Genetic Analysis of Spontaneous Grooming Behavior in the Fruit Fly *Drosophila melanogaster*

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

By
Courtney L. Hannum
Lycoming College
December 2017

Approved by:

(Signature)

(Signature)

(Signature)

(Signature)

This work was supported by a Joanna and Arthur Haberberger Fellowship, awarded to Courtney L. Hannum at Lycoming College, Williamsport, PA.

Table of Contents

Abstract……..……....3

Background and Significance……4

Materials and Methods……..13

Results……22

Discussion……..….28

Figures……41

References……58

Abstract

Repetitive behaviors are commonly associated with several human neurodevelopmental disorders. For example, restricted and repetitive patterns of behavior are a crucial diagnostic criterion for Autism Spectrum Disorder and are often concomitant with other neurodevelopmental disorders such as Fragile X Syndrome. The genetic model organism, *Drosophila melanogaster,* can be used to study neurodevelopmental disorders, as well as natural genetic variations influencing behavior. In this study, we examine genetic variation in flies that contributes to a spontaneous and repetitive behavior: grooming. Results show that there is quantitative variation in grooming among various genetically distinct fly populations from the *Drosophila* Genetics Reference Panel. Through conducting a genome-wide association study, we have identified candidate genes potentially associated with altered grooming that have previously been implicated in nervous system development, mechanosensory projection, and immune system regulation. Preliminary results from RNA interference knockdown experiments of candidate genes, including the genes *Mad*, *psh*, and *bun*, support of a behavioral association with grooming behavior levels and provide sufficient evidence for continued functional validation of candidate genes.

Background and Significance

In many animals, grooming behavior is an innate and crucial aspect of life. This is particularly true of flies, where the presence of foreign debris can hinder an individual's ability to sense the environment, which can negatively impact their ability to successfully mate and find food, ultimately affecting fitness. Well-documented continual variation in grooming behaviors of wild-type fruit flies, *Drosophila melanogaster*, supports the idea that grooming is likely a quantitative trait. Quantitative traits vary among individuals in a population to produce a continuous distribution of phenotypes that are determined by the cumulative influence of multiple genes and the environment. A common method for studying genetic influences of quantitative traits in, *Drosophila*, utilizes the *Drosophila* Genetics Reference Panel (DGRP). This resource consists of more than 200 distinct and genetically isogenized *D. melanogaster* lines with fully sequenced genomes (Huang *et al*., 2016; Mackay *et al*., 2012). DGRP lines were created through inbreeding flies from a single wild-type population to capture, in each line, a subset of the genetic diversity originally present. This quality makes the DGRP useful for genome-wide association studies (GWASs), which allow for the identification of genomic regions associated with a particular quantitative trait. Here, we use a subset of DGRP lines to determine if and how specific genes influence phenotypic expression of grooming behavior toward the goal of elucidating the genetic underpinnings of this complex trait.

Several studies have previously utilized the DGRP to examine the genetics of innate and complex quantitative phenotypes, thereby supporting the utility of GWAS analysis using the DGRP. For example, this methodology allowed for analysis of *Drosophila* sleep patterns during the day and night (Harbison *et al*., 2013). Sleep was

shown to be partially influenced by genetics and is a sexually dimorphic trait. Multiple genes identified by the study have previously been related to human sleeping patterns and disorders (Harbison *et al*., 2013). Specifically, researchers found the *peter pan* gene to affect predominantly day sleep in flies. This gene was formerly determined to be orthologous to the human gene *P2RY11.* In humans, genetic mutations within *P2RY11* can contribute to narcolepsy (Kornum *et al*., 2011). Similarly, the same GWAS determined that *nudE* in flies affects night sleep and is orthologous to *ApoE* in humans, which has an allele known to cause sleep apnea (Gottlieb *et al*., 2004; Harbison *et al*., 2013). More recently, variation of food intake among several DGRP lines has been studied via genome-wide association, revealing sexual dimorphism and genetic correlations. Candidate genes were identified and 77% were validated. Many were found to be significantly correlated with increasing or decreasing food intake in fruit flies and 16 genes, more than half of those studied, had mammalian orthologs known to be related to human metabolism (Garlapow *et al.,* 2015). Such studies exemplify the DGRP's effectiveness in facilitating the genetic understanding of human disorders.

The majority of DGRP studies being conducted employ similar methods. After collecting data regarding a phenotype of interest, results are submitted to the DGRP website in order to identify single nucleotide polymorphisms (SNPs), or genetic variants, correlated with the particular phenotype (http://dgrp.gnets.ncsu.edu/). Specifically, arithmetic means of behavioral quantifications for both males and females of each analyzed DGRP line are submitted to the DGRP website for a GWAS (Huang *et al*., 2016; Mackay *et al*., 2012). Researchers then locate potentially causative genes within the surrounding regions of top associated SNPs. Following the identification of these

candidate genes, researchers can advance their work in a more specific direction through implementing reverse genetic techniques (e.g. Garlapow *et al*., 2015; Harbison *et al.*, 2013; Huang *et al*., 2012). This process has the ability to determine a gene's effect on a given phenotype.

One commonly applied method of reverse genetics is RNA interference (RNAi), a technique to specifically knockdown the expression of a gene of interest (GOI) by reducing or eliminating its corresponding messenger RNA (mRNA) (Yamamoto-Hino & Goto, 2013). RNAi allows for the analysis of consequential phenotypes in a genetic knockdown, thereby enabling an evaluation of normal gene function. This method does not require a change in the genomic sequence, but rather focuses on gene products, making RNAi more feasible for the study of multiple genes in a short period of time. In *Drosophila*, RNAi can be conducted *in vivo* through the use of a GAL4/*UAS* binary expression system (Figure 1). The technique relies on crosses between fly lines expressing the transgenic transcription factor GAL4 and a line containing GAL4's corresponding upstream activating sequence (*UAS*) upstream of a reporter gene of interest (Yamamoto-Hino & Goto, 2013). Progeny of this cross produce the yeast GAL4 transcription factor in known cell and tissue types, which is able to bind the *UAS* and induce expression of the adjacent GOI, in a precise spatiotemporal manner. Differing *GAL4* lines exhibit GAL4 expression during various developmental times and in different tissue types, allowing for detailed analysis of resulting phenotypes. Reporter genes are frequently associated with the *UAS* so that the system's specificity is evident. For example, green fluorescent protein (GFP) reporters lead to fly tissues that glow when placed under ultraviolet light, as a result of GAL4 induced expression, thus verifying the

particular GAL4 expression pattern. In this experiment, we pair specific GAL4 expression with a *UAS*-RNAi construct to regulate its expression.

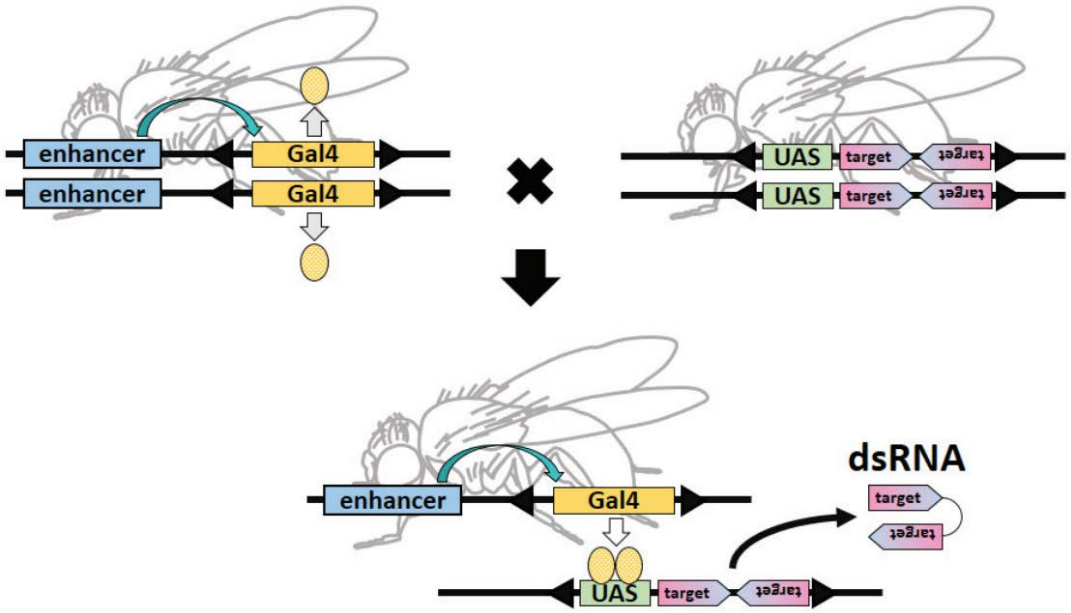

Figure 1: Interactions occurring between GAL4 protein and UAS to induce the expression of a dsRNA hairpin molecule downstream for RNAi. Figure adapted from Yamamoto-Hino & Goto (2013).

RNAi experiments *in vivo* utilize *UAS* lines that contain regions of DNA that, upon induction by GAL4, produce hairpin double stranded RNA (dsRNA) transcripts (Kaya-Copur & Schnorrer, 2016). The protein dicer, then cuts these long dsRNA molecules into short interfering RNA, or silencing RNA (siRNA), that have been purposefully designed as complementary sequences to the GOI mRNA product (Kaya-Copur & Schnorrer, 2016; Yamamoto-Hino & Goto, 2013). These siRNA molecules serve as a template for the RNA induced silencing complex (RISC), which then recognize and cleave the complementary mRNA of the GOI. Following cleavage, mRNA of interest is degraded prior to its translation into functional protein, in turn reducing the protein expression of this gene. This method has wide application, including analyzing the effects of genes implicated by DGRP studies. For example, one experiment used RNAi

knockdown to functionally validate candidate genes determined through a GWAS (Garlapow *et al*., 2015). By knocking down these genes individually, Garlapow *et al*., (2015) were able to deduce which had the greatest effect on their quantitative behavior of interest, namely food intake.

Although many studies have assessed various behaviors using the DGRP, analysis of grooming behavior as a quantitative trait has yet to be examined. Some neural circuits related to grooming behavior have been elucidated by means of circuit analysis of cleaning reflexes, indicating the role of the nervous system in initiating and controlling this behavior (Kays *et al*., 2014). Other findings support the idea of a genetic basis for behavioral “cleaning modules” common movement patterns involved in cleaning various areas of the *Drosophila* body (Seeds *et al*., 2014). There are common patterns of cleaning modules utilized in grooming behavior that are advantageous to fly fitness, illustrating the hierarchical nature of grooming behavior. For instance, the most common grooming pattern involves flies beginning with their eyes and antennae before moving onto their legs, and lastly their thorax. This hierarchical pattern is only observed when grooming is heavily stimulated by dust covering, rather than in spontaneous grooming, which occurs in the absence of a prompting stimuli (Seeds *et al*., 2014). Studies such as these suggest that genetic loci associated with grooming behavior may primarily identify genes related to nervous system function, many of which will likely have human orthologs.

Drosophila is a prominent model organism in diverse branches of biology, particularly genetics. Knowledge regarding the genetic influences on grooming behavior in fruit flies has the potential to impact the study of various human neurodevelopmental

disorders affecting motor abilities. Human neurodevelopmental disorders affect approximately 15% of children in the United States, causing the prevention and treatment of such to be a common area of study (America's Children and the Environment, 2015). Many of these disorders display a wide range of motor symptoms that arise from genetic abnormalities affecting nervous system functioning. Autism Spectrum Disorder (ASD) and related disorders such as Fragile X Syndrome (FXS) and Rett Syndrome are a few characterized by motor skill deficits and abnormalities (American Psychiatric Association, 2013). Symptoms often include restrictive and repetitive patterns of behavior, and, in the case of ASD, these symptoms largely influence a diagnosis. These deficits can prevent individuals from completing everyday tasks and communicating via sign language, thus, making the study of motor skills in regards to neurodevelopmental disorders essential.

The most common monogenic cause of human neurodevelopmental disorders is mutations in the *Fragile X Mental Retardation 1* (*FMR1*) gene that encodes the FMR protein (FMRP) (Heitzer *et al*., 2012). The well-conserved *D. melanogaster* ortholog is known as *dfmr1* (Friedman *et al.,* 2013). In humans, *FMR1* is typically mutated by the occurrence of CGG trinucleotide repeats, leading to decreased production of FMRP. The most common consequence of a *FMR1* mutation in humans is FXS and a large percentage of individuals with FXS concomitantly develop ASD. Research has shown that when the *dfmr1* gene is mutated in flies, the resulting grooming behavior is abnormal. In one study, mutated flies groomed 79% of the five minutes they were observed, in comparison to the control flies who groomed only 9% of the five minutes they were observed (Tauber *et al*., 2011). This increased grooming has been

analogized to the restrictive or repetitive behaviors that define one of the diagnostic criteria of ASD. Excessive grooming of *dfmr1* mutants supports that repetitive behaviors can be observed in flies and humans as a consequence of mutations in orthologous genes, some of which relate to FXS, autism, and potentially other neurodevelopmental disorders.

