Genetic Analysis of Spontaneous Walking Behavior in *Drosophila melanogaster*

Presented to the faculty of Lycoming College in partial fulfillment of the requirements for Departmental Honors in Biology

By
Sean P. McLaughlin
Lycoming College
December 2018

Approved by:

(Dr. Sarah E. Holstein)

(Dr. Chriss E. McDonald)

(Dr. David Broussard)

(Dr. David R. Andrew)

This work was supported by a Joanna and Arthur Haberberger Fellowship, awarded to Sean P. McLaughlin at Lycoming College, Williamsport, PA.

Table of Contents

Abstract

The *Drosophila* Genetic Reference Panel (DGRP) has proven to be a useful tool in developing a quantitative genome for the fruit fly *Drosophila melanogaster*. The development of the DGRP simplified the process of identifying genes associated with quantitative traits, such as behaviors, that result from the cumulative effect of multiple genes. The process of identifying the genes underlying behavior relies on collecting large amounts of data that can takes hours of tedious work to manually analyze. Automated behavioral analysis that employs computer software can dramatically speed up this process and increase analytical throughput. Our study assessed the capabilities of the video-tracking software Ethovision XT® to accurately extract fly behaviors from video data of spontaneously-behaving flies. We proposed that Ethovision XT® is able to extract behavior data in a manner comparable to more laborious manually-extracted data. We applied the automated tracking analysis to previously-recorded videos from a subset of the DGRP stocks and found a continuous variation in spontaneous walking behavior among the populations. These phenotypic data were submitted to the DGRP genome wide association study (GWAS) online analysis pipeline, which identified several genetic polymorphisms of interest for walking behavior. The candidate genes in the regions of these polymorphisms have previously been implicated in synaptic function of the central nervous system and synaptic regulation at the neuromuscular junction. To further test the influence of these genes on the spontaneous walking behavior, we used a reverse genetics technique, RNA inference (RNAi), to reduce expression of the candidate genes and quantified the effects on walking behavior using Ethovision XT®. Results from the RNAi experiments indicate that expression variation of candidate genes does indeed influence spontaneous walking behavior, thereby supporting the results of the GWAS. Future studies will elaborate the role of these genes on variations in spontaneous behavior.

Background

Genome wide association studies (GWAS) are useful tools for modern medicine since they are capable of identifying genes that cause complex genetic diseases, and can in turn lead to more directed treatments (Bush and Moore 2012). Most GWAS studies focus on determining the genes for qualitative phenotypes, such as disease states, while neglecting regions of the genome associated with quantitative traits (Mackay *et al.* 2009). Unlike qualitative traits, which are defined by distinctive characteristics and generally determined by no more than a few genes, quantitative traits show a wide variety within a population, producing a continuous phenotypic distribution influenced by numerous genes. Because they are influenced by small effects of many genes, it is notoriously difficult to elucidate the genetic underpinnings of quantitative traits. One efficient way to study quantitative traits in the fruit fly, *Drosophila melanogaster*, is via the *Drosophila* Genetic Reference Panel (DGRP) (Mackay *et al.* 2012). This resource was created to enable the fine-scale mapping of quantitative trait loci (QTL), which are regions in the genome associated with the variable expression of these traits. The creators of the DGRP sequenced the entire genomes of more than 200 genetically distinct and isogenized *Drosophila melanogaster* lines. These different lines were originally from a single wild-type population, and using genetic techniques were back crossed in such a manner that each line constitutes a unique combination of the standing genetic diversity of the original population. The genomic sequences and genetic polymorphisms of each DGRP line are publicly available, and online tools allow researchers to associate the variation in a quantitative trait with particular genomic loci (Mackay *et al.* 2012). This study will investigate the QTLs associated with spontaneous walking behavior in *Drosophila* to better understand genes that may underlie natural variation in this trait.

The utility of the DGRP lines in determining the genetics of complicated quantitative phenotypes has been verified in several studies, which illustrates the efficacy of GWAS analyses that employ the DGRP. For instance, one study analyzed the behavioral responses to different odorants in *Drosophila* (Arya *et al.* 2015). With the use of over 250,000 flies from 186 different DGRP lines, they were able to identify behavioral responses for 14 chemically diverse odorants. Multiple candidate genes were identified through this analysis, including *Obp99d*, which is involved in olfactory behavior for benzaldehyde (Arya *et al.* 2015). Other candidate genes for complex behavioral phenotypes, such as courtship and mating behaviors, have also been identified through the use of the DGRP (Gaertner *et al.* 2015). Strong gene candidates from this study including *Ser* and *Fur1*, which are implicated in causally influencing a male's decision to stop copulation (Gaertner *et al.* 2015). Studies using DGRP also have translational implications for human-based research. For instance, Weber *et al*. (2012) used 167 lines of DGRP flies to determine novel gene candidates associated with variation in resistance to oxidative stress. There were 17 candidate genes identified in this GWAS analysis, and out of those 17 candidates, 12 have human orthologs, while 10 are further implicated in human diseases (Weber *et al.* 2012). With these studies demonstrating the utility of DGRP's ability to create a better understanding of the genetic basis for quantitative phenotypes, this current study will utilize the DGRP to determine genes that influence spontaneous walking behavior.

One of the benefits of the DGRP is its flexibility in applying large-scale genomic analyses to novel quantitative phenotype. First, a quantitative trait must be identified, defined, and characterized. After the data are collected, the arithmetic means of the behavioral responses for both males and females of each DGRP lines are submitted to the DGRP website to produce a GWAS. This GWAS allows for the identification of single nucleotide polymorphisms (SNPs)

associated with the variation in the phenotype (Huang *et al.* 2016, Mackay *et al.* 2012). With the information provided by the GWAS, researchers can then identify causative genes based on the identified SNPs and their genomic neighborhood. The influence of these genes on the particular phenotype can then be directly tested through reverse genetic techniques such as RNA interference (RNAi) or the CRISPR-Cas9 system (Huang *et al.* 2016, MacKay *et al.* 2012). Processes like these can give a better understanding of the impact from these genes on particular phenotypes, and potentially elucidate some behavioral functions of otherwise uncharacterized genes.

Before being able to identify gene candidates from the GWAS, large amounts of phenotypic data must be obtained. From each study that uses the DGRP, over 100 DGRP lines are used, with thousands of individual flies being screened (Arya *et al.* 2015, Gaertner *et al.* 2015, Weber *et al.* 2012). Analyzing such a large number of flies, particularly with complex behavioral data, can take countless hours of sifting through one behavioral video after another to quantify the desired trait. With this required high throughput of data, an automated tracking system is necessary for studies to be completed in a reasonable time. Ethovision XT® is a well-developed, commercially available tracking software shown to be capable of tracking several different modalities of fly behavior (Kaur *et al.* 2015, Martin 2004, Noldus Information Technology, Wageningen, The Netherlands). These modalities consist of average velocity, total distance traveled, mobility, movement, turn angle and meander. Ethovision XT® is capable of extracting these different modalities through tracking the animal by the center point of its body or by pixel change to determine total distance and time. Using these two sets of data, modalities such as velocity can then be calculated. Although studies have set parameters for extracting data on these modalities, the parameters need to be validated for efficiency based on aspects of the

experimenters' data collection and extraction process. By defining the thresholds based on collected data, Ethovision XT® will be able to extract data on a fly's activity.

Being able to identify causative genes for walking behavior will bring a better understanding into what molecular mechanisms control this highly variable trait. What is known about walking behavior is the location in the fly brain where it is controlled, namely two regions called the central complex and the mushroom bodies (Martin *et al.* 1999). These are implicated in maintaining a state of high probability of walking, but not necessarily in initiating walking bouts. Specifically, two neural circuits that control motor behaviors have been identified in these structures. These circuits comprise of columnar neurons connecting different portions of the central complex, the protocerebral bridge, the ellipsoid body and the noduli, as well as several large-field tangential neurons in the upper part of the fan-shaped body that extends outside of the central complex into the α'-lobe of the mushroom body (Martin *et al.* 1999). Later studies demonstrate that the central complex is associated with higher locomotor control in larvae and adult *Drosophila melanogaster* (Wan *et al.* 2000). Alterations or mutations in the central complex leads to slower moving or less active flies. This is thought to be due to the central complex being the crosslink between the right and left hemispheres allowing for communication to increase swing speed of legs on alternating sides (Wan *et al.* 2000). Along with the location of walking in the brain, a particular gene has also been implicated with spontaneous locomotion. Mutation of the gene *highwire* (*hiw*) shows a defect in adult walking behavior indicating its influence in spontaneous locomotion (Strauss 2002). Molecularly it is a negative regulator of synaptic growth and its mRNA and protein are expressed in most synapses (Strauss 2002).

To further verify the results of GWAS analysis, identified genes of interests are often analyzed using reverse genetics. A common method of reverse genetics is RNA interference

(RNAi), which allows for the knockdown of expression for a gene of interest (GOI). This process occurs by inducing the expression of a double stranded RNA (dsRNA), which then gets processed into small inferring RNA (siRNA) by the enzyme Dicer. Recognition, cleavage, and subsequent silencing of endogenous messenger RNA (mRNA) occurs when the siRNAs act as a template for the RNA-induced silencing complex (RISC), which precisely recognizes target mRNA (Yamamoto-Hino and Goto 2013). Unlike other genetic techniques, this technique allows for the knockdown of a certain gene without changing genomic sequences. Since RNAi focuses on silencing gene expression through the gene products, it allows for the study of more genes in a shorter period of time. RNAi experiments can be carried out *in vivo* for *Drosophila* through the use of the *GAL4/UAS* binary expression system. To carry out this technique, a fly containing the transgenic yeast transcription factor *GAL4* is crossed with a fly containing *GAL4*'s corresponding upstream activator sequence (*UAS*) adjacent to a reporter gene of interest. The progeny of this

