TH, THRA, VDR
dexamethasone	chemical drug	0.0086	27	ACE,APOE,CRHBP,CYP24A1,DACH1,EGR3,ENO1,FAM107A,FOXG1,GPR34, GSTM1, HSPA9, IFNAR2, IL15, LGALS1, MAP2, MSTN, NTS, PPARD, PPARGC1A, PRKCB, RORA, SLCO2A1, SST, THRA, UGT1A6, VDR
tretinoin	chemical - endogenous mammalian	0.000839	25	AHR,ANXA2,APOE,CNR1,CNTNAP2,CUBN,DRD2,IL15, LGALS1,MAP2, NRGN, PA2G7, PML, PPARD, PPARGC1A, PRKCB, RAI14, RXRG, SLC38A2, SCLO2A1, SP7,SST,TH,THRA,VDR

TP53	transcription regulator	0.00488	25	ACE,ANXA2,APOE,CALU,CDH10,CKB,CYP24A1,DLX1,DNAJB4,EGR3, ENSA, GDA, GSTM1, KCNG1, KCNJ4, PHLDA3, PML, PPARD, PPARGC1A, PRKCB, PTPRV, SP7, SUCLG1, THRA, VDR
Muscle Upstream Regulator	**Molecule Type**	**p-value of overlap**	**# molecules affected**	**IPA Molecules in Dataset**
beta-estradiol	chemical - endogenous mammalian	1.79E-09	41	AARS1,ACADVL,ACE,AFF4,AHR,AKAP10,ANXA4,APBA1,APOA1,APOE, ATP5F1B, CA14, CHRM3, CNR1, CUBN, BYP24A1, DNAJC1, ENEO1, GABRG3, IFI30, ILF3, IQGAP2, LUM, MYH10, OGDH, PAPSS2, PMM2, PPARGC1A, PRKCB, RDX, SERPINA1, SERPINA3, SLCA4, SNAP25, SPARC, TF, TP53, TPH2, TXRD1, VCL, VIM
TP53	transcription regulator	1E-08	34	ACADVL,ACE,ALDH1A1,ANXA4,ANXA6,APOA1,APOE, CALU,CYP24A1,DNAJB4, FKBP4, IFI30, MYH10, MYL3, NDUFS3, OGDH, PCK1, PDHB, PDHX, PDP1, PFKFB1, PML, PPARD, PPARGC1A, PRDX6, PRKCB, SDHA, SERPINA3, SLCA4, SP7, TP53, TUBB3, VCL, VIM
lipopolysaccharide	chemical drug	1.45E-05	32	AARS1,ACE,AHR,ALDH1A1,ANXA5,APOA1,APOE,CASR,CCT6A,CUBN, ENO1,EPHX2, IFI30, IFNAR2, IL15RA, MT1, MTMR14, MYL3, PAPSS2, PML, PPARD, PPARGC1A, PRKCB, SDHA, SERPINA1, SERPINA3, SPARC, TF, TP53, VCL, VIM, YWHAQ
tretinoin	chemical - endogenous mammalian	3.16E-06	27	AHR,ALDH1A1,ANXA5,ANXA6,APOE,CNR1,CNTNAP2, CUBN,HPX,ILF3, MAP2, MT1, PCK1,PML, PPARD, PPARGC1A, PRKCB, RAI14, SERPINA1, SLC2A4, SP7, SPARC, TF, TP53, TUBB3, VCL, VIM
dexamethasone	chemical drug	0.000388	27	ACE,ALDH1A1,APOA1,APOE,CYP24A1,ENO1,FKBP4,IFNAR2,ILF3,MAP2, MSTN, MT1, PCK1, PMM2, PPARD, PPARGC1A, PRDX6, PRKCB, SDHA, SERPINA1, SERPINA3, SLC2A4, SPARC, TF, TP53, VCL, VIM