Interestingly, mice models of FXS have also been studied following the creation of null mutations in the orthologous mouse gene, *Fmr1*. Behavioral phenotypes presented by the mice are analogous to those of human FXS patients (McNaughton *et al*., 2008; Mines *et al*., 2010). These behaviors include social isolation and repetitive motor activity. In conducting these studies, excessive grooming behavior was observed in the *Fmr1* null experimental mice when compared to the wild-type mice. Conclusions from social interaction studies suggest that this type of repetitive behavior may be a result of social anxiety caused by the mutation (Heitzer *et al*., 2012; McNaughton *et al*., 2008; Mines *et al*., 2010). Consistent findings between model organisms and human patients support the utility of such organisms for seeking further insight into neurodevelopmental disorders, related neuronal circuitry, and basic brain functioning.

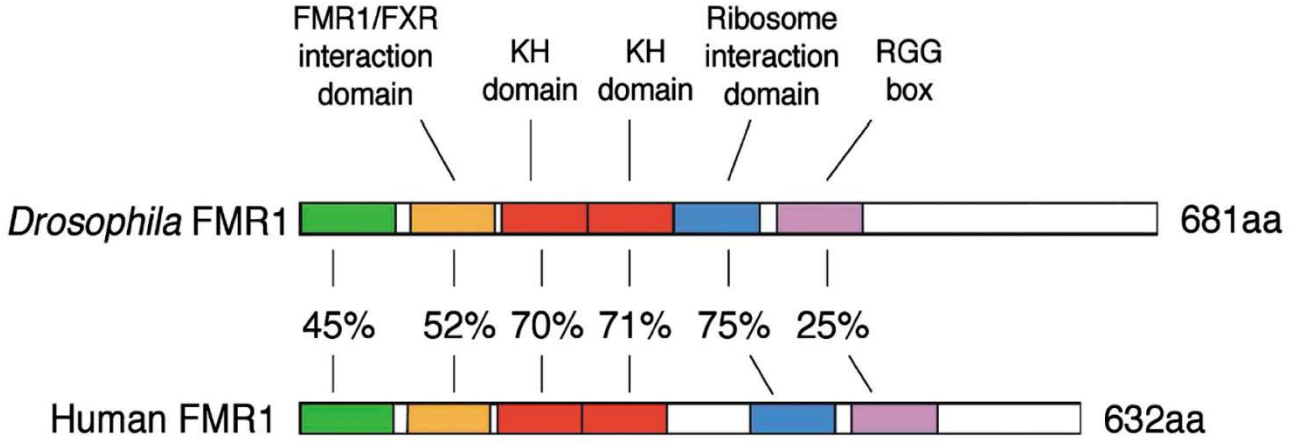

Figure 2: Comparison between *Drosophila* and *Homo sapiens* amino acid sequence identity color coded by domain. Typically, proteins with >25% amino acid identity share the same overall structure. Figure adapted from Gao (2002).

In all three organisms, the orthologous FMR protein is thought to have similar functional roles. FMRP is an RNA-binding protein involved in mRNA metabolism, translational repression, and export from the nucleus (Bagni & Greenough, 2005). Many other proteins interact with FMRP both directly and indirectly. Some of these proteins include cytoplasmic FMRP-interacting protein 1 (CYRIP1) and Staufen, which are both involved with mRNA transport (Bagni & Greenough, 2005). FMRP has also been found to function in the regulation of mRNA expression as well as in RNA-interference pathways, primarily in neurons. A protein that has been identified in humans to be associated with RNA interference through interactions with FMRP is eIF2C2. It has been determined that an orthologous protein in *Drosophila*, AGO1, is particularly important in synaptic plasticity (Bagni & Greenough, 2005; Jin *et al*., 2004). In comparing orthologous proteins of humans and flies, amino acid identity was determined to be very high among each of the 6 domains in FMRP (Gao, 2002). An amino acid identity greater than 25% typically results in the same overall domain structure. Therefore, the functional structures of FMRP and dFMR1 are nearly identical (Figure 2). Such similarities between protein-protein interactions and protein structure further support the use of animal models for the study of neuronal pathways in relation to neurodevelopmental disorders.

Abnormal grooming behavior has been found consistently in animal models of FXS. Being that this is a neurodevelopmental disorder and grooming behavior has been studied in regards to neuronal pathways, our GWAS of grooming behavior as a quantitative trait will likely lead to the identification of genes involved with regulating the central and/or peripheral nervous system. We hypothesize that grooming is a

quantitative behavior of fruit flies showing continuous variation both among and between various DGRP lines. We further hypothesize that reverse genetic studies following the GWAS will functionally validate candidate genes and may yield insights into neuronal function or circuitry that are potentially comparable to that of humans. Such information could lead to an increased understanding of repetitive movements as presentations of human neurodevelopmental disorders.

Materials and Methods

Rearing and Collection of DGRP Animals

A subset of the DGRP comprising 38 genetically isogenized DGRP lines, originally isolated from a single wild-type population, were utilized in this study (Mackay *et al*., 2012). Upon obtaining the DGRP lines from Bloomington *Drosophila* Stock Center (BDSC), they were initially quarantined to prevent the potential spread of mold and/or disease into the existing Andrew Lab fly stocks. After ensuring the health of DGRP lines, expansion populations were created for each line. To do this, 5 virgin females and 10 males were placed into the first vial to maintain a steady pool of animals for the study. Polystyrene *Drosophila* vials measuring 25x95mm were used and each contained approximately 8.5g of Nutri-Fly^{TM} BF food (Genesee Scientific, San Diego, CA). The excess of males allowed for sufficient production of offspring due to high rates of mating. Each day, the 15 flies were transferred into a new vial. This process was continuous until a line was completed by obtaining a minimum of 8 male and 8 female video recordings. Expansions were kept inside the incubator at 25^{o}C with high humidity (over 50%). Circadian rhythm was regulated by a 12-hour light/12-hour dark cycle. These are optimal conditions for rearing healthy *D. melanogaster* (Greenspan, 2004).

Pupae were collected from expansion vials after being distinguished as male or female. Each collected pupa was placed individually into a labeled 16x100mm sterile culture tube containing approximately 2g of Nutri-Fly^{TM} BF food (Genesee Scientific, San Diego, CA). The DGRP line, date, time, sex, and tube label were recorded for each animal. Tubes were placed in the incubator and monitored once per hour during subsequent 12-hour day periods. Flies that eclosed, or emerged from their pupal cases,

within the hour period between monitoring sessions were recorded and used in a video 24 hours later, when adults were 1-day old. Flies with physical abnormalities, such as broken legs or wings that failed to unfurl, were not recorded for use in the study. Flies that eclosed overnight were not used since a specific period of eclosion was unknown.

Behavior Recording of DGRP Lines

Fly behavior was recorded with a video camera (Canon Vixia HF R72), approximately 24 hours after their eclosion was monitored. Flies were recorded in the behavior room maintained at around 25ºC and 70% humidity. Animals were brought to the behavior room at least one hour before their respective recording time, to allow them to acclimate to the environment of the room. Recordings occurred in the dark with only the circular microscope LED desk light on (Model: Mic-2019). Six flies were aspirated into individual wells on a sterile polystyrene 96 well plate and contained by sliding a coverslip over the top (Genesee Scientific, San Diego, CA). Each well had previously been filled with 200μL of 1.5% agar. Agar helped to maintain local humidity and provided a consistent substrate for the flies, while also restricting their vertical movement within the recording arena. Flies were placed in two rows of three and their DGRP line, sex, and well number were recorded. Also recorded were the scorer's initials, recording time, date, room conditions (temperature and humidity), and initial and final plate temperatures. The video camera was set up directly over the wells, in the center of the light, and focused manually. The preparation takes fewer than 5 minutes to perform. Once preparation was complete, the 6 flies were recorded for 10 minutes. Significant observations, such as excessive grooming or standing, were noted. After the videos, flies were individually placed into 0.5mL microcentrifuge tubes (Fisherbrand®)

labeled with their identification number and kept in a -20ºC freezer for subsequent genotyping if required.

Behavior Analysis of DGRP Lines

Periods of grooming bouts were observed in the videos in order to collect quantifiable data on the grooming behavior of *D. melanogaster*. Raw video files were backed up on a Dropbox account, as well as an external storage drive. Videos were condensed using Adobe Media Encoder. The video annotation software, VCode, allowed for manual scoring of the videos (Hagedorn *et al*., 2008). The scoring process was completed by pressing the key corresponding with the correct fly, to mark the beginnings and ends of grooming bouts allowing for later quantification of behavioral expression. Grooming was defined as any stroking of the head, antennae, proboscis, limbs, wings, or thorax using one or more limbs. Grooming behavior began when the limb used was first lifted off of the ground and stopped when the limb being used was removed from the body part being groomed. Grooming was not categorized by separate body parts. If multiple body parts were groomed in succession without standing, walking, or falling in between, it was considered to be the same grooming bout. Two scorers worked to complete the analysis using VCode after both had undergone a statistical comparison of interrater reliability to ensure that results were consistent between scorers. This process did not explicitly blind scorers to fly genotypes, however, the collection and recording procedures sufficiently randomized flies such that scorers had no knowledge of individual fly line designations during the scoring process.

Data Analysis of DGRP Lines

Custom, in-house Perl scripts were utilized to extract grooming data from VCode

output. Data were collected on the number of grooming bouts, grooming bout duration, and percent of total time in the arena spent grooming (grooming index, GI) for each fly. After these data were gathered and combined into one Excel document, figure construction and statistical analysis were performed. Figures were constructed via custom, in-house MATLAB scripts. Ethograms were produced from the VCode analysis. To identify statistical significance, the nonparametric Wilcoxon Rank-Sum test was applied, and significance was assessed as $p < 0.05$.

Conduction of a Genome-Wide Association Study

To conduct the GWAS, average GI values for males and females of each DGRP line were submitted to the DGRP website (http://dgrp.gnets.ncsu.edu/). Results were obtained via email. SNPs associated with the phenotypic averages are provided along with additional information including location, precise variation, type of genomic region, gene name if applicable, and *p*-value. The lowest *p*-values, corresponding to highest significance, were researched using FlyBase to identify interesting genes (Flybase.org; release FB2017_04).

RNAi Functional Validation

From the top associated SNPs, 6 candidate genes were selected based on *p*-value and relevant function for the development of *in vivo* RNAi experiments to infer gene function (Table 1). Transgenic RNAi Project (TRiP-RNAi) fly stocks from the Harvard Medical School collection were used to obtain pre-constructed *UAS*-hairpin fly lines specific for the chosen genes of interest (Hu *et al*., 2016). Specific *GAL4* lines were chosen based on spatiotemporal expression. The *P{GawB}elav; P{Dcr}* line was selected to drive *GAL4* expression, conferred by the *elav* driver, in all neurons during all

developmental time periods. The *P{Dcr}; P{Act5-GAL4*} line was selected due to ubiquitous expression of *GAL4* throughout all developmental time periods conferred by the *actin* driver. *GAL4* driver lines used for RNAi knockdown also contained a *UAS*-Dcr region to overexpress the dicer protein required for siRNA production from the dsRNA hairpins, in accordance to the published recommendations of Kaya-Copur and Schnorrer (2016). All flies were ordered from BDSC (Table 2). All flies were initially quarantined to prevent the potential spread of mold and/or disease into the existing Andrew Lab fly stocks prior to beginning RNAi.

Gene	Region	Basic Function	Average *p*-value
CG15630	Intron	Immunoglobulin-like	1.61E-05
Roe1	Intron	Neurogenesis	1.86E-05
bun	Intron	Cell differentiation Neural cell proliferation Mechanosensory and sensory projection	2.11E-05
Rbp6	Intron	Stem cell development RNA binding	3.03E-05
Mad	Downstream	Cell differentiation Leg morphogenesis Synaptic growth at neuromuscular junctions	4.15E-05
psh	Upstream	Defense response Immune system response	9.26E-05

Table 1: Information regarding genes containing the 6 top associated SNPs identified via a GWAS conducted by the DGRP website. All 6 candidate genes listed have at least 1 TRiP-RNAi transgene line available for order through BDSC.

BDSC 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
25708	*N/A*	*P{w^{+mC}=UAS-Dcr-2.D}1, w^{1118}; P{w^{+mC}=Act5C-GAL4}25FO1/CyO*	*P{Dcr}; P{Act5-GAL4}*	1(X);2
28322	*bun*	*y^{1}v^{1}; P{y$^{+t7.7}$ v$^{+t1.8}$=TRiP.JF02954}attP2*	*y^{1}v^{1} ; P{TRiP.JF02954}*	1(X);3
43183	*Mad*	*y^{1}sc*v^{1}; P{y$^{+t7.7}$ v$^{+t1.8}$=TRiP.GL01527}attP40*	*y^{1}sc*v^{1} ; P{TRiP.GL01527}*	1(X);2
35648	*Mad*	*y^{1}sc*v^{1}; P{y$^{+t7.7}$ v$^{+t1.8}$=TRiP.GLV21013}attP2/TM3, Sb^{1}*	*y^{1}sc*v^{1} ; P{TRiP.GLV21013}*	1(X);3
52877	*psh*	*y^{1}sc*v^{1}; P{y$^{+t7.7}$ v$^{+t1.8}$=TRiP.HMC03615}attP40*	*y^{1}sc*v^{1} ; P{TRiP.HMC03615}*	1(X);2
42589	*CG15630*	*y^{1}sc*v^{1}; P{y$^{+t7.7}$ v$^{+t1.8}$=TRiP.HMS02421}attP40*	*y^{1}sc*v^{1} ; P{TRiP.HMS02421}*	1(X);2
44060	*Roe1*	*y^{1}sc*v^{1}; P{y$^{+t7.7}$ v$^{+t1.8}$ =TRiP.HMS02777}attP40*	*y^{1}sc*v^{1} ; P{TRiP.HMS02777}*	1(X);2
57836	*Roe1*	*y^{1}v^{1}; P{y$^{+t7.7}$ v$^{+t1.8}$=TRiP.HMJ21846}attP40*	*y^{1}v^{1} ; P{TRiP.HMJ21846}*	1(X);2
61324	*Rbp6*	*y^{1}v^{1}; P{y$^{+t7.7}$ v$^{+t1.8}$=TRiP.HMJ23159}attP40*	*y^{1}v^{1} ; P{TRiP.HMJ23159}*	1(X);2

Table 2: All *GAL4* driver lines and TRiP-RNAi transgene lines obtained from BDSC for use in RNAi experiments. Short genotypes are used throughout the text.