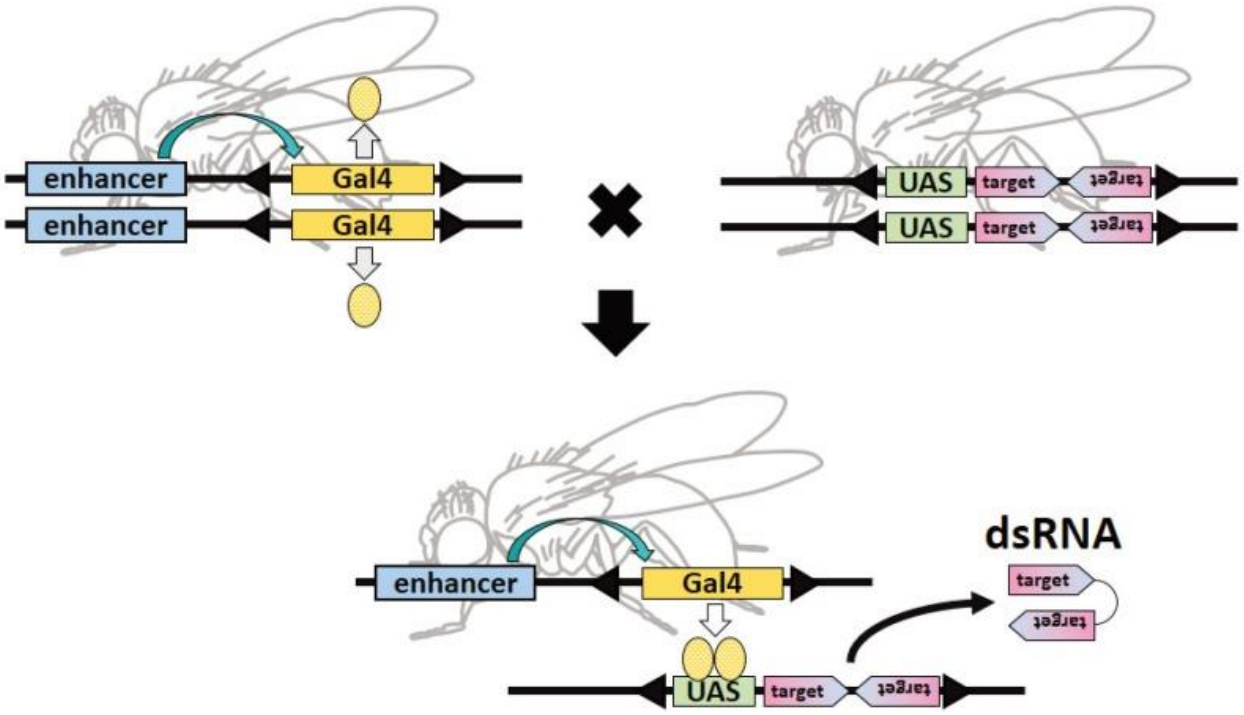

Figure 1. Demonstration of GAL4 protein and UAS interacting to induce the expression of double strand DNA (dsDNA) hairpin for RNAi. Figure modified from Yamamoto-Hino and Goto (2013).

cross will show expression of the *GAL4* transcription factor, which will bind to the *UAS* and induce expression of the *GOI* (Figure 1). Depending on the *GAL4* line, *GAL4* expression will vary predictably at different developmental times or in different tissues, thereby allowing for detailed analysis of resulting phenotypes. An easy way to identify where and when *GAL4* lines are expressed is the use of reporter genes such as green fluorescent protein (GFP). When *GAL4* lines are crossed with flies containing a *UAS*-GFP, the progeny will have the express the visual marker of GFP where the *GAL4* is expressed (Duffy 2002).

The molecular premise of *GAL4/UAS* RNAi experiments *in vivo* begins with the production of a hairpin *UAS*-dsRNA transcripts driven by *GAL4* expression (Kaya-Copur and Schnorrer 2016). This dsRNA is then cut into siRNA by the protein Dicer. These siRNA sequences are complementary to the GOI mRNA, so they can specifically recognize known mRNA sequences. Recognition of the mRNA sequences occur by the RISC identifying the mRNA based on the siRNA templates. Once recognized, the mRNA is cleaved and then degraded before a functional protein can be translated. Being able to reduce the protein expression of specific genes makes RNAi a useful tool for analyzing influential genes that have been implicated in DGRP studies.

Identifying genes that influence spontaneous walking behavior can lead to a better understanding of neurodevelopmental models in *Drosophila melanogaster*. Walking behavior has previously been shown to be a phenotype in fly models of stress, drug addiction, attention deficit disorder (ADHD) and autism (Kaun *et al*. 2012, Ki *et al.* 2017, Kaur *et. al.* 2015). Batsching *et al*. (2016) investigated inescapable stress, finding learned helplessness lowers activity in the fly. This inescapable stress was induced when escape responses were ineffective thus reducing the frequency of escape responses when a painful stimulus was present (Batsching

et al. 2016). Walking behavior is also used to study drug addiction since addictive drugs significantly change this behavior in flies. Cocaine has been shown to increase circling behavior and aberrant walking behavior while MDMA was shown to reduced locomotion in larvae (Kaun *et al.* 2012). Nicotine has also been shown to induce hyperactivity in flies (Kaun *et al.* 2012). Inducing ADHD symptoms from chemicals like imidacloprid also shows an increase movement in flies (Ki *et al.* 2017). Even chemicals linked to ADHD and autism such as bisphenol A, affect walking behavior by reducing walking speed (Kaur *et al.* 2015).

With walking behavior being implicated with these behavioral and neurodevelopmental disorders, a GWAS obtained from analyzing variation in walking behavior has the potential to lead to the identification of genes involved in regulating many aspects of the central and peripheral nervous system. We hypothesize that walking behavior is a quantitative behavior in fruit flies that will show a continuous variation among and between DGRP lines. We further hypothesize that there will be QTL associated with this behavior that we can identify as good candidates for further behavioral genetic analysis. We predict that targeted silencing of associated genes by RNAi will alter spontaneous walking behavior. Ultimately, these analyses seek to identify genetic influences of activity levels which could ultimately inform the field of behavioral neurogenetics in flies and increase their utility as models for studying human neurodevelopmental disorders.

Materials and Methods

Expansion and collection of DGRP flies

This experiment utilized a preliminary subset of the entire DGRP lines. These lines were quarantined upon arrival from Bloomington *Drosophila* Stock Center (BDSC) to prevent the spread of disease and/or mites to other Andrew lab fly stocks. After the quarantine time ensured the stocks were healthy, expansions populations were created by placing 5 virgin females and 10 males of the same line into standard polystyrene *Drosophila* vials containing around 8.5g of NutriFly™ Bloomington formulation food (Genesee Scientific, San Diego, CA). Flies were reared at 25℃ and 80% humidity in an incubator a 12-hour light/12-hour dark cycle. For ten straight days the 15 flies were transferred into a new vial to allow for synchronous populations. These standard conditions are optimal for rearing healthy populations of *Drosophila* (Greenspan, 2004).

Pupae were selected out of the stock vials after identifying the sex. The selected pupae were divided individually into a 16 x 100 mm sterile culture tube from Genesee Scientific containing around 2 g NutriFly™ Bloomington formulation food (Genesee Scientific, San Diego, CA). These tubes were uniquely labeled and the DGRP line, date, time of collection, and sex were recorded. These tubes were placed in the same incubator and monitored every one to two hours from 9:00 AM to 4:00 PM. In this way, each fly could be individually monitored in order to know the time of adult eclosion to within a 2-hour block. Those flies that eclosed between monitoring sessions, which we were confident of their age, were used for behavioral recordings 24-hours later. Flies with physical abnormalities, like deformed legs or wings that have not unfurled, were not selected for the video recordings. Flies that eclosed overnight were not selected for videos since the time of eclosion was not able to be recorded.

Data collection

Videos of the behavioral assay were initially collected for the analysis of grooming behavior from a previous study (Hannum 2017). This prior study focused solely on grooming behavior, but the video data set presented a good opportunity to test and validate automated scoring procedures for this current study. Later video recordings for this current study followed the same protocols, which are described below.

The video recordings of the selected flies occurred approximately 24 hours after adult eclosion was documented. These recordings took place in a behavior room which is climate-controlled to maintain approximately 25℃ temperature with 75% humidity. The flies were brought into the room at least one hour prior to recording to adjust to the recording environment. Recordings took place with the room lights off while having a circular LED desk light on (Model: Mic-2019) so the camera (Canon Vixia HF R72) could have even lighting for the video. Six flies were aspirated into individual wells of a plastic transparent U-bottom 96 well plate from Genesee Scientific for recording. The orientation for these flies in the 96 well plate were two rows by three columns. Each well had 200 µL of 1.5% agar to maintain humidity, provide a consistent substrate, and prevent vertical movement within the recording arena. Each fly was recorded with their DRGP line, sex, and well number in the Multi-Well Scoring template. Recordings on the scorer's initials, time, date, room conditions, and initial and final plate temperatures were documented in the Multi-Well Scoring template as well. The camera was placed directly over the wells and focused manually. The flies were recorded for at least ten minutes. Upon completion of a recording, flies were placed into 0.5 mL microcentrifuge tubes

(Fisherbrand®) labeled with the fly's specific identification number and kept in a -20℃ freezer for future genotyping.

Manual behavioral analysis and verification

Raw files of the videos were saved and backed up on a Dropbox account. The video files were then condensed and converted into an mp4 file format by Media Encoder from Adobe. Once the video was condensed and converted, the computer program VCode (Hagedorn et al., 2008) was used to manually score the videos. In order to build a validation set for automated scoring analysis (described below), we first scored videos from a previous data set (Hannum 2017) for walking behavior. Randomly-selected videos from the top three and bottom three grooming lines of this previous study (Hannum 2017) were used for comparative analysis against automated scores produced by the software program Ethovision XT® (hereafter referred to simply as Ethovision, Noldus Information Technology, Wageningen, The Netherlands). Videos were manually scored for the following activity states: grooming, falling, standing, and walking. Scoring these behaviors was done by pressing a specific key that corresponded with the behavior to mark the beginning and end of that behavior. The most complex behavior among these, grooming, was defined as any stroking of the head, antenna, proboscis, limbs, wings, or thorax by one or more limbs. Walking was defined as the animal moving by the use of their limbs. The behavior falling was defined as when the fly fell from the top of the well until it assumes a normal standing posture. Standing was defined as when the animal was not grooming, falling, or walking. The grooming results were compared statistically by interrater reliability to previous recordings by another recorder on the same videos and flies to ensure consistency and reliability of identification of behaviors.