Discussion

The greatest number of genetic molecules in the mouse genome reside on Chrms 2, 1, 16, 9, 11, and 13 (n= 61, 36, 36, 34, 33, and 22, respectively, Fig. 1). When standardized for length of chromosome, Chrms. 8, 15, 7, 18, and Y (n=20, 15, 20, 1, 1, respectively), have the highest density of genetic molecules (Fig. 1). However, given the multi-chromosomal interaction of some of the molecules (e.g., miRNA) as well as the unmapped epistatic molecules[41], it is unclear whether chromosomal density of these molecules is important. Interestingly, there were 74 genetic variants identified in humans that did not have homologous molecules in the mouse model, the majority on Chrm 18 (human). A major source of these molecules is De Moor and colleague's findings[44] that were associated with exercise participation. These non-homologous molecules may represent specific human-related mechanisms that regulate physical activity. To date, Kostrzewa, et al. is the only cross-species translation attempt in this literature[39]. Certainly, cross-species translation is an area that needs further investigation to determine if there are pathways specific to either the animal or human genetic control of physical activity (Fig. 2).

There are a variety of potential upstream regulators for the molecules collected in this dataset. Interestingly, three of the top five for both brain and muscle are nuclear receptor binding ligands; estradiol (estrogen receptor), dexamethasone (glucocorticoid receptor), and tretinoin (retinoic acid receptor), indicating a potential role of nuclear receptors in the regulation of physical activity. Unlike most intercellular messengers, these ligands can cross the plasma membrane and directly interact with nuclear receptors inside the cell verses being limited to cell surface interactions[70]. Activation of nuclear receptors subsequently causes transcription of genes controlling a variety of cellular processes like cell proliferation, development, metabolism, and reproduction; additionally, some nuclear receptors have also been found to regulate cellular functions within the cytoplasm (e.g. estrogens)[71]. Selective nuclear receptor modulation may be an area of interest for future studies.

The top ranked upstream regulator for both muscle and brain datasets identified in this review was β-estradiol. Since initial efforts in the 1920's[20], it has been well established that sex steroids – most prominently estrogen[19,21], and/or testosterone[17,18], markedly alter physical activity levels. For example, removal of endogenous sex steroids decreased wheel-running by 90% in male animals and 69% in female animals, with administration of exogenous testosterone recovering wheel running activity (103% of baseline in females, 90% of baseline in males)[18]. However, what has not been clear is the mechanism through which these sex steroids work; i.e., is the level of sex steroids – and thus activity – controlled by genetics, or is the effect of sex steroids merely as a 'biological influencer' [a detailed review of the mechanistic differences exists by Bjornstrom et al[72]]. There is evidential support for the capacity of estradiol to increase physical activity levels; however, the mechanisms directly linking estradiol to increased physical activity have not been tested. For example, both testosterone and 17β-estradiol have been shown to enhance insulin-stimulated glucose uptake in female myotubes and to increase palmitate oxidation in both male and female myotubes[73] suggesting a role for estradiol in regulating substrate availability in the muscle. Further, there is extensive literature[74] showing that expression of *Slc2a4* – also known as *Glut4* – is affected by estrogen triggered-mechanisms. Tsao, et al.[9] clearly showed that overexpression of *Glut4* increased wheel-running activity in mice, suggesting that any mechanism through which *Glut4* expression was increased, would like-wise increase physical activity levels. Muscle biopsies obtained from men supplemented with 17β-estradiol demonstrated elevated medium-chain acyl-CoA dehydrogenase (MCAD), PGC1α mRNA, and reduced micro-mRNA, miR-29b (predicted to regulate

PGC-1α) supporting enhanced fat oxidation during exercise[75]. Supporting central mechanisms of estradiol; Peterson, et al. demonstrated estradiol impacts behavioral and synaptic correlates of addiction in female rats through activation of cannabinoid 1 receptor (CB1R - CNR1 gene)[76]. Dubreucq, et al. has demonstrated decreased dopamine activity and subsequent decreased wheel running in CB1R deficient mice[16] supporting a link between estradiol, cannabinoid receptors, and dopamine levels. In short, given the large number of targeted muscle molecules that are regulated in some manner by β-estradiol in both the brain and muscle, it is probable that β-estradiol exerts its physical activity regulating effects through both peripheral and central mechanisms.