RNAi Knockdown Crosses

The creation of specific mRNA knockdown offspring relied on crosses between *GAL4* lines and TRiP-*UAS* lines. Initially, *P{GawB}elav; P{Dcr}* lines were used to identify gene effects solely related to neuronal pathways. Crosses were not conducted in a particular order, but rather by health of the TRiP fly lines. TRiP lines that produced healthy stocks with large population sizes were used first, while other lines took more time to produce a sufficient population size for use in RNAi. Crosses were designed via the recommendations of Kaya-Copur and Schnorrer (2016). Namely, 5 virgin females carrying the *GAL4* driver were crossed with 10 males harboring the *UAS*-TRiP transgene to produce knockdown offspring in 25x95mm polystyrene *Drosophila* vials (Genesee Scientific, San Diego, CA). To collect a sufficient number of animals for

behavioral analysis, expansions were created for each cross by transferring all 15 flies into a new vial each day for 8 days. Cross 1 crossed *P{GawB}elav; P{Dcr}* virgin females with *y¹sc*v¹; P{TRiP.JF02954}* males to target the gene *bun* (Table 3). Cross 2 focused on the gene *Mad* by crossing *P{GawB}elav; P{Dcr}* virgin females with *y¹sc*v¹; P{TRiP.GL01527}* males (Table 4). Cross 3 targeted a different region of *Mad* by crossing *P{GawB}elav; P{Dcr}* virgin females with *y¹sc*v¹; P{TRiP.GLV21013}* males (Table 5). Finally, cross 4 targeted the gene *psh* with a cross of *P{GawB}elav; P{Dcr}* virgin females with *y¹sc*v¹ ; P{TRiP.HMC03615}* males (Table 6).

	Y¹v¹; P{TRiP.JF02954}	***>; P{TRiP.JF02954}***
P{GawB}elav; P{Dcr}	$\frac{P\{GawB\}elav}{y^1v^1};\frac{P\{Dcr\}}{+};\frac{P\{TRiP.JF02954\}}{+}$	$\frac{P\{GawB\}elav}{>};\frac{P\{Dcr\}}{+};\frac{P\{TRiP.JF02954\}}{+}$
P{GawB}elav; P{Dcr}	$\frac{P\{GawB\}elav}{y^1v^1};\frac{P\{Dcr\}}{+};\frac{P\{TRiP.JF02954\}}{+}$	$\frac{P\{GawB\}elav}{>};\frac{P\{Dcr\}}{+};\frac{P\{TRiP.JF02954\}}{+}$

Table 3: Cross 1 - *P{GawB}elav; P{Dcr}* virgin females x *y¹sc*v¹ ; {TRiP.JF02954}* males.

	y¹sc*v¹; P{TRiP.GLO1527}	***>; P{TRiP. GLO1527}***
P{GawB}elav; P{Dcr}	$\frac{P\{GawB\}elav}{y^1sc^*v^1};\frac{P\{Dcr\}}{P\{TRiP.GLO1527\}};\frac{+}{+}$	$\frac{P\{GawB\}elav}{>};\frac{P\{Dcr\}}{P\{TRiP.GLO1527\}};\frac{+}{+}$
P{GawB}elav; P{Dcr}	$\frac{P\{GawB\}elav}{y^1sc^*v^1};\frac{P\{Dcr\}}{P\{TRiP.GLO1527\}};\frac{+}{+}$	$\frac{P\{GawB\}elav}{>};\frac{P\{Dcr\}}{P\{TRiP.GLO1527\}};\frac{+}{+}$

Table 4: Cross 2 - *P{GawB}elav; P{Dcr}* virgin females x *y¹sc*v¹ ; P{TRiP.GL01527}* males.

	$y^1sc^*v^1$; P{TRiP.GLV21013}	$y^1sc^*v^1$; TM3,Sb^1	>; P{TRiP.GLV21013}	>; TM3,Sb^1
P{GawB}elav; P{Dcr}	$\frac{P\{GawB\}elav}{y^1sc^*v^1}; \frac{P\{Dcr\}}{+}; \frac{P\{TRiP.GLV21013\}}{+}$	$\frac{P\{GawB\}elav}{y^1sc^*v^1}; \frac{P\{Dcr\}}{+}; \frac{+}{TM3,Sb^1}$	$\frac{P\{GawB\}elav}{y^1sc^*v^1}; \frac{P\{Dcr\}}{+}; \frac{P\{TRiP.GLV21013\}}{+}$	$\frac{P\{GawB\}elav}{>}; \frac{P\{Dcr\}}{+}; \frac{+}{TM3,Sb^1}$
P{GawB}elav; P{Dcr}	$\frac{P\{GawB\}elav}{y^1sc^*v^1}; \frac{P\{Dcr\}}{+}; \frac{P\{TRiP.GLV21013\}}{+}$	$\frac{P\{GawB\}elav}{y^1sc^*v^1}; \frac{P\{Dcr\}}{+}; \frac{+}{TM3,Sb^1}$	$\frac{P\{GawB\}elav}{y^1sc^*v^1}; \frac{P\{Dcr\}}{+}; \frac{P\{TRiP.GLV21013\}}{+}$	$\frac{P\{GawB\}elav}{>}; \frac{P\{Dcr\}}{+}; \frac{+}{TM3,Sb^1}$

Table 5: Cross 3 - *P{GawB}elav; P{Dcr}* virgin females x *$y^1sc^*v^1$; P{TRiP.GLV21013}* males.

	$y^1sc^*v^1$; P{TRiP.HMCO3615}	>; P{TRiP. HMCO3615}
P{GawB}elav; P{Dcr}	$\frac{P\{GawB\}elav}{y^1sc^*v^1}; \frac{P\{Dcr\}}{P\{TRiP.HMCO3615\}}; \frac{+}{+}$	$\frac{P\{GawB\}elav}{>}; \frac{P\{Dcr\}}{P\{TRiP.HMCO3615\}}; \frac{+}{+}$
P{GawB}elav; P{Dcr}	$\frac{P\{GawB\}elav}{y^1sc^*v^1}; \frac{P\{Dcr\}}{P\{TRiP.HMCO3615\}}; \frac{+}{+}$	$\frac{P\{GawB\}elav}{>}; \frac{P\{Dcr\}}{P\{TRiP.HMCO3615\}}; \frac{+}{+}$

Table 6: Cross 4 - *P{GawB}elav; P{Dcr}* virgin females x *$y^1sc^*v^1$; P{TRiP.HMC03615}* males.

Collection of RNAi Knockdown Animals

Animal collection from the RNAi knockdown crosses followed the same protocols as those previously described for the DGRP lines so that grooming behavior could be compared between all experiments.

Behavior Recording and Analysis of RNAi Knockdown Lines

Data collection from RNAi knockdown crosses was conducted following the same protocols as those described above for the DGRP lines to allow for comparisons. Behavioral analysis of RNAi knockdown lines was also conducted according to the aforementioned protocols of the initial DGRP study.

Data Analysis of RNAi Knockdown Lines

Data were again extracted and analyzed using custom, in-house Perl scripts and custom, in-house MATLAB scripts similar to those used in the data analysis of DGRP lines. To identify pairwise statistical significance between RNAi knockdown lines, the nonparametric Wilcoxon Rank Sum test was applied ($p < 0.05$). One-way analysis of variance (ANOVA) was also applied to the RNAi knockdown lines among the 3 metrics of grooming to determine statistical significance among all 3 lines ($p < 0.05$).

Results

Quantitative Genetics of Grooming Behavior Among and Within DGRP Lines

The 10-minute video recordings provided a total of 808 individual fruit fly observations among 38 different DGRP lines, 34 of which had at least 8 males and 8 females, making them suitable for analysis (Table 7). All flies from the 34 separate lines were manually scored for grooming phenotypes using the video annotation program, VCode (Hagedorn *et al*., 2008). Analysis was conducted on all 34 different DGRP lines to show current results regarding grooming behavior based on various parameters including total grooming index, number of grooming bouts, and average length of grooming bouts.

Grooming index (GI) represents the total percentage of time each fly spent grooming during the 10-minute video. GIs from flies within each of the 34 lines were compared using box-and-whisker plots that show the median, interquartile range (IQR), and whiskers to the 10^{th} and 90^{th} percentiles (Figure 3). The spread of data supports that there is continuous variation both among and within DGRP lines, as expected for a quantitative trait. DGRP line 324 had the lowest median GI (0.7) and line 307 had the highest median GI (12.4), a comparison which is significantly different (Wilcoxon Rank-Sum Test $p = 1.7 \times 10^{-7}$) (Figure 3).

Another parameter of measurement is number of grooming bouts. A grooming bout was categorized as being a single period of grooming, from start to stop. Again, box-and-whisker plot comparisons show a continuous variation of grooming behavior among and within DGRP lines (Figure 4). The order from least to greatest number of grooming bouts, based on median, is not identical to that of GI, however, it is similar. DGRP line 324 had the lowest median for number of grooming bouts (median = 1),

while DGRP line 379 had the highest number of grooming bouts observed (median = 28), a significant difference (Wilcoxon Rank-Sum Test $p = 7.0 \times 10^{-8}$) (Figure 4). DGRP line 307, with the highest GI also resulted in one of the highest number of grooming bouts.

The third parameter of grooming behavior analyzed was the average length of grooming bouts. As with the others, box plots organized from low to high, based on median, show continuous variation both among and within DGRP lines (Figure 5). DGRP line 705 had the lowest median for average length of grooming bouts and DGRP line 307 had the highest. Although the spread of data does not look as drastic in comparison to GI or number of grooming bouts, there was a significant difference between average length of grooming bouts of the DGRP lines with the lowest and highest GI (Wilcoxon Rank-Sum Test $p < 1.0 \times 10^{-4}$) (Figure 6).

Such differences can be more easily visualized using ethograms constructed to compare grooming behavior of DGRP lines with the lowest and highest GI (Figure 7; Figure 8). Each row represents the behavior of an individual fly within the 10-minute video recording. Grey portions represent behaviors other than grooming, such as walking, standing, and falling, while black represent periods of grooming. The lowest scoring line, DGRP line 324, contains 3 flies that do not groom during the 10-minute observation period, 1 fly that grooms once, and 1 fly that grooms twice (Figure 7). Flies of the highest scoring line, DGRP line 307, display more frequent and extended grooming bouts, as represented by the thick black segments (Figure 8). Ethograms show that not only is there a large difference in behavior between DGRP lines, but there are also many differences within the same DGRP line.

Genome-Wide Association Study

Once all flies from the 34 completed DGRP lines had been manually scored, the mean GI for males and females of each line were submitted to the DGRP website (http://dgrp.gnets.ncsu.edu/) to gain access to preliminary GWAS results regarding SNPs associated with observed variation in grooming behavior (Table 8). Submission to the DGRP website provided all SNPs associated, but specifically, 35 SNPs considered to be top associations (Mackey *et al*., 2012). The top 9 SNP associations were identified based on the highest significance. All 9 had *p*-values less than $1x10^{-5}$, making them the most promising for future reverse genetic experiments (Figure 9).

Of the top 9 SNP associations, 6 were identified to be within protein coding genes that had TRiP-RNAi transgene stocks available for order from BDSC (Table 1). These 6 genes are known as the candidate genes resulting from the GWAS. The other 3 were located at transcription factor binding sites (TFBS), which although interesting, are not able to be studied using RNAi as there is no mRNA produced from TFBS regions that can be targeted for degradation. The gene with a TRiP-RNAi transgene stock available for order from BDSC, containing the SNP with the lowest *p*-value in an intronic region, is *CG15630* ($p = 1.61x10^{-5}$) (Figure 10). This gene may be immunoglobulin-like, but currently, is primarily of unknown function and is not yet fully characterized or named. Therefore, this study could reveal a novel discovery regarding the function of *CG15630* in *D. melanogaster* (FB2017_04). *Roe1* also was found to have a top associated SNP within an intronic region and is known to be involved in neurogenesis ($p=1.86x10^{-5}$) (Figure 11; FB2017_04). Bunched (*bun)* was found to have a top associated SNP within an intronic region and is involved in neuronal proliferation

and mechanosensory projection ($p = 2.11x10^{-5}$) (Figure 12; FB2017_04). Study of *bun* may lead to answers regarding how differences in sensory mechanisms may effect repetitive grooming behaviors. The fourth gene with a top associated SNP found in an intronic region is RNA-binding protein 6 (*Rbp6)* and is known to function in stem cell development ($p = 3.03x10^{-5}$) (Figure 13; FB2017_04). *Mad* differed in the fact the SNP was located downstream of the gene rather than in an intron ($p = 4.15x10^{-5}$) (Figure 14). Mothers against dpp (*Mad)* was selected for study because of its known involvement in cell differentiation, leg morphogenesis, and synaptic growth at neuromuscular junctions, all of which seem potentially relevant to grooming behavior (FB2017_04). Persephone (*psh)* was the final candidate gene selected for further study and had a top associated SNP located upstream ($p = 9.26x10^{-5}$) (Figure 15). This gene has a known function in the innate immune system (FB2017_04). It is possible that a correlation may be identified between grooming behavior and health, in regards to the removal of pathogens.