Automated Behavioral Analysis using Ethovision

The condensed and converted mp4 videos were used in the Ethovision analysis. Parameters for analyzing the videos were set constant for the duration of the experiment including the trial control settings, advance parameter in the detections settings, experimental settings, arena settings, track smoothing profile, data profile and some aspects of analysis profile. Experimental settings were set as six for number of arenas, activity analysis on, and unit of distance as millimeters. Arena settings were set by calibrating the length of three wells to 26.2 mm for establishing the dimensions in the video file, drawing the arenas to encapsulate the wells each fly was in, and clicking the validate setup. The advance parameters in the detection settings were set as 0-113 for gray scale, erode first then dilate with erosion as 5 and dilation as 7, minimum detected pixels were set as 1419, and maximum detected pixels were set as 28252. These detection settings were established manually to allow for the software to only recognize the fly without recognizing background noise. The track smoothing profile was set on the smoothing setting.

The parameter used to identify walking was the "movement" variable. There were two parameters within movement that were manipulated to extract results that were comparable to manual scoring. One of the parameters was start and stop velocity which indicated when the software differentiated between the animal not moving and moving. The other parameter was averaging interval which helped recognize minute movements or generalize a behavior over a few seconds. Besides manipulating these variables for the software to identify the behavior of walking, every other parameter was set as previously described.

Ethovision Data Collection and Verification of Walking Behavior

The mp4 video files were uploaded to the Ethovision software for the extraction of walking behavior. For each video every individual arena was defined based on their size and location of each well in the frame of the video. Through trial and error of various software settings (described below), Ethovision was able to extract walking behavior with appropriate movement parameter settings. The first round of tests kept the averaging interval constant at 1 and changed the start stop velocity. The next tests then manipulated the averaging interval once finding that a start velocity of 0.71 mm/s and a stop velocity of 0.70 mm/s were most comparable to manual scores. Increasing the averaging interval increases the 95% confidence interval, thus leading us to us an averaging interval of 1 with a start velocity of 0.71 mm/s and a stop velocity of 0.70 mm/s. The start and stop velocity indicated that the fly was determined to be "moving" if it was moving faster than 0.71 mm/s and "not moving" if it was moving slower than 0.70 mm/s. Once all the arenas were defined and all the videos within one file, Ethovision was run with the computer power settings on until the program was finished, and data were recorded automatically.

Genome Wide Association Study

GWAS results were obtained by submitting the average walking indexes for males and females of each DGRP line to the DGRP website (http://dgrp2.gnets.ncsu.edu/). Within a few hours results were returned through email. The GWAS identified each single nucleotide polymorphism (SNP) in the entire genomic data set with a p-value associated with the phenotypic averages. Other information provided on the SNPs were their location, precise variation, type of genomic region, and gene name if available. Custom written Perl scripts were

used to input the SNP data to produce a Manhattan plot. The genes of interest from the GWAS were identified by the SNPs with a p-value lower than 5×10^{-5}. For the information on the genes of interest, the database FlyBase was used (flybase.org; release FB2018_05).

Data Analysis

Custom written Perl scripts allowed for the extraction of data from the scored videos from VCode. Data was collected on the percent of total time spent on walking (walking index, WI) for each fly. These WI's from manual VCode data were used to compare to Ethovision data from the same videos. This data was analyzed by using a geometric mean regression line with a 95% confidence interval in MATLAB (MathWorks®, Natick, MA, USA).

MATLAB was used to display the walking data into ordered box-and-whisker plots. These plots were ordered from lowest median of walking index to highest median of walking index. A nonparametric Wilcoxon Rank-Sum test was implemented in MATLAB to determine if there was a statistically significant difference between the highest and lowest lines.

RNAi Functional Validation

Based on the top associated SNPs from the GWAS, four candidate genes were chosen for RNAi experiments based on both their p-value and whether their function related to movement in *Drosophila* (Table 1). For *Drosophila in vivo* RNAi experiments, the *Gal4-UAS* system was utilized to knockdown gene expression in a tissue-specific manner to infer gene function based on time of development and region of the body. Pre-constructed *UAS*-hairpin fly lines for the selected candidate genes were obtained from the Transgenic RNAi project (TRiP-RNAi) fly stocks from Harvard Medical School collection (Hu *et al*. 2016). Based on specific

spatiotemporal expression, *GAL4* lines were obtained from BDSC. The three *GAL4* lines selected were *P{GawB}elav; P{Dcr}*, which has neuron-specific expression; *P{Dcr}; P{Act5c-GAL4}*, which has ubiquitous expression; and *P{Dcr}; P{GAL4-Mef2}*, which expresses only in muscle cells. Each *GAL4* line contained a *UAS-Dcr* so the protein Dicer will be overexpressed in RNAi experiments. Dicer is important for siRNA production from the dsRNA hairpins so an overexpression increases the amount of siRNA thus increasing the chance of a behavioral change. All flies were obtained through BDSC (Table 2) and quarantined to prevent the potential spread of mold and/or disease into existing fly stocks prior to the beginning of RNAi experiments.

Gene	Region of associated SNP	Basic Function	Average p-value
Naa35	Upstream	N-terminal protein amino acid acetylation	5.03E-06
DAT	UTR 5'	Dopamine Transporter	5.03E-06
Lmpt	Intron	Zinc ion binding	5.70E-06
Ndl	Synonymous coding	Serine-type peptidase	5.84E-06
Hiw	Synonymous coding/Intron	Zinc ion binding, synaptic growth at neuromuscular junction	1.83E-05

Table 1. Information on genes from the 5 top associated SNPs identified by the GWAS. Each gene candidate has at least one TRiP-RNAi transgene line available in the BDSC

BSDC Designation	Gene of Interest	Full Genotype	Short Genotype	Affected Chromosome
25750	N/A	P{w$^{+mW.hs}$=GawB}elav^{C155}w^{1118};P{w^{+mC}=UAS-Dcr-2.D}2	P{GawB}elav; P{Dcr}	1(X); 2
25756	N/A	P{w^{+mC}=UAS-Dcr-2.D}1, w^{1118}; P{w^{+mC}=GAL4-Mef2.R}R1	P{Dcr}; P{GAL4-Mef}	1(X); 3
28031	*Hiw*	y^{1}v^{1}; P{y$^{+7.7}$v$^{+1.8}$=TRiP.JF02866)attP2	P{TRiP.JF02866}	1(X); 3
28790	*Lmpt*	y^{1}v^{1}; P{y$^{+7.7}$v$^{+1.8}$=TRiP.JF03218)attP2	P{TRiP.JF03218}	1(X); 3
31256	*DAT* (Long)	y^{1}v^{1}; P{y$^{+7.7}$v$^{+1.8}$=TRiP.JF01197)attP2/TM3, Ser^{1}	P{TRiP.JF01197}	1 (X); 3
50619	*DAT* (Short)	y^{1}sc*v^{1}; P{y$^{+7.7}$v$^{+1.8}$=TRiP.HMC02986)attP2	P{TRiP.HMC02986}	1(X); 3
61901	*Naa35*	y^{1}v^{1}; P{y$^{+7.7}$v$^{+1.8}$=TRiP.HMJ23456)attP40	P{TRiP.HMJ23456}	1(X); 2

Table 2. Every *GAL4* driver lines and TRiP-RNAi transgene lines were obtained from BDSC. Short genotypes are used throughout the text.

RNAi Crosses

To obtain animals with specific mRNA knockdowns, crosses between the *GAL4* lines and the *TRiP-UAS* lines were necessary. Crosses between the *GAL4* lines and the *TRiP-UAS* lines were based on the health of the lines. Only *GAL4* and *TRiP* lines that produced an abundance of animals from both sexes were able to be used in crosses. Based off of the recommendations of Kaya-Copur and Schnorrer (2016), the crosses consisted of at least 5 *GAL4* virgin females with at least 10 *UAS-TRiP* males to produce offspring in 25 x95 mm polystyrene *Drosophila* vials (Genesee Scientific, San Diego, CA.). For validation of the influence of the RNAi process, crosses were also created with *GAL4* virgins and wild type (w^{1118}) males and wild type virgin females with the *UAS-TRiP* males using the same method stated above. To increase the number of offspring, expansions were created for each cross by transferring the animals in the cross to a new vial every two days for 14 days. Cross 1 crossed *P{GawB}elav; P{Dcr}* virgin females with *P{TRiP.JF02866}* males targeting the gene *hiw* (Table 3). For the negative controls for *hiw*,

cross 2 was virgin *w*^{1118} females crossed with *P{TRiP.JF02866}* (Table 4). Cross 3 focused on the gene *Naa35* by crossing *P{GawB}elav; P{Dcr}* virgin females with *P{TRiP.HMJ23456}* males (Table 5). Cross 4 was the negative control for the gene *Naa35* by crossing virgin *w*^{1118} females with *P{TRiP.HMJ23456}* males (Table 6). The last cross between *P{GawB}elav; P{Dcr}* virgin females and *w*^{1118} males was a negative control for the GAL4 line *elav* (Table 7).

	y^1,v^1; P{TRiP.JF02866}	>;P{TRiP.JF02866}
w^{1118}, P{GawB}elav; P{Dcr}	$\frac{w1118,P\{GawB\}elav}{y1,v1};\frac{P\{Dcr\}}{+};\frac{P\{TriP.JF02866\}}{+}$	$\frac{w1118,P\{GawB\}elav}{>};\frac{P\{Dcr\}}{+};\frac{P\{TriP.JF02866\}}{+}$
w^{1118}, P{GawB}elav; P{Dcr}	$\frac{w1118,P\{GawB\}elav}{y1,v1};\frac{P\{Dcr\}}{+};\frac{P\{TriP.JF02866\}}{+}$	$\frac{w1118,P\{GawB\}elav}{>};\frac{P\{Dcr\}}{+};\frac{P\{TriP.JF02866\}}{+}$

Table 3: Cross 1 - *w*^{1118}, *P{GawB}elav;P{Dcr}* virgin females x *y*^{1},*v*^{1}*; P{TRiP.JF02866}* males.

	y^1,v^1; P{TRiP.JF02866}	>;P{TRiP.JF02866}
w^{1118}	$\frac{w1118}{y1,v1};\frac{+}{+};\frac{P\{TriP.JF02866\}}{+}$	$\frac{w1118}{>};\frac{+}{+};\frac{P\{TriP.JF02866\}}{+}$
w^{1118}	$\frac{w1118}{y1,v1};\frac{+}{+};\frac{P\{TriP.JF02866\}}{+}$	$\frac{w1118}{>};\frac{+}{+};\frac{P\{TriP.JF02866\}}{+}$