Another potential regulator, TP53, a nuclear transcription regulator – was 2nd and 5th ranked potential upstream regulator of the muscle and brain molecules, respectively. TP53 is most often classified as a tumor suppressor gene and is activated by a variety of cellular stresses including oxidative stress at which time it becomes a transcription regulator (40). It has recently been suggested that moderate intensity exercise stimulates the cancer protective functions of TP53 (41). Thus, it is likely that physical activity can cause TP53 to accumulate in the stressed cell and regulate the various target genes in the physical activity dataset. However, since TP53 is not upregulated until physical activity is undertaken, TP53's effects would be a product of exercise, rather than being a regulator of physical activity.

Limitations: There are several limitations that must be considered when interpreting the results of this systematic review. First, while we worked to be as inclusive as possible, it is probable that there are studies that are missing from this summary. Certainly, there is no one set of key terms used by all the authors in this field; some agreed-upon set of terms would help similar future efforts.

Second, during the time-span of the considered papers, ranging from publication in the early 2000s to 2020, the methodology employed in genetic investigations changed dramatically. One significant example is the switch from using microsatellites to SNP as genomic markers. While we made the best efforts to standardize measurements, such as using the most recent builds (at time of data analysis) of the murine and human genomes, there were some results that could not be standardized[42] and thus, were left out of the analysis. In a similar vein, the minimal epistatic data that are available[41,45] was left out of the analysis primarily because of uncertainty of how to classify these data within the parameters of the network analysis. Removing these data from the analysis should not be construed to indicate they are not important – indeed the epistatic

data are the only available data of their kind in the field – but rather are examples of the limits of our analysis paradigm.

Third, a significant limitation both in the composing of and interpretation of the analyses is that there is a tremendous amount of data available. While we recognize our limits for interpretation and the length limitations of a written study, we have provided not only the data-files used in the network analyses as supplemental information, but also, extensive supplementary tables to enable other investigators to add to the dataset and rerun the analysis, or to gather the interpretative-narrative in their own work. Fourth, it must be remembered that like much work in the genetic realm, the results are mostly associative with few causal studies available. While some of the data used in this study were causative[9,10,15,16,53], much of the data are associative, especially the genomic variants we identified. Given there have been past difficulties noted in the literature in regards to defining mechanisms based on associative findings – especially genomic variant designs[77] – inferences based on genomic variant data should be made cautiously. Further, a continuing limitation of designs focusing on genomic variation is the uncertainty of whether these variants translate into different functional molecules. The addition of proteomic, transcript, and microRNA data compliments the genomic variation data and can add support for those genomic variants. If anything, our results guide future studies that focus on causative mechanisms.

Last, the work is limited by the tools available. Even though tools such as IPA are powerful, these tools are still limited in their scope and the genetic molecules available in their databases. For example, between 40-45% of the target molecules identified from the literature are not identifiable in the IPA database and so it is possible that other critical networks have been missed in this report. Additionally, given the magnitude of potential targets of each miRNA, it is possible that IPA does not contain full network linkages for all the miRNAs in the dataset and as such may not be the best approach to investigating the potential role of miRNAs in regulating physical activity.

In conclusion, the current paper describes our approach to gather the significant physical activity genetic regulating molecules from the literature, and to integrate all the known possible data into common upstream regulators related to physical activity control. Given the established consensus that physical activity is regulated both at the central and peripheral levels, potential upstream regulators are presented for both central (e.g., brain) and peripheral (e.g., muscle) molecules. Based on these analyses, the upstream regulator

analyses may be a great tool for future researchers to build research questions and investigate the broader

scope of mechanisms directly linking upstream regulators to regulation of physical activity.