In Vivo RNAi Knockdown

To date, RNAi knockdown has been conducted on 3 of the top 6 candidate genes selected, *bun*, *Mad*, and *psh*. *CG15630*, *Roe1*, and *Rbp6* have not yet been studied due to time and experimental constraints. Data analysis of the grooming behavior observed for the 3 knockdown lines measured the same parameters of grooming as in the initial DGRP study: GI, number of grooming bouts, and average length of grooming bouts. Results show consistency between all 3 parameters of grooming, as *Mad* is consistently the lowest and *bun* is consistently the highest in all measured phenotypes.

GI of the RNAi knockdown lines shows statistically significant variation among

the 3 studied (ANOVA $p = 1.794 \times 10^{-7}$). TRiP_GLV21013_elav, which represents the RNAi knockdown line targeting *Mad* mRNA, has the lowest GI (median = 3.66). TRiP_JF02954_elav, which represents the RNAi knockdown line targeting *bun* mRNA, has the highest GI (median = 9.54). TRiP_HMCO3615_elav, which represents the RNAi knockdown line targeting *psh* mRNA, has a GI that falls in between the other 2 lines (median = 7.49). In all pairwise comparisons, results for GI were significantly different between these 3 RNAi experiments (see Figure 16 and associated *p*-values).

The parameter number of grooming bouts showed similar statistically significant variation among the 3 RNAi knockdown lines (ANOVA $p = 1.9376 \times 10^{-6}$). TRiP_GLV21013_elav (*Mad*) has the lowest number of grooming bouts (median = 10), TRiP_JF02954_elav (*bun*) has the highest number of grooming bouts (median = 22), and TRiP_HMCO3615_elav (*psh*) falls between the other 2 (median = 17). As with GI, pairwise comparisons indicate that each of these populations are significantly different (see Figure 17 and associated *p*-values).

Lastly, the average length of grooming bouts was compared to find the same order of RNAi knockdown lines. Variation among lines was also found to be statistically significant (ANOVA $p = 0.0412$). However, the differences between lines were not as large as other metrics, the lowest (TRiP_GLV21013, median = 2.319) and highest (TRiP_JF02954, median = 2.7795) (see Figure 18 and associated *p*-values). This is similar to the initial DGRP results in which the average length of grooming bouts did not show as much of a variation between lines (Figure 5).

In comparison to the GIs of the DGRP lines, TRiP_GLV21013_elav (*Mad*) shows similar results to those of the lowest grooming DGRP lines. The GIs of

TRiP_JF02954_elav (*bun*) shows similar results to those of the highest grooming DGRP lines. The GIs of TRiP_HMCO3615_elav (*psh*) do not seem to fall on either extreme end in comparison to the DGRP lines. Together, such results suggest that there is a difference between these lines suggesting that certain proteins do function to regulate grooming behavior. Specifically, a decrease in *Mad* expression corresponds to less grooming behavior, while a decrease in *bun* expression corresponds to heightened grooming behavior. Such differences can be more easily visualized with ethograms showing the behavior of individual flies within the 10-minute video recording (Figure 19).

Discussion

A Genetic Basis for Grooming Behavior in D. melanogaster

Given the continuous natural variation observed in the quantification of grooming behavior in this study, our results support the hypothesis that grooming behavior in *Drosophila melanogaster* is a quantitative trait. Many similarities are observed between the 3 measured parameters of grooming including total GI, number of grooming bouts, and average length of grooming bouts. The lowest scoring lines for total grooming index (GI) tend to also exhibit the lowest number of grooming bouts and average length of grooming bouts. For example, DGRP line 324 has the lowest median for GI and the lowest median for number of grooming bouts. DGRP line 307 had the highest medians for both GI and length of grooming bouts. Box-and-whisker plots between the different parameters show the same lines to be low scoring and the same lines to be high scoring, among all three metrics. As the GI is a function of the number and length of grooming bouts, these results indicate that both influence total grooming. Consistency in results support that there are particular DGRP lines that groom less than others. Such differences are predicted to be primarily due to genetic variation because the environment is strictly controlled for all tested DGRP lines. The only known differences between various DGRP lines are genetic. Therefore, phenotypic differences are most likely due to genetic variants, such as SNPs.

Overall, average length of time per grooming bout between DGRP lines, shows less variation than the other 2 parameters measured. This is supported by ethograms produced for the lowest and highest lines because the majority of grooming bouts in both tend to show black lines of similar width, representing similar amounts of time

spent grooming. Hence, the factor primarily influencing the GI of *D. melanogaster* seems to be the number of grooming bouts. However, there are other differences between the ethograms of the lowest and highest lines. Ethograms from DGRP line with the highest GI have multiple wide black segments, showing longer grooming bouts. The lower grooming lines appear to groom for only short time periods, very infrequently, and some not at all. Such disparate lines support the idea that grooming is both variable among and within DGRP lines, again, suggesting a varied genetic influence on the behavior.

Given promising results obtained in this study, which support a genetic basis for the natural variation of grooming behavior, averages for the GI of males and females for each line were submitted to the DGRP website for the conduction of a GWAS (Mackay *et al*., 2012). Top associations, which are of the largest interest, show that many of the SNPs associated with the variation in grooming are related to TFBSs and genes with functions that seem relevant to predicted pathways. TFBSs are expected to be commonly associated with GWASs because genetic variations within TFBSs can alter the capabilities of key protein/DNA interactions. If a given variant alters binding abilities of transcription factors, then the corresponding gene will be regulated differently, potentially leading to altered gene expression. Although these TFBS regions are commonly associated, they are difficult to study since they cannot be analyzed using RNAi knockdown because these regions do not directly produce mRNA. If the corresponding gene regulated by a given transcription factor was known, then that gene could be studied using RNAi. However, current technologies do not easily allow for the identification of every interaction, as transcription factors can regulate upstream or

downstream genes. Therefore, this factor presents a limitation to the study of TFBSs.

Due to such limitations, only SNPs located in known protein coding genes were selected for further study using RNAi. The first candidate gene to be analyzed was *bun*, which has a human ortholog known as TSC22 Domain Family Member 1 (*TSC22D1*) that functions to regulate development (OMIM #607715). Previous work suggests that bun has similar functions to those of TSC22D1 proteins and is related to the ability to affect transcription factor activity in regulating various pathways (FB2017_04). One of the more critical roles is in the *decapentaplegic* signaling pathway, which induces multiple cell fates early in development (Dobens *et al*., 2000). Specifically, within this pathway *bun* has been found involved in eye development and limiting the size of the operculum, a region of the egg, by integrating signals from decapentaplegic (DPP) and epidermal growth factor (EGF) (Dobens *et al*., 2000; Treisman *et al*., 1995). Another function of *bun* is the formation of dorsal appendages via interactions between growth factors. *bun*'s functioning in relation to dorsal appendages ensure proper dorsal cell fates and peripheral nervous system functioning (Dobens *et al*., 1997). It is not unlikely that the organization of developmental patterning can influence behaviors such as grooming because alterations could lead to variations in motor abilities caused by improper development of the motor or nervous systems in areas such as the appendages. Likewise, abnormalities in body part location may simply cause more or less grooming behavior based on physical abilities.

The bun protein has also been studied in regards to mushroom body formation (Kim *et al*., 2009). The mushroom body is a distinct invertebrate brain region specialized in sensory integration, primarily from olfactory stimuli, and learning (Ito *et al*., 1997).

Again, this is a possible cause of grooming behavior variation since differences in how sensory information is interpreted could alter the desire to clean via grooming. For example, if a certain olfactory cue typically induces grooming, a variation in the mushroom bodies could cause flies to no longer recognize the cue in the same way, consequently, causing less grooming. Genes such as *bun*, that have been found to influence a wide range of central functions are likely to effect a range of behaviors including our behavior of interest; however, the interpretation of exact mechanisms is consequently more challenging.

The second candidate gene analyzed in this study was *Mad*, a transcription factor involved in mediating responses to growth factors such as DPP (Xie & Spradling, 1998). These results suggest an accurate association to grooming since both *bun* and *Mad* have been found to be involved in the same pathways. Consistent results such as these suggest that the TGF beta signaling pathway and *decapentaplegic* pathway, which are both involved in early developmental patterning and growth, likely influence grooming behavior (Dobens *et al* 1997; Dworkin & Gibson, 2006; Marquez *et al*., 2001; Xie & Spradling, 1998). *Mad* mutants have been found to result in abnormal formation of various body parts including eyes, wings, and legs, similar to consequential phenotypes of *bun* mutants (Dobens *et al*., 1997; Dworkin & Gibson, 2006; Marquez *et al*., 2001; Treisman *et al*., 1995). Interestingly, *Mad* also positively regulates synaptic growth at neuromuscular junctions, as determined during a study analyzing acyl-CoA synthetase's (ACSL4) role in the development of X-linked neurodevelopmental disorders (Lui *et al*., 2014). Previous associations to human disorders make *Mad* an interesting candidate for study via RNAi. The human ortholog of *Mad* is Mothers against decapentaplegic

homolog 1 (*SMAD1*), which has an amino acid identify of 76% to its fly counterpart and plays similar roles in the same pathways within humans (OMIM #601595). Such similarities further indicate the potentials of studying *Mad* in this model organism.

The third candidate gene analyzed thus far is *psh*, which encodes a serine protease crucial to innate immune responses (El Chamy *et al*., 2008; Jang *et al*., 2006). Enzymatic functions of psh become active following microbial infection and indirectly activate the Toll pathway (Jang *et al*., 2006). An active Toll pathway cascade produces antimicrobial peptides and cytokines to fight infection (El Chamy *et al*., 2008). Similar conclusions have been made between the findings of various studies, suggesting that this gene surely affects a fly's ability to defend against microbial pathogens like bacteria and fungi (e.g. Buchon *et al*., 2009; El Chamy *et al*., 2008; Jang *et al*., 2006; Yamamoto-Hino & Goto 2016). Although the nervous system is the expected system primarily affecting grooming behavior, it is not surprising that the immune system potentially influences it as well. Pathways that identify and fight pathogens may play a role in regulating grooming behavior because it could function to prevent potential microbial infection. However, the mechanisms that may contribute are currently unknown and are of debate in the field (personal communication, Julie H. Simpson, University of California Santa Barbara). There may be a positive association between stimulation of the Toll pathway and enhanced grooming behavior to induce the removal of microbial pathogens to prevent further infection. Study of *psh* may lead to novel findings that directly relate these systems.

RNAi Functional Validation

Analysis from the 3 RNAi knockdown lines completed supports the hypothesis

that certain genes, identified via genome-wide association using the DGRP, play a role in altering metrics of grooming behavior. Specifically, results show that proteins encoded by candidate genes of high interest affect overall frequency of grooming. In assessing the consequential grooming behavior, potential pathways involved may be predicted.

DGRP lines that display increased numbers of grooming bouts and an overall higher grooming index may have genetic variants causing a heightened sensory response or an increase in motor neuron stimulation. Such responses would likely lead to the desire to groom more frequently. There are countless ways in which these pathways could be affected including variations in inhibitory or excitatory neurotransmitters, differences regarding neuronal connections/projections, and abnormalities in other related systems such as the immune system.

Results show that knockdown of *bun* mRNA led to a heightened grooming behavior as measured by GI, number of grooming bouts, and length of grooming bouts. This response is determined to be high based on comparisons to previous DGRP results and results of the other 2 RNAi knockdown lines. Based on known functions of *bun*, the mushroom bodies of the observed flies may have interpreted sensory information in a way that caused them to have a greater desire to groom (Kim *et al*., 2009). It is also possible that the lack of this protein led to atypical physical development involving the appendages and nervous system, which may have caused issues in the neuronal network responsible for regulating this motor behavior (Dobens *et al*., 2000; Treisman *et al*., 1995). An increase in grooming behavior has previously been associated with models of human neurodevelopmental disorder presentations such as

restricted and repetitive movements (Tauber *et al*., 2011).

RNAi knockdown of *Mad* expression led to an overall decrease in grooming behavior in comparison to the other 2 RNAi knockdown lines and the DGRP lines. Based on this consequential behavior and the known function of *Mad*, it is reasonable to infer that there was decreased synaptic growth at the neuromuscular junctions, a prediction that could be verified anatomically in future studies (Lui *et al*., 2014). Decreased innervation of muscles could potentially manifest as decreased motor behavior in general, observed in our study as reduced grooming. Being that there is a human ortholog of *Mad*, it is possible that errors in the orthologous gene may lead to motor impairment in humans as well.