Table 4: Cross 2 - *w*^{1118} virgin females x *y*^{1},*v*^{1}*; P{TRiP.JF02866}* males.

	y^1,v^1; P{TRiP.HMJ23456}	>;P{TRiP.HMJ23456}
w^{1118}, P{GawB}elav; P{Dcr}	$\frac{w1118,P\{GawB\}elav}{y1,v1};\frac{P\{Dcr\}}{+};\frac{P\{TriP.HMJ23456\}}{+}$	$\frac{w1118,P\{GawB\}elav}{>};\frac{P\{Dcr\}}{+};\frac{P\{TriP.HMJ23456\}}{+}$
w^{1118}, P{GawB}elav; P{Dcr}	$\frac{w1118,P\{GawB\}elav}{y1,v1};\frac{P\{Dcr\}}{+};\frac{P\{TriP.HMJ23456\}}{+}$	$\frac{w1118,P\{GawB\}elav}{>};\frac{P\{Dcr\}}{+};\frac{P\{TriP.HMJ23456\}}{+}$

Table 5: Cross 3 - *w*^{1118}, *P{GawB}elav;P{Dcr}* virgin females x *y*^{1},*v*^{1}*; P{TRiP.HMJ23456}* males

	y^1,v^1; P{TRiP.HMJ23456}	>;P{TRiP.HMJ23456}
w^{1118}	$\frac{w1118}{y1,v1};\frac{+}{+};\frac{P\{TriP.HMJ23456\}}{+}$	$\frac{w1118}{>};\frac{+}{+};\frac{P\{TriP.HMJ23456\}}{+}$
w^{1118}	$\frac{w1118}{y1,v1};\frac{+}{+};\frac{P\{TriP.HMJ23456\}}{+}$	$\frac{w1118}{>};\frac{+}{+};\frac{P\{TriP.HMJ23456\}}{+}$

Table 6: Cross 4 - w^{1118} virgin females x y^1,v^1; *P{TRiP.HMJ23456}* males

	w^{1118}	>
w^{1118}, P{GawB}elav;P{Dcr}	$\frac{w1118,P\{GawB\}elav}{w1118};\frac{P\{Dcr\}}{+};\frac{+}{+}$	$\frac{w1118,P\{GawB\}elav}{>};\frac{P\{Dcr\}}{+};\frac{+}{+}$
w^{1118}, P{GawB}elav;P{Dcr}	$\frac{w1118,P\{GawB\}elav}{w1118};\frac{P\{Dcr\}}{+};\frac{+}{+}$	$\frac{w1118,P\{GawB\}elav}{>};\frac{P\{Dcr\}}{+};\frac{+}{+}$

Table 7: Cross 5 - w^{1118}, *P{GawB}elav;P{Dcr}* virgin females x w^{1118} males

Behavior Recordings and Analysis of RNAi Knockdown lines

Collection for the RNAi knockdown crosses consisted of the same protocols as those previously described for the DGRP lines so that walking behavior could be compared for all experiments.

Video recording, data collection and behavioral analysis for the RNAi knockdown crosses followed the same protocols as those previously described for the DGRP lines so that walking behavior could be compared among experiments. Instead of using MATLAB to create the box-and whisker plots for the RNAi behavior visualization, violin plots created from the website http://shiny.chemgrid.org/boxplotr/.

Results

Validation of automated scoring methods

To validate the scorer's ability to correctly identify and quantify *Drosophila melanogaster* behaviors, the results from manually scoring 21 previously scored videos were compared. The statistical analysis used to compare between the scorers identification of grooming behavior was a geometric regression analysis with a significance level of 0.05. The slope of the line was 0.9333 with a 95% confidence interval of 0.8585 to 1.0147. The intercept of the line was -0.5003 with a 95% confidence interval of -0.9205 to -0.1137 (Figure 2). Both scores were found to be statistically similar and verified that the scorer is capable of identifying behaviors of *Drosophila melanogaster*.

Walking indexes between Ethovision-scored and manually-scored videos were compared to validate Ethovision's capabilities of scoring spontaneous walking behavior. For the analysis, 22 flies were used to compare Ethovision's scores against manually scored videos. A geometric mean regression analysis with a significance level of 0.05 was again used to statistically compare the two. Different averaging intervals and thresholds within the movement variable in Ethovision were tested to determine which setting in Ethovision could be comparable to manual scoring (Table 1 and Table 2). These variables were the only variables manipulated, and all other variables were maintained at the settings previously described. The best setting was Ethovision trial setting 7 which was set for a start velocity of 0.71 mm/s and a stop velocity of 0.70 mm/s (Table 2 and Figure 3). This setting had a slope of 1.0252 with an intercept of 0.0181. The intervals for the slope and intercept with their 95% confidence interval are shown in Table 2. When testing for different averaging intervals with a standard combination of start and stop velocity, the averaging with the smallest 95% confidence interval was an averaging interval of 1

(Table 1). The data found that Ethovision scores were not significantly different than manual scores implying that Ethovision can be used to score for spontaneous walking behavior during the rest of the experiment.

Genome Wide Association with Walking Behavior

The walking index data for each DGRP was represented by box-and-whisker plots. These box-and-whisker plots were oriented based on their median scores from the line with the lowest walking index median (DGRP 307) to the line with the highest walking index median (DGRP 852) (Figure 5). The highest line had a walking index significantly higher than the lowest line (Wilcoxon Rank-Sum test; p-value = 6.66×10^{-9}). Due to the statistical differences between the walking indexes of different DGRP lines, we can conclude that there is indeed, natural variation in spontaneous walking behavior present among the 35 lines of DGRP flies.

Once the mean walking indexes of both sexes from each DGRP line were computed, the data were sent to the DGRP GWAS analysis pipeline (Table 10). The p-values associated with each individual SNP in the genome regarding the walking behavior data was converted into a Manhattan plot (Figure 5), which shows the probability of every single SNP in relation to its location in the genome. The threshold for assessing if the SNP are putatively associated with spontaneous walking behavior in this experiment is defined as any p-value lower than 5×10^{-5}. A subset of relevant genes that fell below this threshold are shown in Table 3.

RNAi Knockdown of Target Gene Expression Altered Walking Behavior

Data from the spontaneous walking behavior of the two RNAi experiments were measured by Ethovision identifying the walking indexes of each cross. During the RNAi experiments, the collection of the w^{1118} x *P{GawB}elav; P{Dcr}* progeny did not follow protocol due to time restrictions. Flies were recorded past the 24-30 hour post-eclosion limit, and the

results from this control cross are therefore considered only as preliminary. For the *hiw* gene, there was a significant difference between the experimental RNAi knockdown progeny (P{TRiP.JF02866}> *P{GawB}elav; P{Dcr}*) and control flies (*P{TRiP.JF02866}>w*^{1118}) (Kruskal-Wallis test, *p*-value = 0.00155). This result indicated that the experimental crossed had a significantly lower walking index (Figure 6). When comparing the experimental progeny to the other *elav* driver control cross (*w*^{1118}*>P{GawB}elav; P{Dcr*}), there was no statistically significant difference (*p*-value = 0.77206).

The RNAi experiment on the gene *Naa35* showed statistical significant differences between the experimental progeny and both control groups. The experimental progeny (*P{TRiP.HMJ23456}>P{GawB}elav;P{Dcr*}) had a statistically higher walking index than both (*P{TRiP.HMJ23456*}>w^{1118}) (*p*-value = 0.01385) and the *GAL4* driver background line (*w*^{1118}*>P{GawB}elav;P{Dcr})* (*p*-value = 0.00035).

Discussion

Validating Automated Behavior Identification

In order to build a data set to use in benchmarking automated scoring procedure, we first set out to validate manually-scored videos from a previous study with the scorer of this current study. We did this in order to train the scorer in identifying behaviors and confirm the scorer's ability to identify various behavior. This was important for being able to identify whether the software was properly recognizing behaviors with different parameters. The geometric mean regression comparison resulted in a slope of 0.9333 and an intercept of -0.5003 (Figure 1). These data indicate both a minor proportional and systematic biases towards present manual scoring. The intercept displays a slight systematic bias which can be attributed to the differing methods in how the videos were analyzed between these two studies. The current study scored for four behaviors on one fly at a time with frame-by-frame resolution, while the previous study scored only grooming behaviors on up to three flies at a time in a less time-conscious manner. Not only does focusing on one fly tend to increase the grooming index, but the focused frame-by-frame analysis increases precision as well. These differing methods could explain the systematic bias towards manual scoring of the current study. Finding that the two scorers do not significantly differ in their behavioral assessments verifies the ability of this author to accurately assess behaviors in our assay.

Ethovision was utilized to extract data on spontaneous walking behavior in a scorer-independent way. Multiple combinations of parameters were initially manipulated and tested to determine which settings most accurately recapitulated the results of manual scoring in our test data videos. We were able to determine settings that ultimately displayed very little to no systematic or proportional biases when compared to manually-scored results (Figure 4). These

results supported the hypothesis that Ethovision is indeed capable of rapidly and accurately assessing walking behaviors in flies. Adjusting Ethovision to score for spontaneous walking behavior allowed the experiment to utilize this automated pipeline to score all 769 flies and 157 video files in a fraction of the time and with much less effort than prior studies. While these results are beneficial for extracting data on spontaneous walking behavior in flies, these results also show promise for Ethovision's capabilities in assessing more complex behaviors, such as grooming.

A Genetic Basis for Spontaneous Walking Behavior in Drosophila

From the settings previously described, Ethovision was able to extract data on walking from all 157 videos within 15 hours. Since the average time to manually score one fly was ~33 minutes, it would have taken ~423 hours to manually record all 769 flies. This amount of tedious manual data analysis would have taken months, thus verifying Ethovision's utility in increasing behavioral analysis throughput for flies.

The natural variation observed in spontaneous walking between populations in this study support the hypothesis that walking behavior in *Drosophila* is a quantitative trait. The differing distribution of walking indexes each line indicates that there is continuous variation among populations, which strongly suggests that there are genetic influences regulating variation in spontaneous walking (Figure 5). Moreover, our behavioral assay reduces environmental influences of behavior by strictly controlling handling and environmental conditions, further supporting the idea that genetic differences between line cause differences in behavior. With environmental factors held constant, the only known difference between the DGRP lines are

genetic. These differences in phenotypes can then be directly related to genetic variation with genome-wide association analysis.