Citations

1. Knab AM, Bowen RS, Hamilton AT, Gulledge AA, Lightfoot JT. Altered dopaminergic profiles: implications for the regulation of voluntary physical activity. *Behav Brain Res*. 2009;204(1):147-152. doi:10.1016/j.bbr.2009.05.034

2. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. *Public Health Rep*. 1985;100(2):126-131.

3. Lightfoot JT, De Geus EJC, Booth FW, et al. Biological / Genetic Regulation of Physical Activity Level: Consensus from GenBioPAC. *Med Sci Sports Exerc*. 2018;50(4):863-873. doi:10.1249/MSS.0000000000001499

4. Joosen AM, Gielen M, R V, Westerterp KR. Genetic analysis of physical activity in twins. *Am J Clin Nutr*. 2005;82(6):82 6 1253.

5. Perusse L, Tremblay A, C L, Bouchard C. Genetic and environmental influences on level of habitual physical activity and exercise participation. *Am J Epidemiol*. 1989;129:1012-1022.

6. KNAB AM, BOWEN RS, HAMILTON AT, LIGHTFOOT JT. PHARMACOLOGICAL MANIPULATION OF THE DOPAMINERGIC SYSTEM AFFECTS WHEEL-RUNNING ACTIVITY IN DIFFERENTIALLY ACTIVE MICE. *J Biol Regul Homeost Agents*. 2012;26(1):119-129.

7. Rhodes JS, Garland T. Differential sensitivity to acute administration of Ritalin, apormorphine, SCH 23390, but not raclopride in mice selectively bred for hyperactive wheel-running behavior. *Psychopharmacology*. 2003;167(3):242-250. doi:10.1007/s00213-003-1399-9

8. Rhodes JS, Hosack GR, Girard I, Kelley AE, Mitchell GS, Garland T. Differential sensitivity to acute administration of cocaine, GBR 12909, and fluoxetine in mice selectively bred for hyperactive wheel-running behavior. *Psychopharmacology (Berl)*. 2001;158(2):120-131. doi:10.1007/s002130100857

9. Tsao TS, Li J, Chang KS, et al. Metabolic adaptations in skeletal muscle overexpressing GLUT4: effects on muscle and physical activity. *FASEB J*. 2001;15(6):958-969. doi:10.1096/fj.00-0381

10. Ferguson DP, Dangott LJ, Schmitt EE, Vellers HL, Lightfoot JT. Differential skeletal muscle proteome of high- and low-active mice. *Journal of Applied Physiology*. 2014;116(8):1057-1067. doi:10.1152/japplphysiol.00911.2013

11. Rendeiro C, Masnik AM, JG M. Fructose decreases physical activity and increases body fat without affecting hippocampal neurogenesis and learning relative to an isocaloric glucose diet. *Sci Rep*. 2015;5(9589):10 1038 09589.

12. Vellers HL, Letsinger AC, Walker NR, JZ G, Lightfoot JT. High fat high sugar diet reduces voluntary wheel running in mice independent of sex hormone involvement. *Front Physiol*. 2017;8(628):10 3389 2017 00628.

13. Kelly SA, Nehrenberg DL, Hua K, Garland T, Pomp D. Functional genomic architecture of predisposition to voluntary exercise in mice: expression QTL in the brain. *Genetics*. 2012;191(2):643-654. doi:10.1534/genetics.112.140509

14. Ferguson DP, Schmitt EE, Lightfoot JT. Vivo-Morpholinos Induced Transient Knockdown of Physical Activity Related Proteins. *PLoS One*. 2013;8(4):e61472. doi:10.1371/journal.pone.0061472

15.   Ferguson DP, Dangott LJ, Vellers HL, Schmitt EE, Lightfoot JT. Differential protein expression in the nucleus accumbens of high and low active mice. *Behavioural Brain Research*. 2015;291:283-288. doi:10.1016/j.bbr.2015.05.035

16.   Dubreucq S, Durand A, I M. Ventral tegmental area cannabinoid type-1 receptors control voluntary exercise performance. *Biol Psychiatry*. 2012;73(9):10 1016 2012 10 025.