Decreased expression of *psh* also seemed to lead to a slight decrease in grooming behavior, rather than an increase. The grooming behavior did not seem to be as extreme as the other 2 RNAi knockdown lines, however, it seems comparable to that of DGRP lines in the lower half of the continuous variation. A decrease in grooming associated with decreased expression of *psh* could be related to the activation of the Toll pathway discussed above. If psh is not present, then the Toll pathway may not be activated correctly in response the presence of microbial pathogens (El Chamy *et al*., 2008). It is likely that this pathway is typically involved in stimulating grooming behavior and that without proper activation of Toll, this does not occur and less grooming behavior is observed than normal.

Improvements and Future Work

To further, support the idea that observed alterations in grooming behavior are due to knockdown of mRNA products, behavioral analysis of TRiP-RNAi lines, without

crosses with the *GAL4* driver, should also be conducted as a negative control. This is because there may be differences within the genetic background between the TRiP-RNAi lines and the DGRP lines utilized in our studies. In order to control for these genetic differences that may ultimately influence phenotypic expression of grooming we would like, in the future, to establish baseline grooming metrics for the parental *GAL4* and *UAS* lines. This would allow a more direct comparison of changes in grooming behavior upon RNAi knockdown. However, our results still show significant differences between the 3 RNAi knockdown lines, which is most likely associated with the decrease in gene expression due to the fact that these TRiP-RNAi lines are derived from a relatively recent homozygous experimental stock designed specifically for the development of these lines (Hu *et al.*, 2016).

In conducting crosses for RNAi, Cross 2 did not produce animals that could be studied using our behavioral analysis. The intention of Cross 2 was to induce knockdown of *Mad* expression. In this cross, both male and female parental (P0) flies selected did not have curly wings or a known *CyO* mutant allele. However, resulting offspring (F1) all displayed curly wings. Further research into the TRiP-RNAi transgene lines did not help elucidate these aberrant and unexpected results. The only information explaining the Curly phenotype is a statement on the Harvard website: “may be segregated by CyO” (https://fgr.hms.harvard.edu/fly-in-vivo-rnai). Details regarding why the F1 generation in Cross 2 had curly wings is currently unknown, but may relate to incomplete penetrance of the Curly phenotype and variable, temperature-sensitive expressivity. Further information is required prior to using these segregated RNAi-TRiP lines in our studies on grooming. Flies with physical abnormalities, including curly wings,

were excluded from all previous data collection because their mutant phenotypes likely represent a confounding variable. Upon realizing some lines had this unknown segregating factor, we chose not to initially work with 5 other TRiP-RNAi lines that were received for this study, since they were also listed as possibly being segregated by CyO. This left the 3 healthy lines discussed above for use in RNAi due to the fact that they had known genotypes and no inconclusive phenotypes. One of the lines used was an alternative TRiP-RNAi line for the same gene, *Mad*, so it could be studied regardless of the first attempt that produced curly winged animals.

In the future, these segregated lines should be researched to understand why the F1 generation results in curly winged flies to attempt crosses that produce wild-type wings so that behavior can be analyzed. This may require contacting a professional from the Harvard Medical School Department of Genomics or BDSC to confirm genotypes of the 5 remaining lines so that they too can be used in future experiments. In addition to conducting *in vivo* RNAi on remaining lines, all 8 could also be crossed with the *P{Dcr}; P{Act5-GAL4}* driver to compare phenotypic consequences to those crossed with the *P{GawB}elav; P{Dcr}* line. If resulting grooming behavior is drastically altered in only the *P{GawB}elav; P{Dcr}* line, then it is most likely that the nervous system is the origin of the phenotypic variation caused by a particular gene as the *elav* driver is expressed only throughout the nervous system. If resulting grooming behavior is drastically altered in only the *P{Dcr}; P{Act5-GAL4}* line, then the effect of a gene is likely not predominantly affecting the nervous system as this particular *actin* driver is ubiquitously expressed. Lastly, if results between both driver lines cause the same drastic changes in grooming behavior, there are likely multiple interacting pathways

involved. Through this additional comparison, more detailed conclusions about gene function can be inferred.

Another next step relates to the fact that only a subset of the DGRP lines were used for this study in order to show that grooming behavior is in fact a quantitative trait appropriate for GWAS and further research of particular genes. Over 150 DGRP lines are available for use in a continuation in order to determine most accurately which SNPs are highly associated with grooming behavior. If more lines were to be used, an automated scoring process would be beneficial to increase the efficiency of analysis. Technology may initially only allow a program to identify times within the 10-minute videos where flies are not walking. By doing so, it would allow for portions within the videos, where flies are surely not grooming, to be eliminated to in turn decrease the duration of time spent scoring each fly. Currently, the Andrew Lab has obtained access to EthoVision video tracking software and another student is working to appropriately program it for our needs (Noldus *et al*., 2001).

Another improvement to be made to this experiment involves the conduction of genotypic and proteomic confirmation of RNAi knockdown. Due to time limitations this has not already been done. To ensure that the TRiP-RNAi lines are properly leading to degradation of mRNA products from our GOI we could first conduct reverse transcription real time polymerase chain reaction (RT-PCR or qPCR). This technique monitors the amplification of a specific transcript during PCR by using a fluorescent dye. A control gene, typically *GapDH* because it is expressed highly in all cell types is used. qPCR measures based on amplifications necessary to reach a given threshold based on the control. The lower this measurement, the more highly a gene is expressed. We

should be able to conduct qPCR for offspring of the RNAi crosses and compare gene expression for the GOI to that of wild-type flies of the same age. The goal of this method is to see a significant decrease in gene expression since mRNA for our target genes should be degraded by RISC.

A more accurate confirmation would be obtained through Western blotting because it focuses on the actual levels of protein produced. In this case, protein levels should be significantly decreased in our experimental flies compared to wild-types because mRNA is not present in the same amounts for translation into protein. This method works by tagging proteins extracted from the flies with antibodies in order to only identify specific proteins of interest. As a result of the antibody tags, the proteins can be separated based on size via a protein gel and later visualized on a film. This approach can be difficult and more time-consuming and thus, is less ideal for our research.

Although RNAi knockdown is an efficient tool for reverse genetic studies in regard to time and money, it does have flaws. The efficiency of gene silencing is often inconsistent and may affect phenotypic outcomes, possibly causing false negatives or positives (Yamamoto-Hino & Goto, 2013). There is also the potential of false positives when conducting RNAi due to similar sequences between an off target mRNA and the dsRNA associated with RISC. An emerging genome editing tool, Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-CRISPR associated protein (Cas9), may offer more precise and accurate gene targeting technique. This native bacterial defense system has been modified for laboratory use through the design of a programmable guide RNA (gRNA) (Doudna and Charpentier 2014). The gRNA is

complementary to a chosen target site located within a gene of interest. Here, at a particular location the gRNA, consisting of CRSIPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA), direct the Cas9 endonuclease to cut double stranded DNA (dsDNA). Endogenous cellular mechanisms facilitate DNA repair following dsDNA breaks. One mechanism repairs DNA imprecisely by non-homologous end joining (NHEJ), which commonly causes indels and consequently, leads to frameshift mutations. This mechanism can be employed with CRSIPR-Cas9 to produce null mutations. This is unlike RNAi in the fact that the CRISPR system directly alters the genomic sequence and can lead to complete gene inactivation. It is possible that RNAi can be used as a preliminary tool to screen for interesting genes and then follow-up studies can be conducted using CRISPR-Cas9 to analyze gene function based on complete gene knockout.

Similarly, CRISPR-Cas9 can be used to conduct allelic replacement of various DGRP lines. This method would work to replace DGRP lines containing a minor allele at a specific SNP location with the reference allele. Resulting grooming behavior could be analyzed to reveal whether or not replacement of the minor allele altered the grooming behavior in a specific direction. This information would provide conclusions regarding the effects of variation at a SNP. This would differ from RNAi studies and CRISPR-Cas9 null knockouts because they both rely on a decrease in gene expression rather than a site specific mutation. It is likely that grooming behavior would differ between an intragenic mutation and complete loss of gene expression.

Conclusion

Overall, this study provides sufficient preliminary results in support of a continuation of this work. One of the main goals of the research was to establish a solid foundation for future work in the Andrew Lab at Lycoming College. Completion of these experiments, which resulted in a successful GWAS and RNAi knockdown, offer a set of functional methods to interrogate behavior in *Drosophila*. Results suggest that this work is reasonable for undergraduate students and creates a basis for new independent and group experiments. Thus far, work with the DGRP shows that grooming behavior is a quantitative trait that shows continuous variation among and between lines. Such information led to the identification of genomic regions associated with the phenotype of interest and continued analysis supports the involved of 3 specific genes. In particular, expression levels of *Mad*, *bun*, and *psh* have been found to effect grooming behavior. The methodology of the study seems to be accurate because the candidate genes studied all have been associated with mechanisms plausibly related to grooming. Although current results are preliminary, they show that using this genetic system to study grooming behavior has potential for novel discoveries related to understanding how specific genetic variants impact the function of nervous systems and immune systems. These findings may be related to human neurodevelopmental disorders given the previous connections to grooming behavior.

Bloomington Stock ID	DGRP Designation	Males		Females		Total
		M		F		
25174	DGRP_208	DGRP_208_M	11	DGRP_208_F	18	29
25175	DGRP_301	DGRP_301_M	8	DGRP_301_F	9	17
25176	DGRP_303	DGRP_303_M	12	DGRP_303_F	9	21
25177	DGRP_304	DGRP_304_M	12	DGRP_304_F	10	22
25179	DGRP_307	DGRP_307_M	10	DGRP_307_F	14	24
25180	DGRP_313	DGRP_313_M	17	DGRP_313_F	9	26
25181	DGRP_315	DGRP_315_M	14	DGRP_315_F	10	24
25182	DGRP_324	DGRP_324_M	9	DGRP_324_F	9	18
25183	DGRP_335	DGRP_335_M	10	DGRP_335_F	10	20
25184	DGRP_357	DGRP_357_M	15	DGRP_357_F	10	25
25185	DGRP_358	DGRP_358_M	9	DGRP_358_F	12	21
25186	DGRP_360	DGRP_360_M	9	DGRP_360_F	10	19
25187	DGRP_362	DGRP_362_M	0	DGRP_362_F	0	0
25188	DGRP_375	DGRP_375_M	11	DGRP_375_F	8	19
25189	DGRP_379	DGRP_379_M	14	DGRP_379_F	15	29
25190	DGRP_380	DGRP_380_M	14	DGRP_380_F	8	22
25191	DGRP_391	DGRP_391_M	12	DGRP_391_F	9	21
25192	DGRP_399	DGRP_399_M	8	DGRP_399_F	16	24
25193	DGRP_427	DGRP_427_M	9	DGRP_427_F	11	20
25194	DGRP_437	DGRP_437_M	10	DGRP_437_F	12	22
25195	DGRP_486	DGRP_486_M	12	DGRP_486_F	11	23
25197	DGRP_517	DGRP_517_M	12	DGRP_517_F	13	25
25198	DGRP_555	DGRP_555_M	10	DGRP_555_F	11	21
25199	DGRP_639	DGRP_639_M	9	DGRP_639_F	2	11
25200	DGRP_707	DGRP_707_M	10	DGRP_707_F	12	22
25201	DGRP_712	DGRP_712_M	11	DGRP_712_F	9	20
25202	DGRP_730	DGRP_730_M	0	DGRP_730_F	0	0
25203	DGRP_732	DGRP_732_M	14	DGRP_732_F	10	24
25204	DGRP_765	DGRP_765_M	2	DGRP_765_F	2	4
25205	DGRP_774	DGRP_774_M	13	DGRP_774_F	9	22
25206	DGRP_786	DGRP_786_M	9	DGRP_786_F	10	19
25207	DGRP_799	DGRP_799_M	10	DGRP_799_F	12	22
25208	DGRP_820	DGRP_820_M	12	DGRP_820_F	12	24
25209	DGRP_852	DGRP_852_M	10	DGRP_852_F	13	23
25210	DGRP_859	DGRP_859_M	9	DGRP_859_F	9	18
25445	DGRP_365	DGRP_365_M	14	DGRP_365_F	12	26
25744	DGRP_705	DGRP_705_M	10	DGRP_705_F	8	18
25745	DGRP_714	DGRP_714_M	9	DGRP_714_F	13	22
						Total = 767

Table 7: Spreadsheet showing all flies collected and video recorded for behavioral analysis by DGRP line designation. Flies not suitable for analysis have been excluded prior to the construction of this table. Green boxes show completed lines with at least 8 males and 8 females. Yellow boxes show lines with too few flies for analysis. Red boxes show lines with no flies collected.