Our GWAS results identified genes that influence spontaneous walking in flies. SNPs with a significant association with walking behavior were selected for further testing through RNAi. Besides the significance of the SNP, location and function were taken into consideration when choosing the SNPs for RNAi experiments. For our assessment, SNPs involved with known protein coding genes were investigated since our aim is to observe the effects of a knockdown in protein expression. The function of the SNP also had to be associated with the nervous system or muscular system since each system contributes to walking behavior. Based on this criteria, two genes were selected for RNAi testing.

The first gene candidate identified was *hiw* due to there being multiple significant hits in the region of this gene on the X chromosome (Figure 5). The gene *hiw* has a human ortholog known as MYC-binding protein 2 (MYCBP2) or PAM, which has high expression in axons and dendrites in the brain (OMIM #610392). Research on *hiw* indicates similar localization of expression in axons and dendrites of flies (Wan *et al.* 2000, Wu *et al*, 2005). *hiw* has been implicated to control the structural and functional development of synapses throughout larval development. In normal synaptic growth, *hiw* is exclusively found in the presynaptic cell (Wu *et al.* 2005). Besides localization of *hiw* in the brain, previous work has shown it to be influential in the neuromuscular junction (NMJ). Its role at the NMJ is to negatively regulate the BMP signaling cascade that is required for normal growth and function. This is demonstrated by mutants of *hiw* having more synaptic boutons at the NMJ (McCabe *et al.* 2004). The overgrowth of the synapse for the NMJ happens when autophagy occurs. Autophagy down-regulates *hiw* which leads to the development of the NMJ. Autophagy and *hiw* converge on the same Wnd-

dependent MAPK signaling pathway that regulates NMJ development (Shen and Ganetzky 2009). This interaction between autophagy and *hiw* can be prevented by the expression of *Rae1*. Rae1 is necessary to promote *hiw* abundance by protecting *hiw* from autophagy (Tian *et al.*, 2011). With all of the influence *hiw* has on the NMJ it is no surprise that *hiw* mutants have been implicated with adult walking defects (Wan *et al.* 2000). Thus, the knockdown of the mRNA product for such an influential gene in NMJ development could be predicted to reduce the spontaneous walking behavior in *Drosophila*. This is indeed the results we observed when we knocked down expression using RNAi as compared to the RNAi control line. These results warrant further analysis in future studies.

The other gene of focus in the experiment is *Naa35*, which had the lowest *p*-value in the GWAS results indicating the highest association with spontaneous walking behavior (Table 1). *Naa35* does not have a human ortholog, however; *Naa38* and *Naa30* are human genes that interact with *Naa35* to make up the N-acetyltransferase complex (NatC) (OMIM #617990 and OMIM #617989). In humans, *Naa35* is an auxiliary subunit in the NatC while *Naa30* is the catalytic subunit of the complex (Polevoda *et al.* 2009). When the NatC has a null mutation in human cell lines, cells show reduced growth and the induction of apoptosis (Starheim *et al.* 2009). The result is cell death from apoptosis due to the stabilization of p53 when the NatC is null mutated. While there are a few studies on the *Naa35* orthologous genes in humans, there are no current studies investigating the role of *Naa35* in *Drosophila*. It can be assumed that *Naa35* has similar effects in *D. melanogaster* as it does in humans, however, without any previous evidence this can only be an assumption.

The results from the two RNAi knockdown experiments partially support the hypothesis that the genes identified in the GWAS from the DGRP experiment have an influence in affecting

spontaneous walking behavior. Through the use of RNAi knockdown, the results indicate that the protein expression of certain candidate genes have an impact in time spent walking. Potential neural pathways that initiate or maintain walking behavior may be predicted through the expression of these genes and proteins. This could be a future avenue of research for the Andrew lab.

Due to genotypic variability between DGRP lines, increased walking indexes could have been caused by heightened sensory stimulation, increased motor neuron stimulation or increased motor neuron output. These variants could lead flies to walk for longer periods of time. These pathways can be influenced by numerous variations in the amount of neurotransmitters secreted, differences in synaptic connections, abnormalities in higher processing, neuron-to-neuron communication, or abnormalities at the neuromuscular junction. Due to the countless pathways with no singular basis that influence the difference in walking behavior among the DGRP lines, the results from the DGRP cannot yet give insight on any specific pathway affecting walking behavior. Further functional analysis of the genes implicated in this current study would need to be conducted.

RNAi knockdown of *hiw* expression show a significant decrease in walking behavior for the comparison to one of the negative controls. The experimental knockdown had a significant decrease in walking index compared to the background control cross (Figure 6). These results are consistent with previous work with *hiw* which has shown mutants of *hiw* to display walking defects (Wan *et al.* 2000). The lack of *hiw* protein expression that is responsible for normal growth and function of the NMJ could lead to dysregulation in the pathways that are responsible for walking behavior (McCabe *et al.* 2004). In the comparison of P{TRiP.JF02866}> P{GawB};P{Dcr} to w^{1118}>P{PGawB};P{Dcr} there was no significant difference between the

walking indexes (Figure 6). With there being no significant difference between the *hiw* knockdown and the control with the *GAL4 elav*, the *GAL4* elav could be the leading factor in the decrease between P{TRiP.JF02866}>P{GawB}elav;P{Dcr} and P{TRiP.JF02866}>w^{1118}. However, with the *GAL4 elav* control animals being over the limit of the normal recording age these results can only lead to a speculation and not a definitive answer. Previous research has indicated that behavior of *D. melanogaster* changes at different stages in the life cycle, even altering during the life of an adult fly (Tauber *et al.* 2011).

Results from the *Naa35* RNAi knockdown experiment demonstrated a significant increase in walking index. The P{TRiP.HMJ23456}>P{GawB}elav;P{Dcr} has a significant increase in walking behavior compared to both negative controls P{TRiP.HMJ23456}>w^{1118} and w^{1118}>P{GawB}elav;P{Dcr} (Figure 7). The significant increase between the experimental progeny and the P{TRiP.HMJ23456}>w^{1118} progeny suggest that *Naa35* may play a role in neurons that suppress spontaneous walking behavior if *Naa35* has the same function in flies as it does in humans (Starheim *et al.* 2009). If the knockdown of *Naa35* leads to neuron death along with reduced neuron growth and results in an increase in walking behavior, then the role of *Naa35* in *D. melanogaster* spontaneous walking behavior is to suppress walking behavior. It may suppress the output of motor signals or suppress the sensory signals that trigger walking behavior. Although there is a significant increase comparing the experimental cross with the *GAL4 elav* negative control, the animals were not recorded within the age limits, which suggests that these preliminary result show animal behavioral changes as age increases (Tauber *et al.* 2011).

To further support the identifications of the causes of the changes in walking behavior in the RNAi knockdown experiments, a positive control must be analyzed. In this experiment, two

negative controls were implemented to observe the impact of the *UAS* and the *GAL4* driver on walking behavior. Each individual *UAS* and *GAL4* was crossed with a wild type fly (*w*^{1118}) to create a hemizygotic animal to replicate the hemizygosity of both *UAS* and *GAL4* constructs in the experimental cross. For a positive control there should be a *GAL4* that is ubiquitously expressed so every cell will have an mRNA knockdown of the specific gene. *GAL4* lines targeting actin are ubiquitously expressed in every cell. Such lines exist, such as the Bloomington stock center line #25708 P{UAS-Dcr-2.D}1,w^{1118};P{Act5C-GAL4}25FO1/CyO. This stock was ordered and supposed to be implemented into the experiment as the positive control, however, the stock was not viable despite severe coddling. With this stock not producing a sustainable amount of animals, future experiments should utilize other *GAL4* actin drivers for a positive control for RNAi knockdown experiments.

Although this experiment focused on the implications of the candidate genes in the nervous system through the use of the *GAL4 elav*, future experiments should determine the effect these genes have in muscles. Certain genes, such as *hiw*, show a dual influence in the nervous system and the muscular system. To identify whether these genes in the muscular system have an impact on walking behavior, the TRiP *UAS* lines of these genes need to be crossed with *GAL4* muscle drivers. One muscle driver, P{Dcr}; P{GAL4-Mef}, was ordered and added to the laboratory stocks for the use of this RNAi knockdown experiment. Similarly to the *GAL4* actin drivers, this stock was not an optimal stock for this experiment. During the expansion of the *GAL4 Mef* animals, there was very few female animals produced. Due to *GAL4* animals needed to be virgin females for the experimental crosses, these animals were not used in this experiment. For future RNAi experiments, to investigate candidate genes in muscular system either new

GAL4 muscle drivers need to be utilized or the abundance of *GAL4 Mef* females need to be optimized.

To increase the accuracy of future experiments, it is optimal to analyze the entirety of the DGRP lines and not just a subset as we did in this study. The DGRP has over 200 isogenized lines that are available for research while this study only used a partial subset of their inventory for the sake of expedience. Utilizing more lines from DGRP can confirm the idea that spontaneous walking behavior is a quantitative trait and uncover more gene candidates that influence the natural variation of walking behavior. Previous research in this laboratory was limited in the number of DGRP lines that were able to be screened since the entirety of behavioral analysis was manual. With the introduction of Ethovision cutting behavioral analysis down from what used to take over 400 hours of analysis to taking within 24 hours, future experiments can now investigate on a much broader scale. The goal in future experiments is to not only automate the process of analyzing spontaneous walking behavior, but also more complex behaviors, such as grooming.

An improvement for the RNAi knockdown experiments involves validating if and where the specific *GAL4* drivers we used are expressing *GAL4*. *GAL4* drivers are supposed to be specifically expressed in certain area of the organisms at specific points in development. For instance, the *GAL4 elav* is supposed to be expressed in every neuron in the adult fly. A way to ensure the *GAL4* is working properly is crossing the *GAL4* with a fly containing a *UAS* of green fluorescent protein (GFP). Progeny of this cross will contain GFP for visualization and confirmation of where and when the *GAL4* is expressed. These animals can be viewed under fluorescent microscopes or under confocal microscopes to specifically indicate where the *GAL4* lines are expressing.