17.   Bowen RS, Ferguson DP, Lightfoot JT. Effects of Aromatase Inhibition on the Physical Activity Levels of Male Mice. *J Steroids Horm Sci*. 2011;1(1):1-7. doi:10.4172/2157-7536.S1-001

18.   Bowen RS, Knab AM, Hamilton AT, McCall JR, Moore-Harrison TL, Lightfoot JT. Effects of Supraphysiological Doses of Sex Steroids on Wheel Running Activity in Mice. *J Steroids Horm Sci*. 2012;3(2):110. doi:10.4172/2157-7536.1000110

19.   Gorzek JF, Hendrickson KC, Forstner JP, RJ L, AL M, Lowe DA. Estradiol and tamoxifen reverse ovariectomy-induced physical inactivity in mice. *Med Sci Sports Exerc*. 2007;39(2):248-257.

20.   Richter CP. Animal behavior and internal drives. *Q Rev Biol*. 1927;2(3):307-343.

21.   EJ R, GS W. Role of estrogens in androgen-induced spontaneous activity in male rats. *J Comp Physiol Psychol*. 1975;89(6):573-579.

22.   Misra BB, Langefeld CD, M O, Cox LA. Integrated omics: Tools, advances, and future approaches. *J Mol Endocrinol*. Published online 2018:10 1530-18-0055.

23.   Dawes MJ, Moore-Harrison T, Hamilton AT, et al. Differential gene expression in high- and low-active inbred mice. *Biomed Res Int*. 2014;2014(361048):10 1155 2014 361048.

24.   Dawes MJ, Kochan KJ, PK R, Lightfoot JT. Differential miRNA expression between inherently high- and low-active mice. *Physiol Rpts*. 2015;3(7).

25.   Lightfoot JT, Turner MJ, Pomp D, Kleeberger SR, Leamy LJ. Quantitative trait loci for physical activity traits in mice. *Physiol Genomics*. 2008;32(3):401-408. doi:10.1152/physiolgenomics.00241.2007

26.   Lightfoot JT. Exercise Genomics. LS P, SM R, eds. *The genomic regulation of physical activity*. Published online 2011:45-72.

27.   Sayers EW BJ Bolton EE et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res*. 2022;50(D1):10 1093 1112.

28.   *National Center for Biotechnology Information*. GRCH38.p13; 2019.

29.   *National Center for Biotechnology Information*. GRCM39; 2020.

30.   Cunningham F, Allen JE, J A. Ensembl 2022. *Nucleic Acids Res*. 2022;50(D1):10 1093 1049.

31.   Manzoni C, Kia DA, Vandrovcova J, et al. Genome, transcriptome and proteome: The rise of omics data and their integration in biomedical sciences. *Brief Bioinform*. 2018;19(2):10 1093 114.

32.   Krämer A, Green J, J P, Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics*. 2014;30(4):10 1093 703.

33.   Bruneau M, Walsh S, Selinsky E, et al. A genetic variant in IL-15Rα correlates with physical activity among European–American adults. *Mol Genet Genomic Med*. 2018;6(3):401-408. doi:10.1002/mgg3.368

34.   Keeney BK, Meek TH, Middleton KM, LF H, Garland TJ. Sex differences in cannabinoid receptor-1 (CB1) pharmacology in mice selectively bred for high voluntary wheel-running behavior. *Pharmacol Biochem Behav*. 2012;101(4):10 1016 2012 02 017.