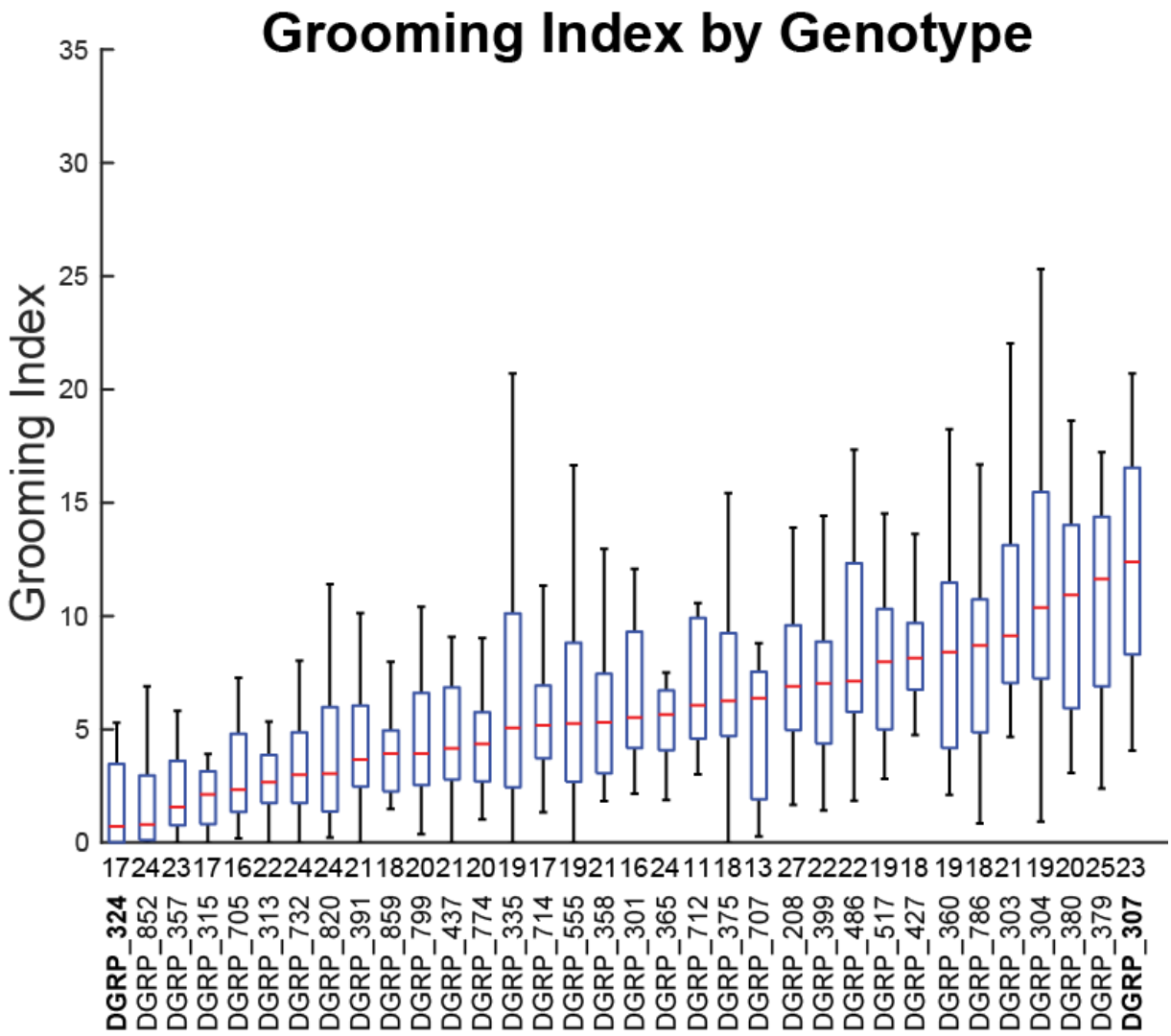

Figure 3: Grooming Index represents total time spent grooming. DGRP lines organized based on median from lowest to highest. *n* values are displayed above line designations on the x-axis. Box-and-whisker plots show the IQR (25th-75th percentiles) in the box with the red line at the median and whiskers extending from the 10th to 90th percentiles.

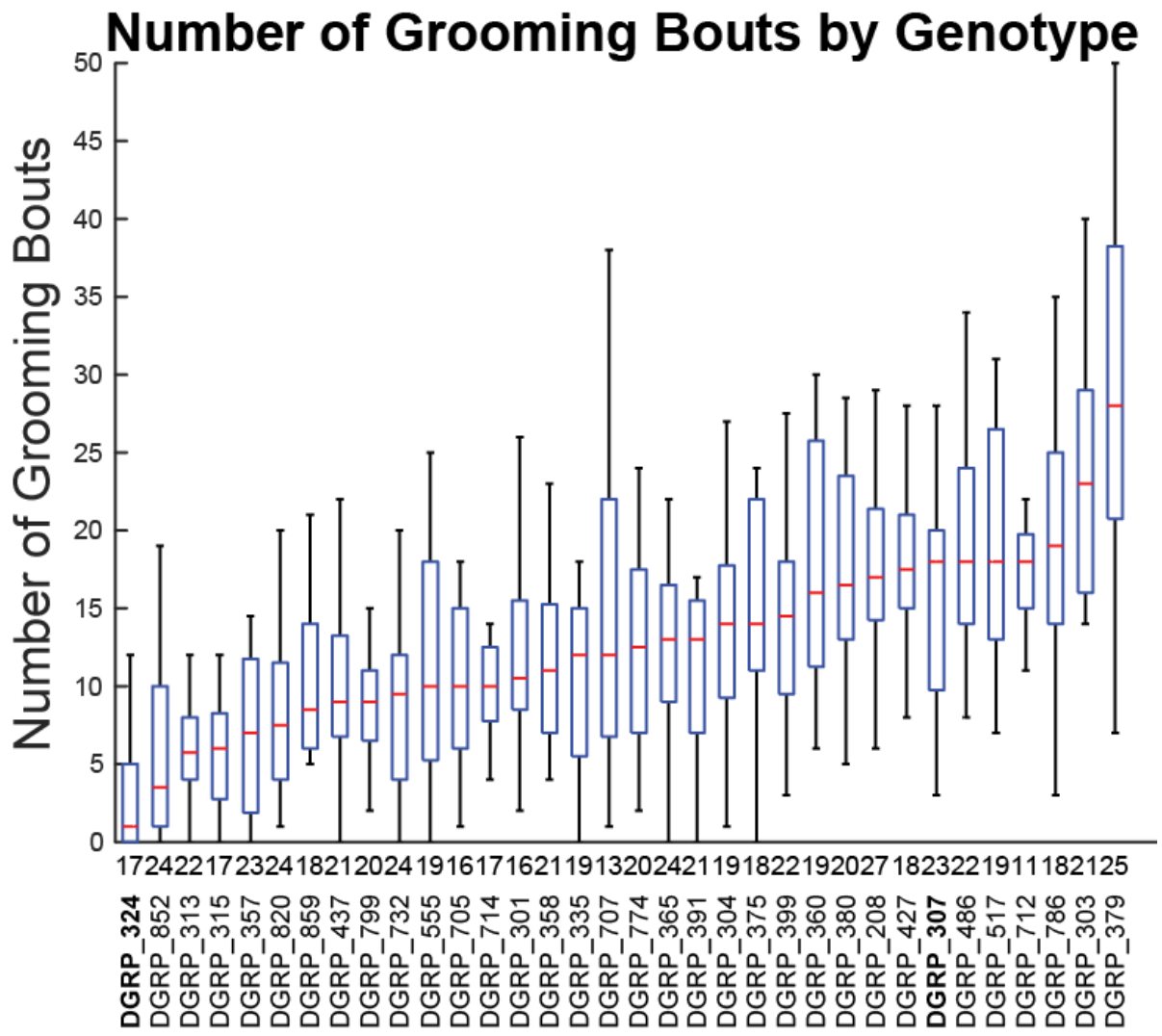

Figure 4: Number of separate grooming bouts, organized lowest to highest based on median. *n* values displayed above DGRP line designations on the x-axis Box-and-whisker plots show the IQR (25^{th}-75^{th} percentiles) in the box with the red line at the median and whiskers extending from the 10^{th} to 90^{th} percentiles.

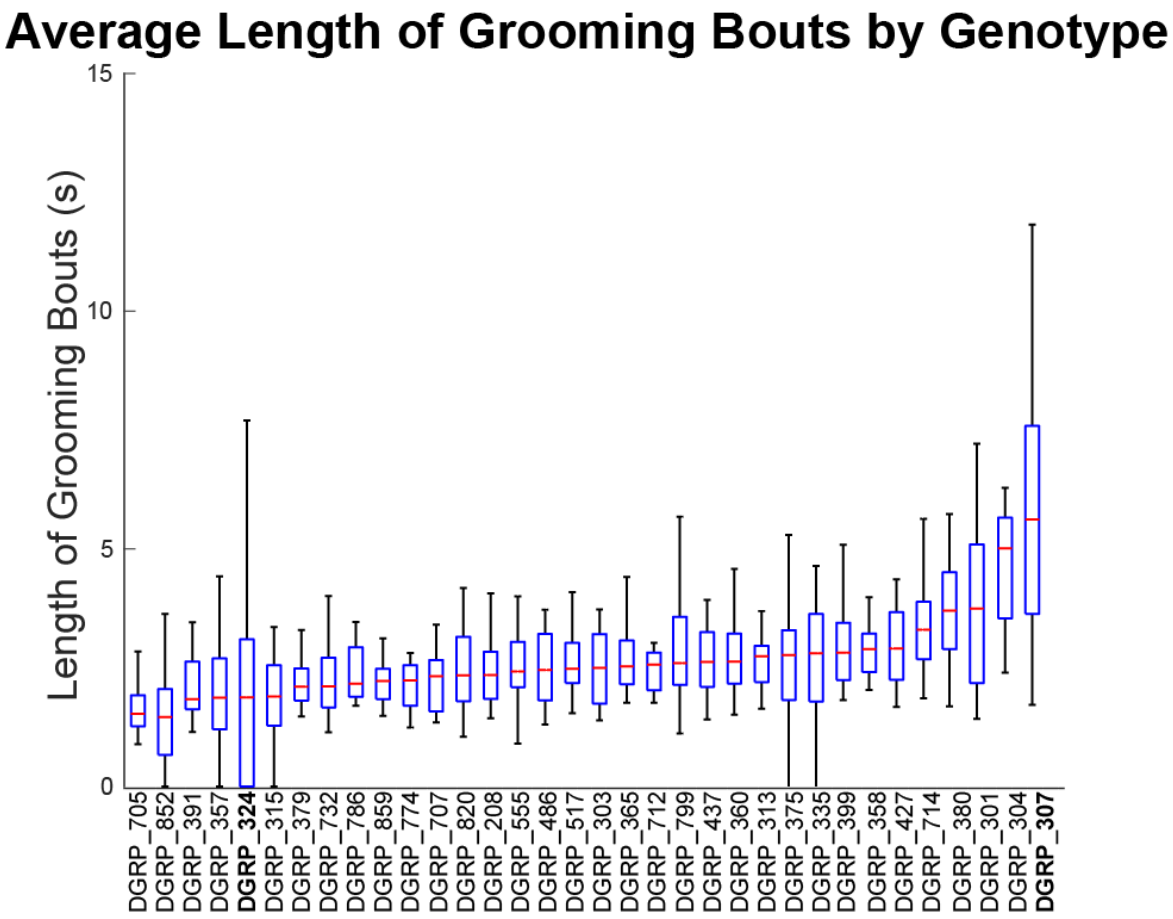

Figure 5: Average length of grooming bouts for each DGRP line, organized lowest to highest based on median. Box-and-whisker plots show the IQR (25^{th}-75^{th} percentiles) in the box with the red line at the median and whiskers extending from the 10^{th} to 90^{th} percentiles.

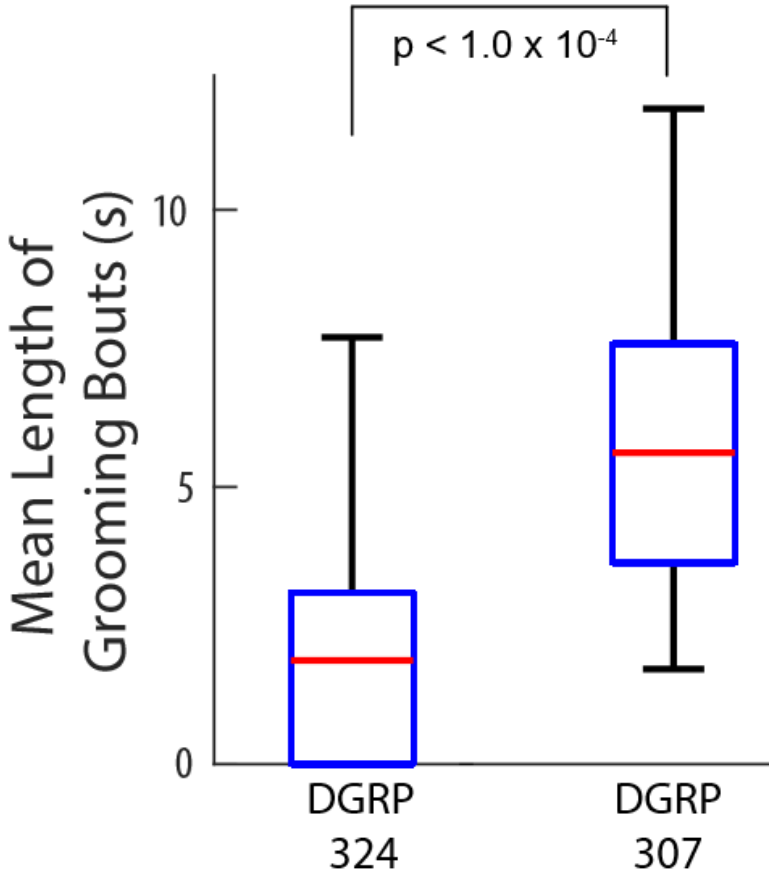

Figure 6: Average grooming bout length of flies from the 2 most disparate DGRP lines for the parameter GI. Box-and-whisker plots show the IQR (25th-75th percentiles) in the box with the red line at the median and whiskers extending from the 10th to 90th percentiles. Wilcoxon Rank-Sum Test used to determine *p*-value.