Another procedure that can validate the success of the RNAi knockdown experiment involves the confirmation of genotypic interference. Due to time constraints with this experiment, this procedure was not able to be accomplished. A method called reverse transcription real time polymerase chain reaction (RT-PCR or qPCR) can ensure that the TRiP-RNAi experimental crosses led to the degradation of the mRNA products of the candidate genes. qPCR uses fluorescent dye to observe the amplification of a specific mRNA products during PCR. A control gene that is highly expressed ubiquitously, typically GapDH, is used to calibrate the qPCR and identify how many amplifications of that control gene are necessary to reach a significance threshold. This experiment could utilize qPCR through comparing mRNA products of the RNAi knockdown crosses to the mRNA products of the control crosses. For validation of the degradation of the mRNA products from the RNAi knockdowns, there should be significant decrease in mRNA products compared to the control crosses.

Along with ensuring the knockdown of genotypic expression for this experiment, proteomic confirmation of the RNAi knockdown experiment should also be validated. Western blotting is a method that could identify the levels of protein expression in our RNAi experiments. A western blot comparing the RNAi experimental crosses and the control crosses would have a significant decrease in in the RNAi experimental crosses due to the degradation of mRNA products of the genes of interest. It is able to identify protein expression level through tagging proteins with antibodies specific to the proteins of interest. These antibodies could then be separated by size in a protein gel and visualized on a film later. The process of western blotting is cumbersome and time consuming, which is not optimal for experiments under such a time constraint like this one.

As RNAi knockdown is the most beneficial reverse genetics tool in terms of limited time and money, flaws do occur when using RNAi. False positives and false negatives are a possibility due to the fact that the efficiency of gene silencing can be inconsistent, and its impact on phenotypes may not be apparent if a certain threshold is not met (Yamamoto-Hino and Goto 2013). False positives occur when off target mRNA has similar sequences with the dsRNA so that mRNA is targeted for degradation instead of the target mRNA. A more efficient reverse genetic tool that is becoming more available is Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated protein (Cas9) (Doudna and Charpentier 2014). Originally a bacterial defense system, this mechanism has been manipulated in the laboratory by designing specific guide RNA (gRNA). This gRNA targets genes of interest by it being complementary to a target site on the gene. When the gRNA pairs with the gene of interest, the Cas9 endonuclease cuts the double stranded DNA (dsDNA). The cell recognizes this break in the DNA and aims to fix the breakage. Ways in which the cell fixes the problem is non-homologous end joining (NHEJ) or homology direct-repair (HDR). NHEJ creates insertions and deletions in the genomic code leading to null mutations of the targeted gene. This method is useful for experiments aiming to knockout a gene such as this one. HDR is where the cell recognizes the location of the breakage on the other copy of that DNA strand in the cell and uses it as a template to repair the broken DNA strand. Researchers have used HDR to insert desired genes into the genome by creating DNA strand that can be recognized as the extra DNA copy but with the selected gene imbedded in the synthetic DNA strand. CRISPR differs from the RNAi method by affecting the genome of the organism instead of altering the mRNA expression, which could be more efficient in knocking down gene expression. Future experiments could

utilize the CRISPR-CAs9 system if preliminary RNAi experiments do not appear to be working efficiently.

Given the evidence provided in this study, further investigation into genes of interest for spontaneous walking behavior is imperative. Through the validation of Ethovision, conducting experiments on walking behavior is an accomplishable task for an undergraduate student. Results from the DGRP walking experiment shows that spontaneous walking behavior is a quantitative trait due to the continuous variation among and between lines. From this a GWAS was conducted from the walking behavior data that indicated multiple genes involved in contributing to this behavior. RNAi knockdown experiments on two genes of interest, *hiw* and *Naa35*, demonstrated that each individual gene is involved in influencing walking behavior. With the identification and verification of gene variants influencing walking behavior, these findings could lead to better fly models of neurodevelopmental disorders.

Figures and Legends

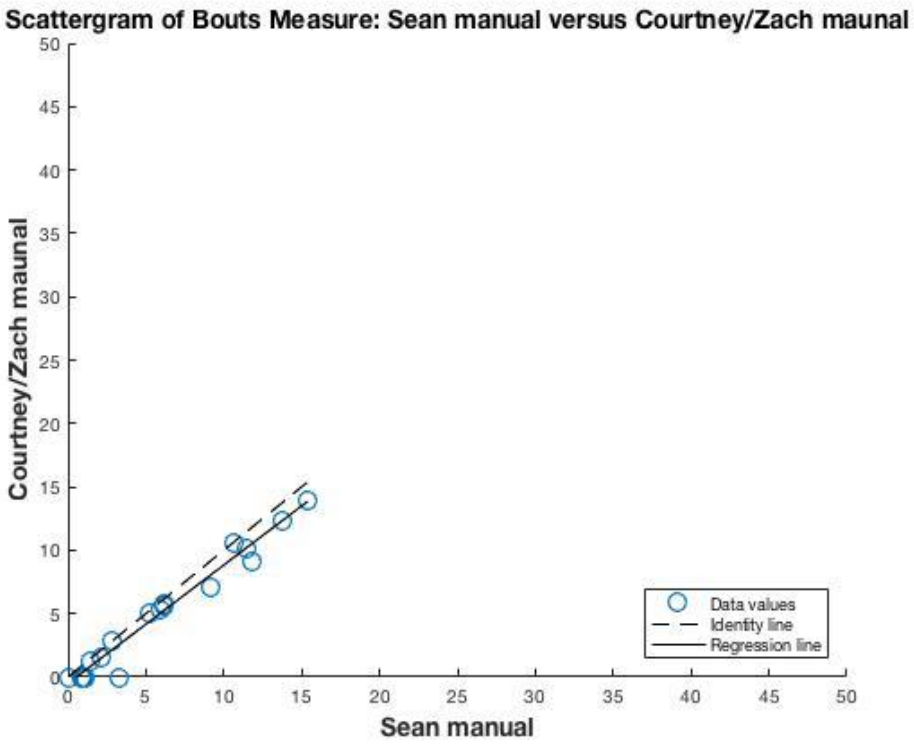

Figure 2 Comparison between manually scored (Sean Manual) and previous scored (Courtney/Zach) grooming indices of 21 samples using a geometric mean regression test with a significance level of 0.05. Slope was 0.9333 [0.8585, 1.0147] and an intercept of 0.5003 [-0.9205, 1.1137]. These intervals include 1 for the slope and 0 for the intercept so there is neither systematic nor proportional bias.

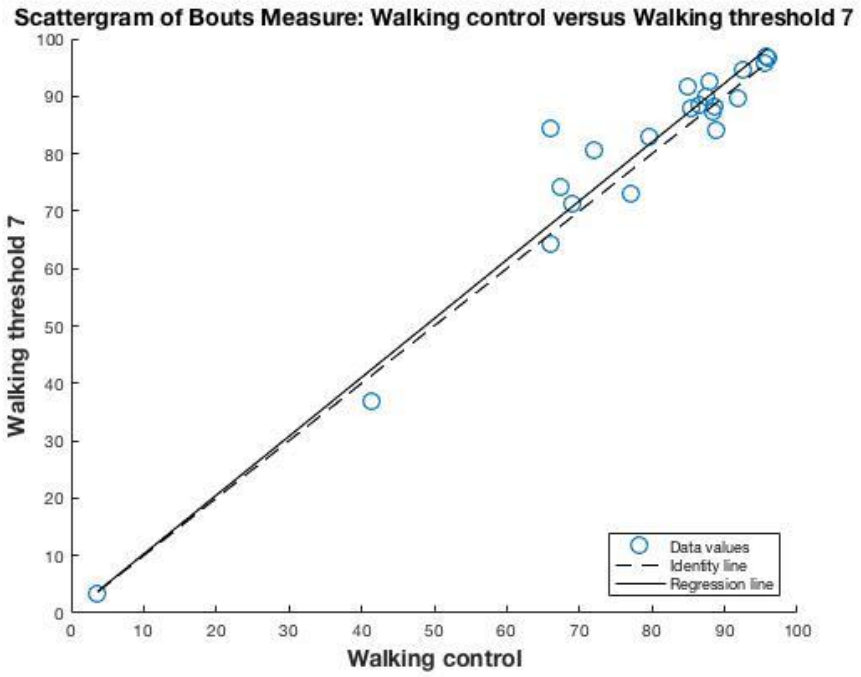

Figure 3 Comparison of Walking Indexes between manually scored videos and Ethovision scored videos on threshold 7 of 22 samples using a geometric mean regression test with a significance level of 0.05. Slope was 1.0252 [0.9190, 1.1437] and an intercept of 0.0181 [-9.1970,8.2787]. These intervals include 1 for the slope and 0 for the intercept so there is neither systematic nor proportional bias.

Averaging interval	WI slope	WI slope interval	WI intercept	WI intercept interval
1	1.0516	[0.9415, 1.1746]	-5.8377	[-15.4047, 2.7275]
2	1.0521	[0.9416, 1.1755]	-5.7279	[-15.3305, 2.8663]
3	1.0521	[0.9412, 1.1761]	-5.6362	[-15.2766, 2.9883]
4	1.0578	[0.9461, 1.1827]	-5.6565	[-15.3715, 3.0326]
5	1.073	[0.9617, 1.1971]	-6.1234	[-15.7813, 2.5328]

Table 8 Slope and intercept with 95% confidence intervals from different averaging interval settings within movement. The analysis test used was a geometric mean regression on 22 samples. The averaging interval with the smallest interval was and averaging interval of 1. The slope interval was 0.2331 and the interval for the intercept was 18.1322 indicating this averaging interval has the most precision.