35.   Caetano-Anollés K, Rhodes JS, Garland T, et al. Cerebellum Transcriptome of Mice Bred for High Voluntary Activity Offers Insights into Locomotor Control and Reward-Dependent Behaviors. *PLoS One*. 2016;11(11):e0167095. doi:10.1371/journal.pone.0167095

36.   Klimentidis YC, Raichlen DA, J B. Genome-wide association study of habitual physical activity in over 377,000 UK Biobank participants identifies multiple variants including CADM2 and APOE. *Int J Obes (Lond*. 2018;42(6):10 1038 41366-018-0120-0123.

37.   Cai G, Cole SA, N B. A quantitative trait locus on chromosome 18q for physical activity and dietary intake in Hispanic children. *Obesity*. 2006;14(9):1596-1604. doi:10.1038/oby.2006.184.

38.   Comuzzie AG, Cole SA, Laston SL, et al. Novel Genetic Loci Identified for the Pathophysiology of Childhood Obesity in the Hispanic Population. *PLoS One*. 2012;7(12):e51954. doi:10.1371/journal.pone.0051954

39.   Kostrzewa E, Brandys MK, HA L, Kas MJ. A candidate syntenic genetic locus is associated with voluntary exercise levels in mice and humans. *Behav Brain Res*. 2014;10(1016/j.bbr.2014.05.006).

40.   Kumar KG, Dicarlo LM, Volaufova J, AR Z, Richards BK. Increased physical activity cosegregates with higher intake of carbohydrate and total calories in a subcongenic mouse strain. *Mamm Genome*. 2010;21(1-2):10 1007 00335-009-9243-0.

41.   Leamy LJ, D P, Lightfoot JT. An epistatic genetic basis for physical activity traits in mice. *J Hered*. 2008;99(6):10 1093 045.

42.   De Moor MHM, Posthuma D, Hottenga JJ, Willemsen G, Boomsma DI, De Geus EJC. Genome-wide linkage scan for exercise participation in Dutch sibling pairs. *Eur J Hum Genet*. 2007;15(12):1252-1259. doi:10.1038/sj.ejhg.5201907

43.   Leamy LJ, Pomp D, Lightfoot JT. A search for quantitative trait loci controlling within-individual variation of physical activity traits in mice. *BMC Genet*. 2010;11(1):83. doi:10.1186/1471-2156-11-83

44.   De Moor MHM, Liu YJ, Boomsma DI, et al. Genome-wide association study of exercise behavior in Dutch and American adults. *Med Sci Sports Exerc*. 2009;41(10):1887-1895. doi:10.1249/MSS.0b013e3181a2f646

45.   Leamy LJ, D P, Lightfoot JT. Epistatic interactions of genes influence within-individual variation of physical activity traits in mice. *Genetica*. 2011;139(6):10 1007 10709-011-9586-9589.

46.   Doherty A, Smith-Byrne K, Ferreira T, et al. GWAS identifies 14 loci for device-measured physical activity and sleep duration. *Nat Commun*. 2018;9(1):5257. doi:10.1038/s41467-018-07743-4

47.   Lightfoot JT, Leamy L, Pomp D, et al. Strain screen and haplotype association mapping of wheel running in inbred mouse strains. *J Appl Physiol (1985)*. 2010;109(3):623-634. doi:10.1152/japplphysiol.00525.2010

48.   Eisener-Dorman AF, Grabowski-Boase L, Steffy BM, T W, Tarantino LM. Quantitative trait locus and haplotype mapping in closely related inbred strains identifies a locus for open field behavior. *Mamm Genome*. 2010;21(5-6):10 1007 00335-010-9260-.

49.   Lin X, Chan KK, YT H. Genetic determinants for leisure-time physical activity. *Med Sci Sports Exerc*. 2018;50(8):10 1249 0000000000001607.