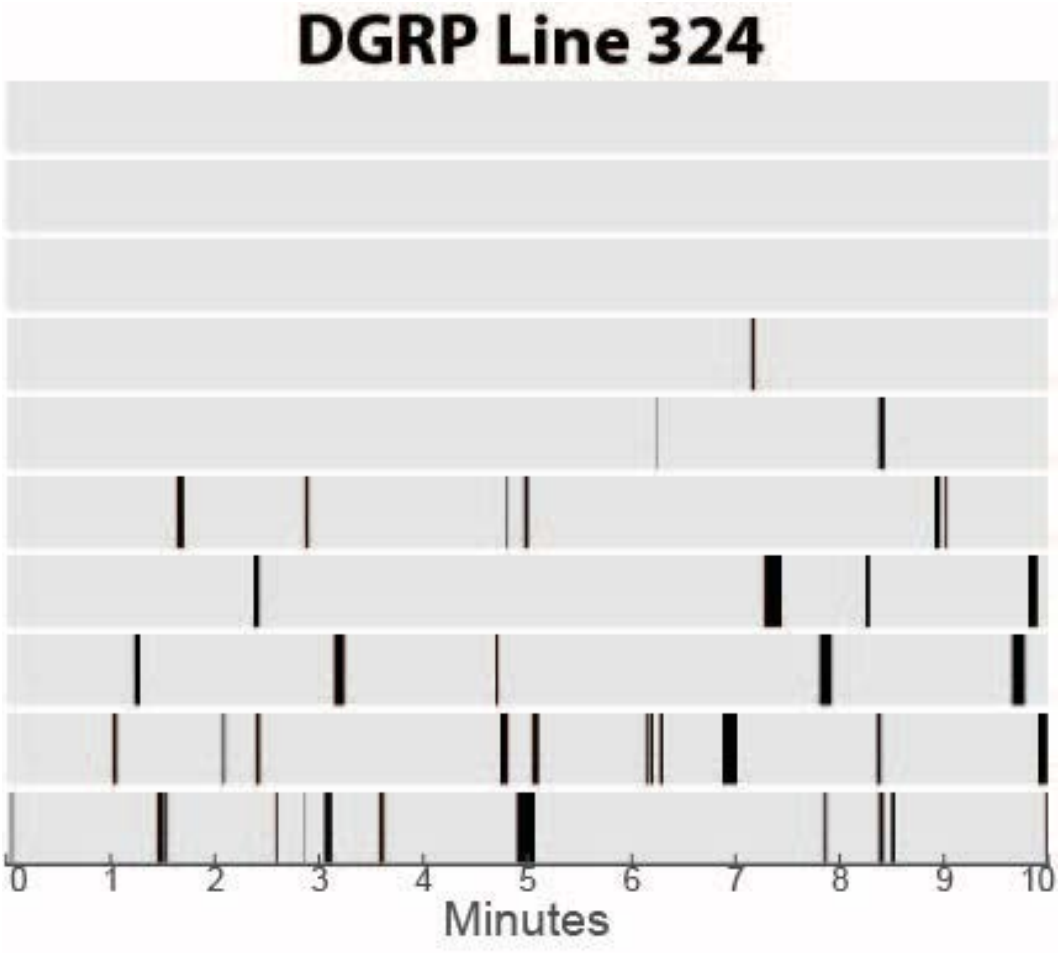

Figure 7: Ethograms from 10 individual flies from the DGRP line with the lowest median GI (DGRP 324). Each row represents an individual fly's behavior (Black: Grooming; Grey: Other) during the 10-minute observation period.

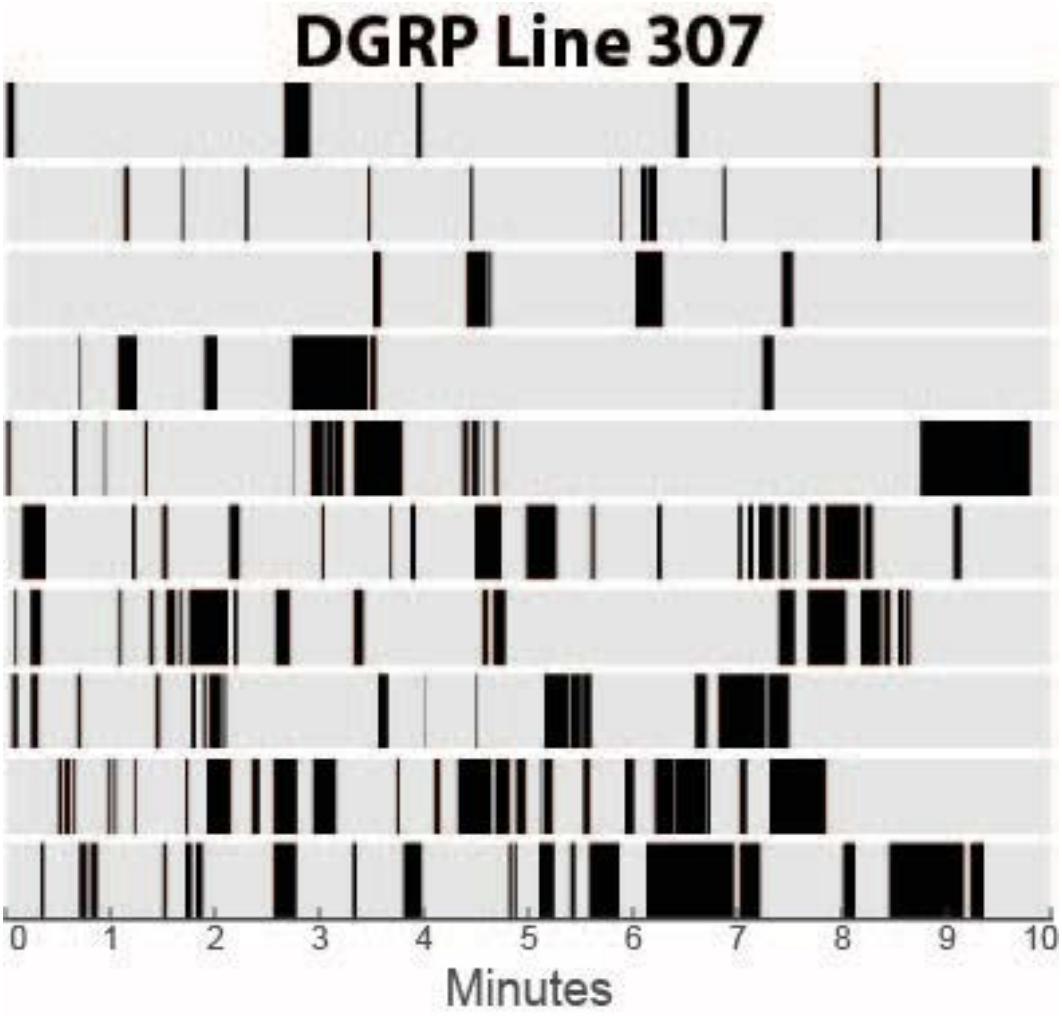

Figure 8: Ethograms from 10 individual flies from the DGRP line with the highest median GI (DGRP 307). Each row represents an individual fly's behavior (Black: Grooming; Grey: Other) during the 10-minute observation period.

DGRP Line	Mean GI Males	Mean GI Females
208	7.242	7.285
301	6.036	6.582
303	11.544	8.831
304	13.088	9.391
307	11.814	14.811
313	2.544	3.195
315	2.136	2.093
324	0.517	2.839
335	8.506	4.140
357	2.104	5.994
358	5.218	5.968
360	8.099	8.516
365	5.498	8.060
375	6.580	8.941
379	12.766	11.633
380	9.706	11.762
391	3.396	6.056
399	5.776	7.360
427	8.228	8.516
437	4.645	4.816
486	7.648	8.851
517	6.540	10.193
555	5.057	6.289
705	2.643	3.449
707	6.151	4.737
712	7.896	8.885
714	7.396	6.551
732	3.415	3.283
774	3.458	5.593
786	6.583	10.381
799	3.976	5.863
820	3.593	4.138
852	0.657	2.980
859	3.768	4.386

Table 8: Mean GI for the males and females of each DGRP line, used for submission of results to the DGRP website.

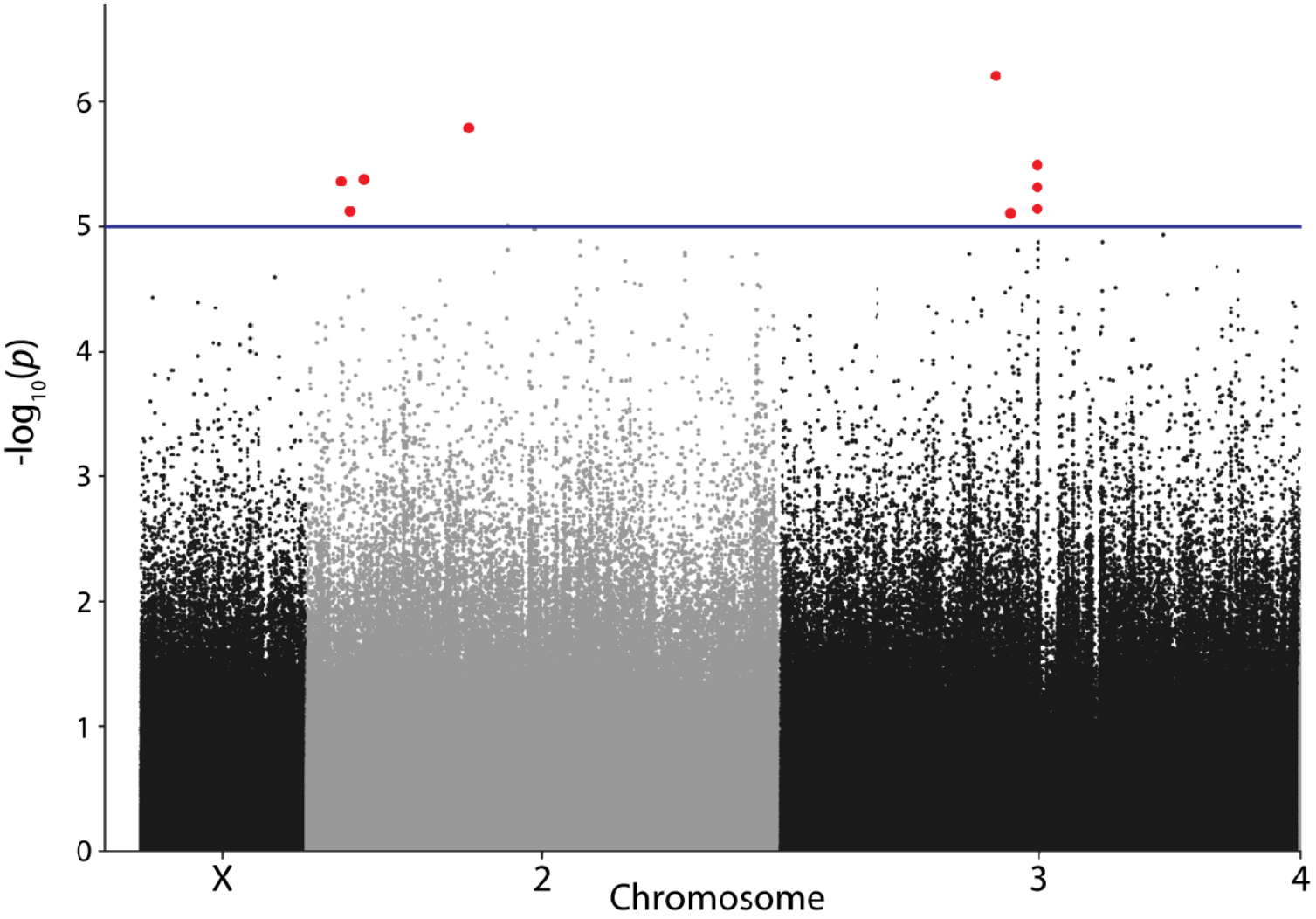

Figure 9: Manhattan plot displaying negative logarithm of the associated *p*-value for each SNP across the genome using results provided by the GWAS conducted through the DGRP website (Mackay *et al*., 2012). Lower *p*-values (higher y-values) indicate a stronger association between that particular SNP and the measured phenotype, grooming behavior. Red dots represent SNP associations of highest significance ($p < 10^{-5}$).