Thresholds	WI slope	WI slope interval	WI intercept	WI intercept interval
1.00mm/s-0.99mm/s	1.0516	[0.9415, 1.1746]	-5.8377	[-15.4047, 2.7275]
0.95mm/s-0.94mm/s	1.0469	[0.9353, 1.1718]	-4.6499	[-14.3647, 4.0295]
0.91mm/s-0.90mm/s	1.0442	[0.9368, 1.1639]	-4.1076	[-13.4178,4.2520]
0.85mm/s-0.84mm/s	1.0407	[0.9340, 1.1595]	-3.1334	[-12.3778, 5.1636]
0.81mm/s-0.80mm/s	1.0376	[0.9307,1.1452]	-2.1464	[-11.4089, 6.1624]
0.75mm/s-0.74mm/s	1.0267	[0.9205, 1.1452]	-0.7805	[-9.9983, 7.4835]
0.71mm/s-0.70mm/s	1.0252	[0.9190, 1.1437]	0.0181	[-9.1970, 8.2787]

Table 9 Slope and intercept with 95% confidence intervals from different threshold settings within movement. The analysis test used was a geometric mean regression on 22 samples. The threshold 0.71 mm/s – 0.70 mm/s had the slope closest to 1 and the intercept closest to 0. This suggest it is the best threshold to use to prohibit systematic bias and proportional bias.

DGRP Line	Male WI	Female WI
208	77.65	78.46
301	78.97	77.73
303	73.83	77.46
304	57.68	62.71
307	55.44	25.97
313	88.49	82.23
315	89.83	86.25
324	93.84	86.79
335	49.86	40.83
357	78.39	72.07
358	76.19	78.97
360	79.49	81.37
365	86.80	67.33
375	80.08	80.00
379	76.50	66.59
380	77.54	78.34
391	91.21	84.90
399	80.74	69.88
427	74.41	62.88
437	87.01	82.37
486	82.25	81.09
517	72.55	66.31
555	71.29	68.51
705	88.82	73.87
707	78.27	65.55
712	83.72	79.47
714	69.17	60.09
732	83.25	77.32
765	70.03	61.97
774	85.91	78.24
786	85.35	76.03
799	79.22	82.46
820	89.11	82.33
852	94.79	87.43
859	86.65	78.56

Table 10 Mean WI for males and females of every DGRP line from the experiment. Submitted in this format to the DGRP website.

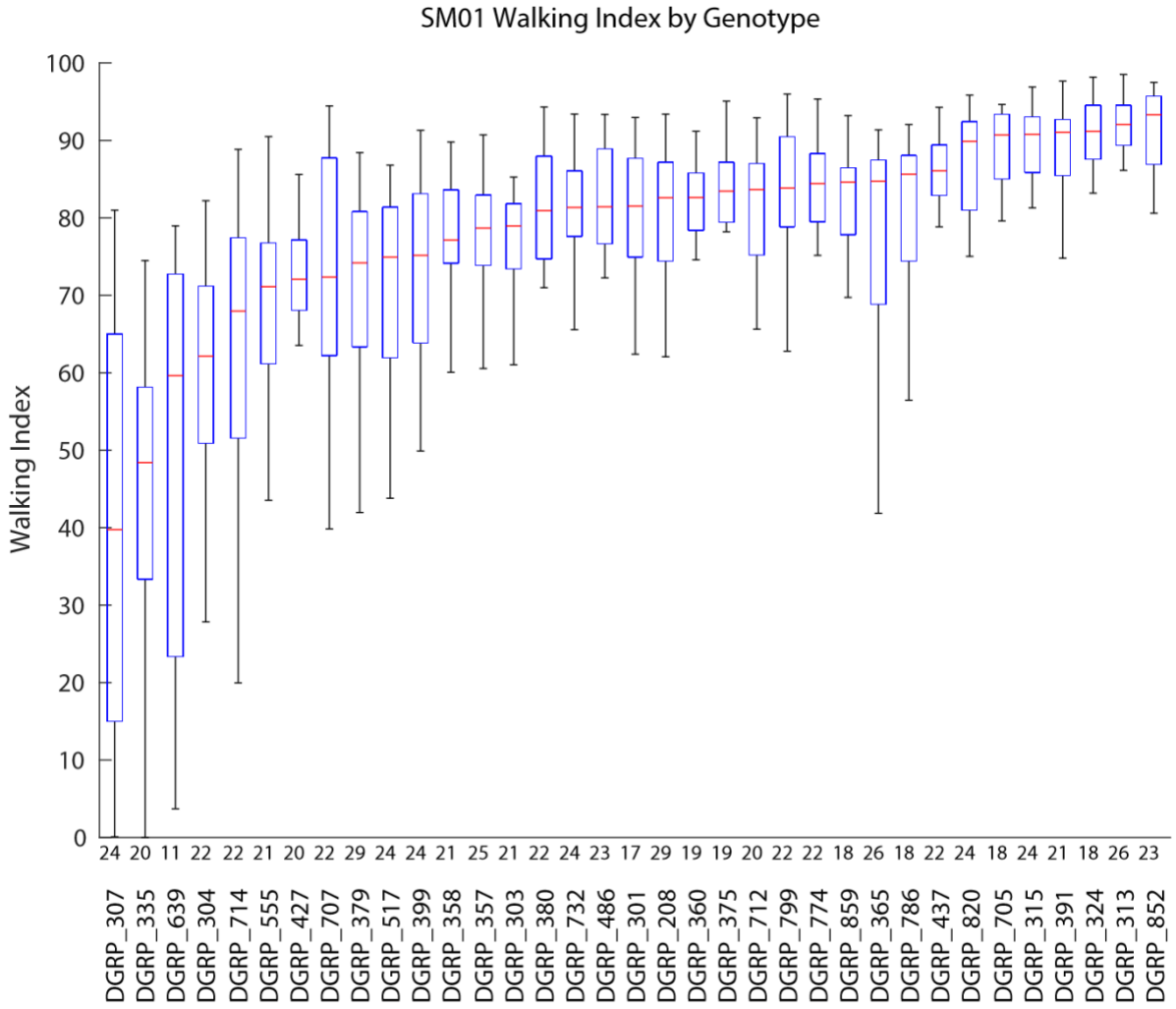

Figure 4 Variation of walking behavior from the different DGRP lines. The box-and-whisker plots were ordered by the median of each DGRP line. There is a significant difference between the lowest line DGRP 307 and the highest line DGRP 852 (p-value = 6.66 x 10^{-9}).

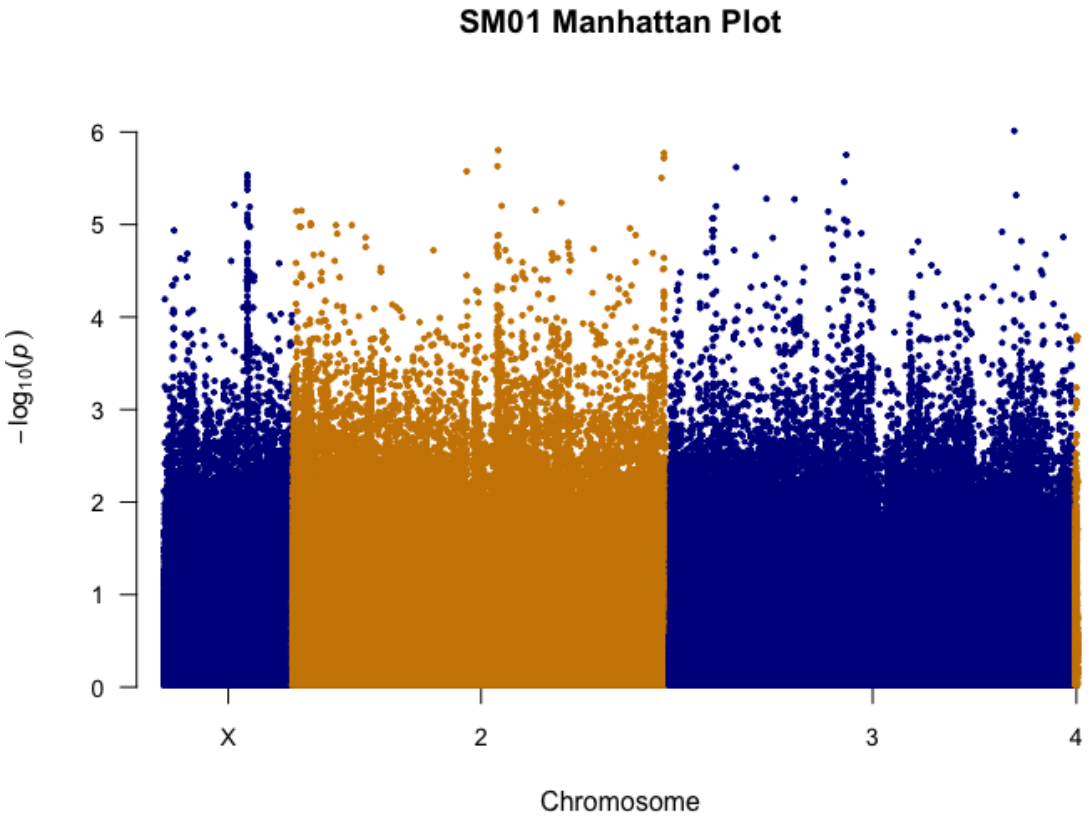

Figure 5 Manhattan Plot of each individual SNP in the genome. Each dot represents a SNP in the entire genome. The higher the dot is on the Y-axis, the more the SNP is associated with walking behavior. The threshold for this experiment was any *p*-value lower than 5×10^{-5}.