50.   Mäestu J, Lätt E, Rääsk T, et al. Ace I/D polymorphism is associated with habitual physical activity in pubertal boys. *J Physiol Sci*. 2013;63(6):427-434. doi:10.1007/s12576-013-0280-9

51.   Nehrenberg DL, Wang S, Hannon RM, Garland T, Pomp D. QTL underlying voluntary exercise in mice: interactions with the "mini muscle" locus and sex. *J Hered*. 2010;101(1):42-53. doi:10.1093/jhered/esp066

52.   Furuse T, Takano-Shimizu T, Moriwaki K, T S, Koide T. QTL analyses of spontaneous activity by using mouse strains from Mishima battery. *Mamm Genome*. 2002;13(8):10 1007 00335-002-2168-5.

53.   Pistilli EE, Bogdanovich S, F G. Loss of Il-15 receptor alpha alters the endurance, fatigability, and metabolic characteristics of mouse fast skeletal muscles. *J Clin Invest*. 2011;121(8):10 1172 44945.

54.   Gielen M, Westerterp-Plantenga MS, Bouwman FG, et al. Heritability and genetic etiology of habitual physical activity: a twin study with objective measures. *Genes Nutr*. 2014;9(4):415. doi:10.1007/s12263-014-0415-5

55.   Saul MC, Majdak P, Perez S, Reilly M, Garland T, Rhodes JS. High motivation for exercise is associated with altered chromatin regulators of monoamine receptor gene expression in the striatum of selectively bred mice. *Genes Brain Behav*. 2017;16(3):328-341. doi:10.1111/gbb.12347

56.   Hartmann J, Garland T Jr, Hannon RM, Kelly SA, G M, Pomp D. Fine mapping of "mini-muscle," a recessive mutation causing reduced hindlimb muscle mass in mice. *J Hered*. 2008;99(6):679-687.

57.   Simonen RL, Rankinen T, L P. A dopamine D2 receptor gene polymorphism and physical activity in two families. *Physiology and Behavior*. 2003;78(4-5):751-757.

58.   Hillis DA, Yadgary L, GM W. Genetic basis of aerobically supported voluntary exercise: Results from a selection experiment with house mice. *Genetics*. 2020;216(3):10 1534 120 303668.

59.   Stefan N, Vozarova B, A DP. The GLN223ARG polymorphism of the leptin receptor in pima Indians: Influence on energy expenditure, physical activity and lipid metabolism. *Int J Obes Relat Metab Disord*. 2002;26(12):10 1038 0802161.

60.   Jean A, Laurent L, J B. The nucleus accumbens 5-HTR4-cart pathway ties anorexia to hyperactivity. *Transl Psychiatry*. 2012;2:e203:10 1038 2012 131.

61.   Kas MJ, Mooij-Van Malsen JG, M K. High-resolution genetic mapping of mammalian motor activity levels in mice. *Genes Brain Behav*. 2009;8(1):10 1111 1601-183 2008 00435.

62.   Umemori J, Nishi A, Lionikas A, et al. QTL analyses of temporal and intensity components of home-cage activity in KJR and C57Bl/6J strains. *BMC Genet*. 2009;10(40):10 1186 1471-2156-10-40.

63.   Kelly SA, Nehrenberg DL, JL P. Genetic architecture of voluntary exercise in an advanced intercross line of mice. *Physiol Genomics*. 2010;42(2):10 1152 00028 2010.

64.   Wilkinson AV, Gabriel KP, Wang J, et al. Sensation-seeking genes and physical activity in youth. *Genes Brain Behav*. 2013;12(2):181-188. doi:10.1111/gbb.12006

65.   Williams EG, Mouchiroud L, Frochaux M, et al. An evolutionarily conserved role for the aryl hydrocarbon receptor in the regulation of movement. *PLoS Genet*. 2014;10(9):e1004673. doi:10.1371/journal.pgen.1004673

66.   Kelly SA, Nehrenberg DL, Hua K, Garland T, Pomp D. Quantitative genomics of voluntary exercise in mice: transcriptional analysis and mapping of expression QTL in muscle. *Physiol Genomics*. 2014;46(16):593-601. doi:10.1152/physiolgenomics.00023.2014

67.   Yang HS, Vitaterna MH, Laposky AD, K S, Turek FW. Genetic analysis of daily physical activity using a mouse chromosome substitution strain. *Physiol Genomics*. 2009;39(1):00066 2009.

68.   Keeney BK, Raichlen DA, Meek TH, et al. Differential response to a selective cannabinoid receptor antagonist (SR141716: Rimonabant) in female mice from lines selectively bred for high voluntary wheel-running behaviour. *Behav Pharmacol*. 2008;19(8):10 1097 0 013 32831 3 6.

69.   Yang HS, Shimomura K, Vitaterna MH, Turek FW. High-resolution mapping of a novel genetic locus regulating voluntary physical activity in mice. *Genes, Brain and Behavior*. 2012;11(1):113-124. doi:10.1111/j.1601-183X.2011.00737.x

70.   Jin L, Li Y. Structural and functional insights into nuclear receptor signaling. *Advanced Drug Delivery Reviews*. 2010;62(13):1218-1226. doi:10.1016/j.addr.2010.08.007

71.   Sever R, Glass CK. Signaling by Nuclear Receptors. *Cold Spring Harb Perspect Biol*. 2013;5(3):a016709. doi:10.1101/cshperspect.a016709

72.   Björnström L, Sjöberg M. Mechanisms of Estrogen Receptor Signaling: Convergence of Genomic and Nongenomic Actions on Target Genes. *Molecular Endocrinology*. 2005;19(4):833-842. doi:10.1210/me.2004-0486

73.   Salehzadeh F, Rune A, M O, Al-Khalili L. Testosterone or 17beta-estradiol exposure reveals sex-specific effects on glucose and lipid metabolism in human myotubes. *J Endocrinol*. 2011;210(2):10 1530-10-0497.

74.   Gregorio KCR, CP L, Machado UF. Estrogen and glycemic homeostasis: The fundamental role of nuclear estrogen receptors ESR1/ESR2 in glucose transporter GLUT4 regulation. *Cells*. 2021;10(1):10 3390 10010099.

75.   Maher AC, Akhtar M, Tarnopolsky MA. Men supplemented with 17beta-estradiol have increased beta-oxidation capacity in skeletal muscle. *Physiol Genomics*. 2010;42(3):342-347. doi:10.1152/physiolgenomics.00016.2010

76.   Peterson BM, Martinez LA, Meisel RL, Mermelstein PG. Estradiol impacts the endocannabinoid system in female rats to influence behavioral and structural responses to cocaine. *Neuropharmacology*. 2016;110(Pt A):118-124. doi:10.1016/j.neuropharm.2016.06.002

77.   J F, TF M. Genetic architecture of quantitative traits in mice, flies, and humans. *Genome Res*. 2009;19(5):723-733.

Table Legend

Table 1: Citations reporting genetic molecules associated with/causative of physical activity

Table 2: Top 5 Upstream Regulators for Brain and Muscle Subsets

Figure Legends

Figure 1 – Map of Mouse/Human Genetic Molecules Associated With/Causal for Physical Activity.

Figure 2 – Map of Genetic Molecules Associated with Physical Activity.

Supplemental Content:

Supplemental Table 1: All molecules associated with / causative of physical activity.

Supplemental Table 2: Mapped molecules in database.

Supplemental Table 3: All upstream regulators of brain molecules

Supplemental Table 4: All upstream regulators of muscle molecules

Data Files:

Supplemental Content Data File 1: All molecules associated with / causative of physical activity in the brain (.csv).

Supplemental Content Data File 2: All molecules associated with / causative of physical activity in the muscle (.csv).

Figure 1 – Map of Mouse/Human Genetic Molecules Associated With/Causal for Physical Activity

Figure 2 – Map of Genetic Molecules Associated With Physical Activity – Human Genome