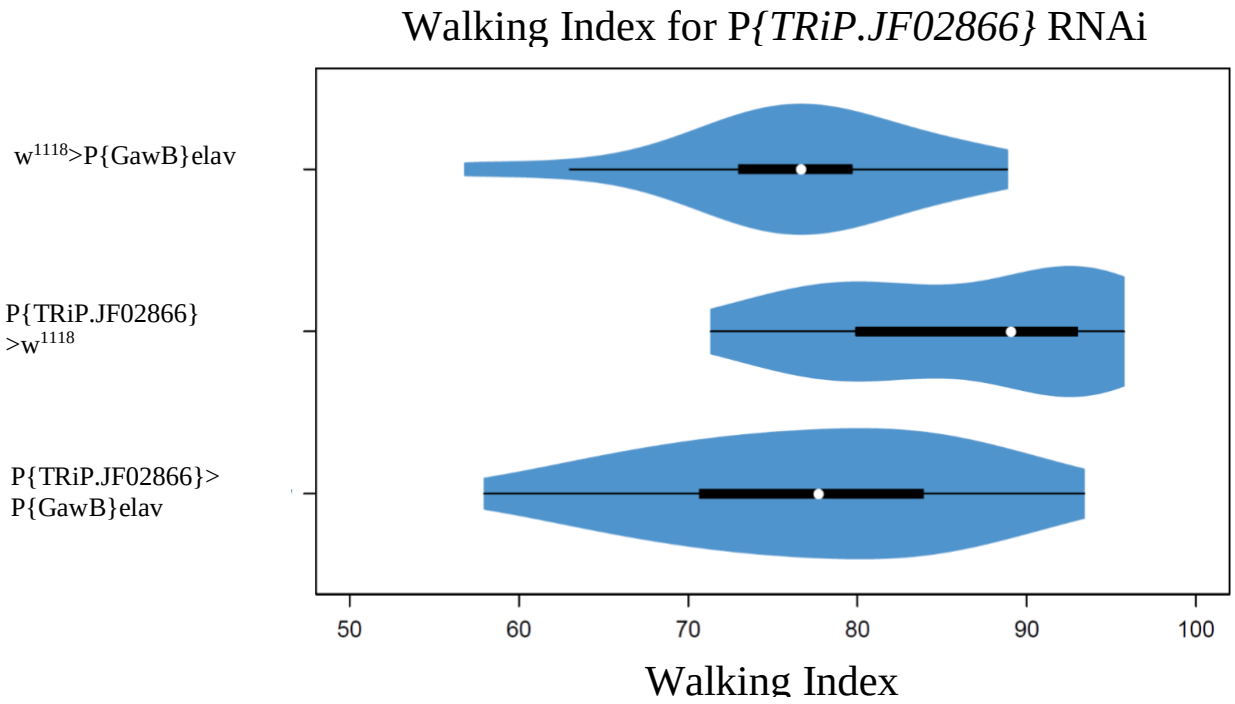

Figure 6 Violin Plots of the walking indexes for the RNAi experiments on *hiw*. There is a significant difference between *hiw*>*elav* and *hiw*>w^{1118} (*p*-value = 0.00155). There was no significant difference between *hiw*>*elav* and w^{1118}>*elav* (*p*-value = 0.77206).

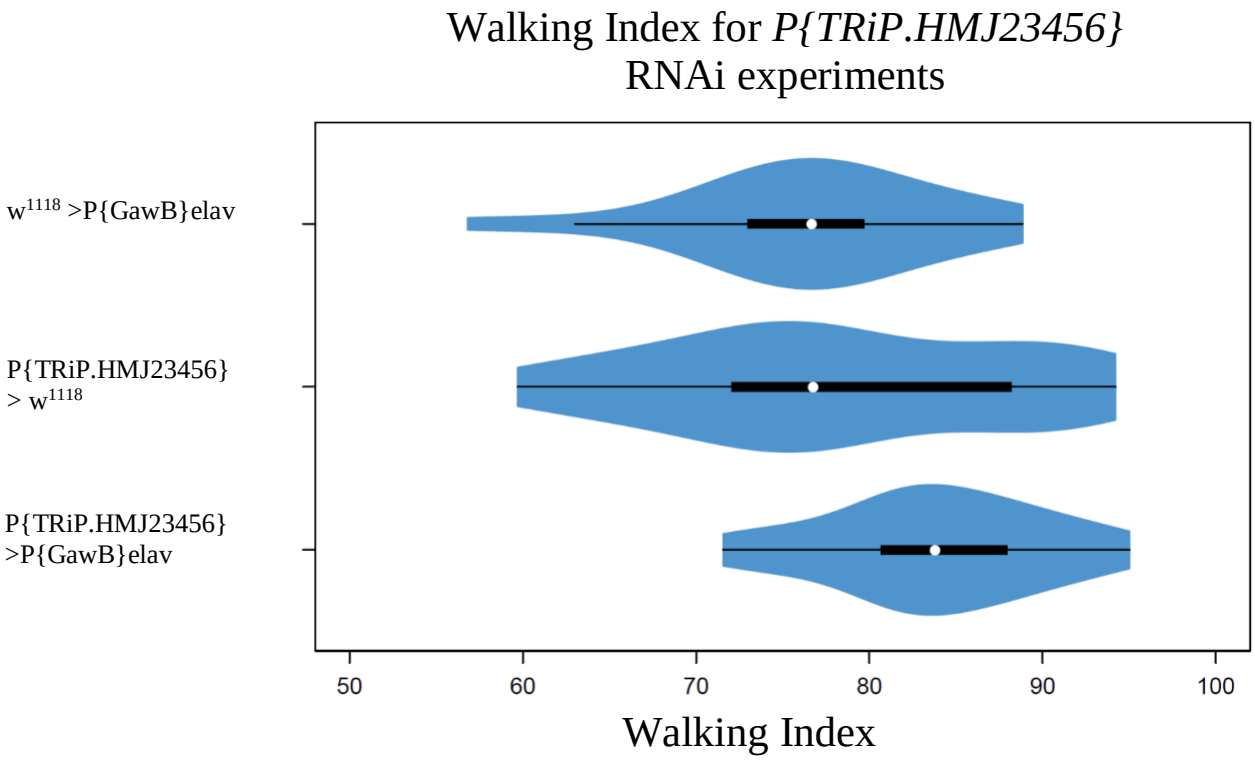

Figure 7 Violin Plots of the walking indexes for the RNAi experiments on *Naa35*. There is a significant difference between *Naa35*>*elav* and *Naa35*>w^{1118} (*p*-value = 0.01385). There was a significant difference between *Naa35*>*elav* and w^{1118}>*elav* (*p*-value = 0.00035).

References

Arya, G. H., M. M. Magwire, W. Huang, Y. L. Serrano-Negron, T. F. C. Mackay, *et al*., 2015 The genetic basis for variation in olfactory behavior in *Drosophila melanogaster*. Chem Senses. 40: 233-243.

Batsching, S., R. Wolf, and M. Heisenberg. 2016 Inescapable stress changes walking behavior in flies – learned helplessness revisited. PLoS One 11: e0167066.

Duffy, J. B. 2002 GAL4 system in Drosophila: A fly geneticist's swiss army knife. Genesis 34: 1-15.

Gaertner, B. E., E. A. Ruedi, L. J. McCoy, J. M. Moore, M. F. Wolfner, *et al*., 2015 Hertiable variation in courtship patterns in *Drosophila melanogaster*. G3. 5: 531-539.

Huang, W., A. Massouras, Y. Inoue, J. Peiffer, M. Rámia, *et al*., 2014 Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. Genome Res. 24: 1193-1208.

Kaun, K. R., A. V. Devineni, and U. Herberlein. 2012 *Drosophila melanogaster* as a model to study drug addiction. Hum Genet. 131: 959-975.

Kaur, K., A. F. Simon, V. Chauhan, and A. Chauhan 2015 Effect of bisphenol A on *Drosophila melanogaster* behavior – A new model for the studies on neurodevelopmental disorders. Behav Brain Res. 284: 77-84.

Kaya-Copur, A., and F. Schnorrer. 2016 A guide to genome-wide in vivo RNAi applications in *Drosophila*. *Drosophila*: Methods and Protocols. 1478: 118-143.

Ki, S., H-S. Lee, and Y. Park. 2017 Perinatal exposure to low-dose imidacloprid causes ADHD-like symptoms: Evidence from an invertebrate model study. Food Chem Toxicol. 110: 402-407.

Kim, S., H.S. Lee, and Y. Park. 2017 Perinatal exposure to low-dose imidacloprid causes ADHD-like symptoms: Evidence from an invertebrate model study. Food Chem Toxicol. 110:402-407.

Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, *et al*., 2012 The *Drosophila melanogaster* Genetic Reference Panel. Nature. 482: 173-178.

Martin A. R., T. Raabe, and M. Heisenburg. 1999 Central complex substructures are required for the maintenance of locomotor activity in *Drosophila melanogaster*. J Comp Physiol A. 185: 277-288.

McCabe, B. D., S. Hom, H. Aberle, R. D. Fetter, G. Marques, *et al.,* 2004 Highwire regulates presynaptic BMP signaling essential for synaptic growth. Neuron. 41: 891-905.

Morozova, T. V., W. Huang, V. A. Pray, T. Whitham, R. R. H. Anholt, *et al*., 2015 Polymorphisms in early neurodevelopmental genes affect natural variation inalochol sensitivity in adult *Drosophila*. BMC Genomics. 16: 865.

Polevoda, B., T. Arnesen, and F. Sherman. 2009 A synopsis of eukaryotic N^{α}- terminal acetyltransferases: nomenclature, subunits and substrates. BMC Proc. 3: S2.

Shen, W., and B. Ganetzky. 2009 Autophagy promotes synapse development in *Drosophila*. J Cell Biol. 187: 71.

Starheim, K. K., D. Gromyko, R. Evjenth, A. Ryningen, J. E. Varhaug, et al., 2009 Knockdown of human N^{α}- terminal acetyltransferases complex C leads to p53-dependent apoptosis and aberrant human Arl8b localization. Mol Cell Biol. 29: 3569-3581.

Strauss, R. 2002 The central complex and the genetic dissection of locomotor behaviour. Curr Opin Neurobiol 12: 633-638.

Tauber, J. M., P. A. Vanlandingham and B. Zhang. 2011 Elevated levels of the vesicular monoamine transporter and a novel repetitive behavior in the *Drosophila* model of Fragile X Syndrome. PLoS One 6: e27100.

Tian, X., J. Li, V. Valakh, A. DiAntonio, and C. Wu. 2011 Drosophila Rae1 controls the abudance of the ubiquitin ligase Highwire in post-mitotic neurons. Nat Neurosci. 14: 1267-1275.

Wan, H. I., A. DiAntonio, R. D. Fetter, K. Bergstrom, R. Strauss, *et al*. 2000 Highwire regulates synaptic growth in *Drosophila*. Neuron 26: 313-329.

Weber, A. L., G. F. Khan, M. M. Magwire, C. L. Tabor, T. F. C. Mackay, *et al*., 2012 Genome-wide association analysis of oxidative stress resistance in *Drosophila melanogaster*. PLoS One. 7: e34745.

Wu, C., Y. P. Wairkar, C. A. Collins, and A. DiAntonio. 2005 *Highwire* function at the *Drosophila* Neuromuscular Junction: Spatial, Structural, and Temporal Requirements. J Neurosci. 25: 9557-9566.

Yamamoto-Hino, M., and S. Goto. 2013 In vivo RNAi-based screens: Studies in model orgnaimsms. Genes. 4: 646-665.