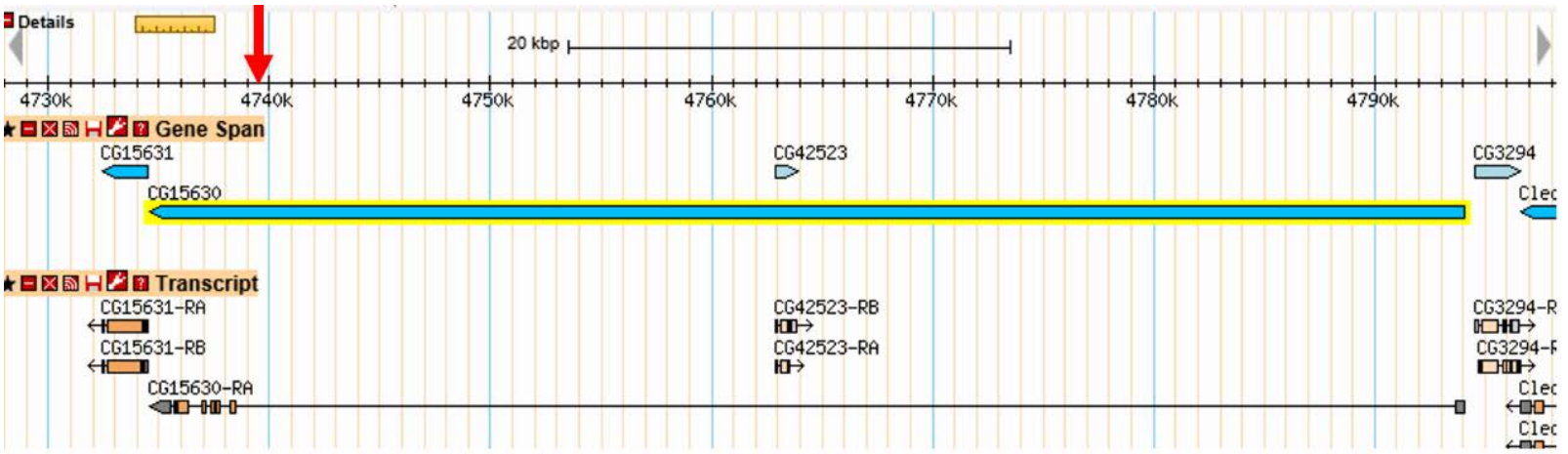

Figure 10: FlyBase view of *CG15630* location in *D. melanogaster* genome. Location of associated SNP identified with red arrow (FB2017_04).

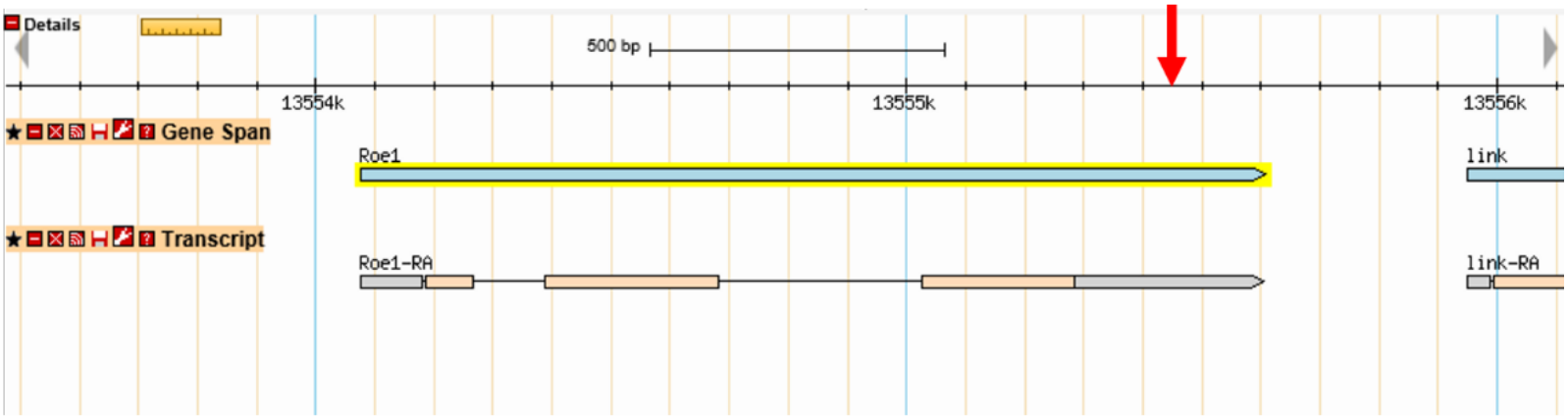

Figure 11: FlyBase view of *Roe1* location in *D. melanogaster* genome. Location of associated SNP identified with red arrow (FB2017_04).

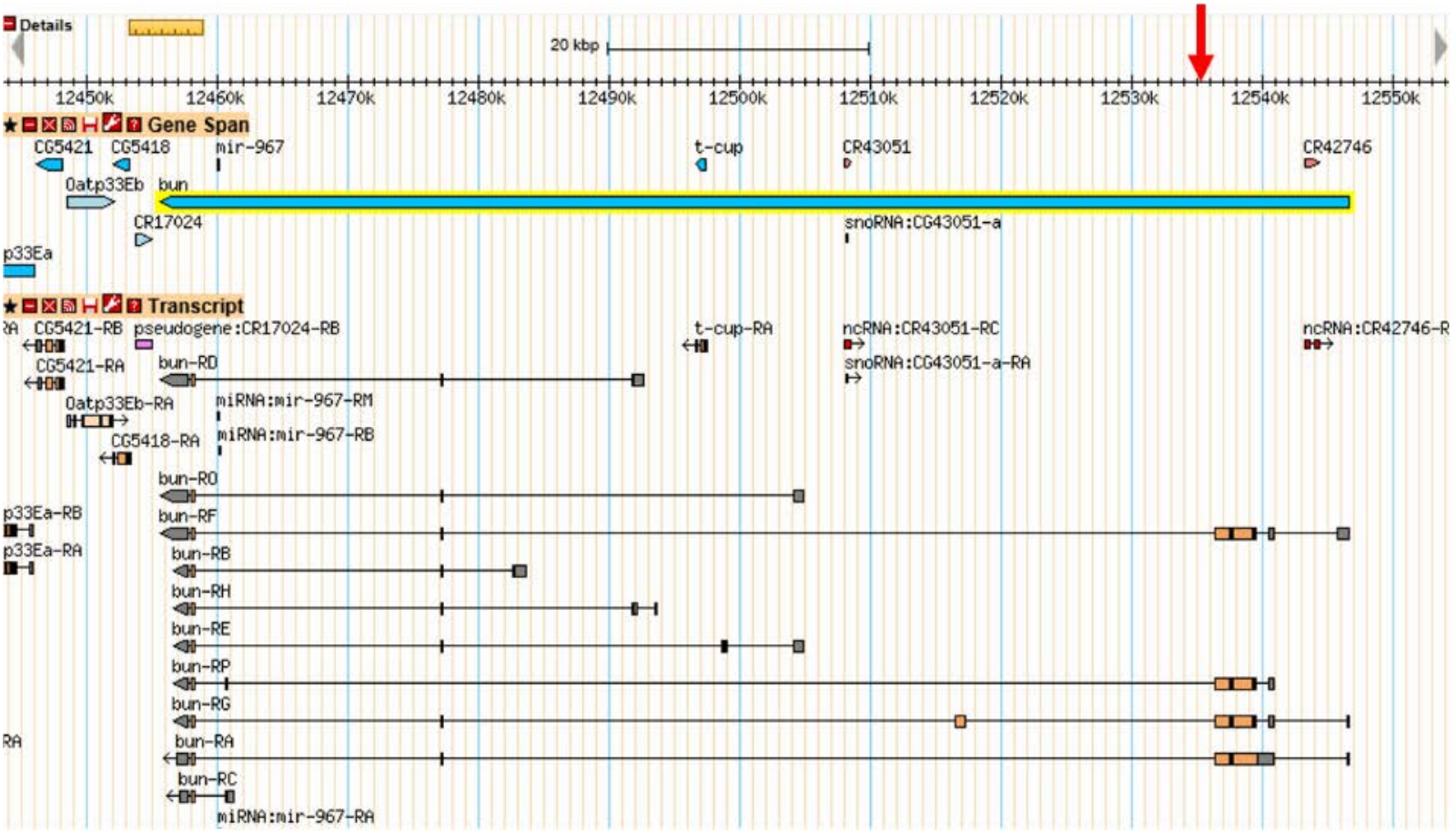

Figure 12: FlyBase view of *bun* location in *D. melanogaster* genome. Location of associated SNP identified with red arrow (FB2017_04).

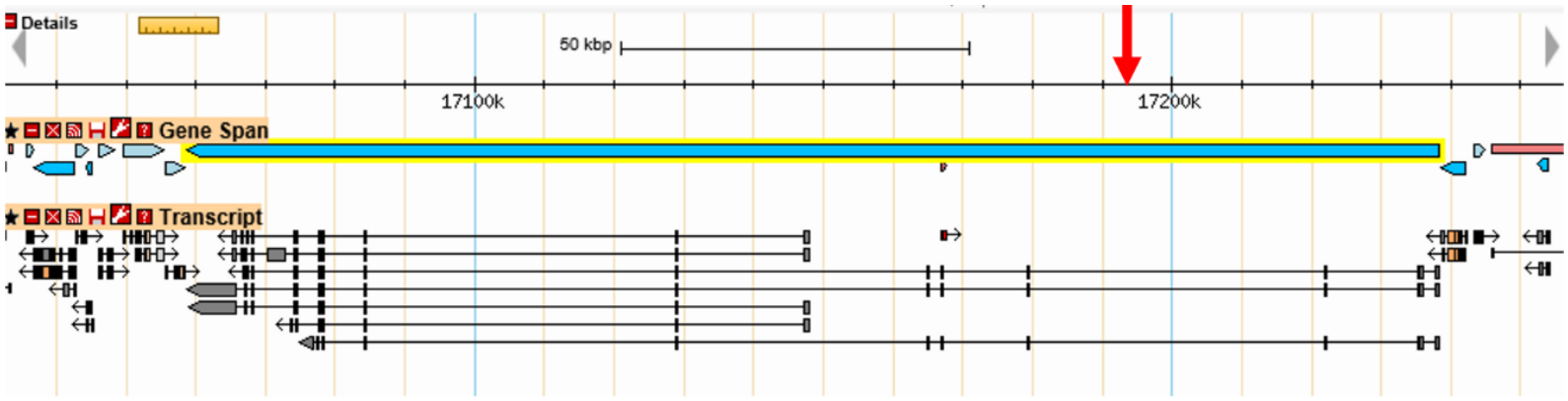

Figure 13: FlyBase view of *Rbp6* location in *D. melanogaster* genome. Location of associated SNP identified with red arrow (FB2017_04).

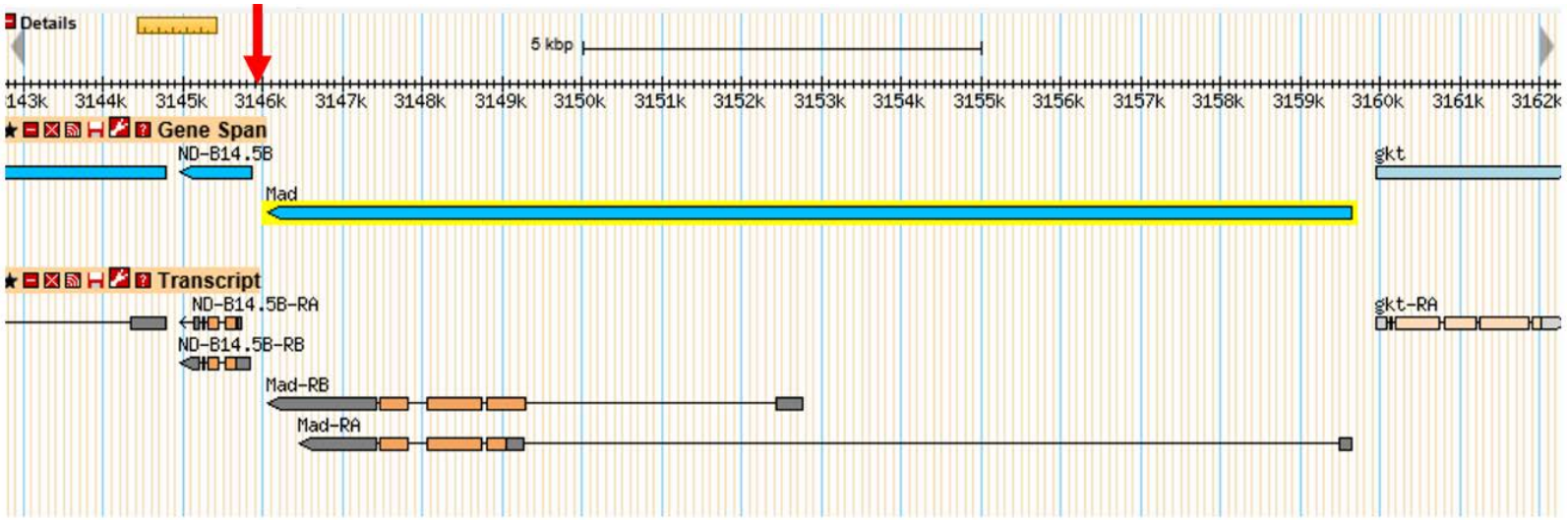

Figure 14: FlyBase view of *Mad* location in *D. melanogaster* genome. Location of associated SNP identified with red arrow (FB2017_04).

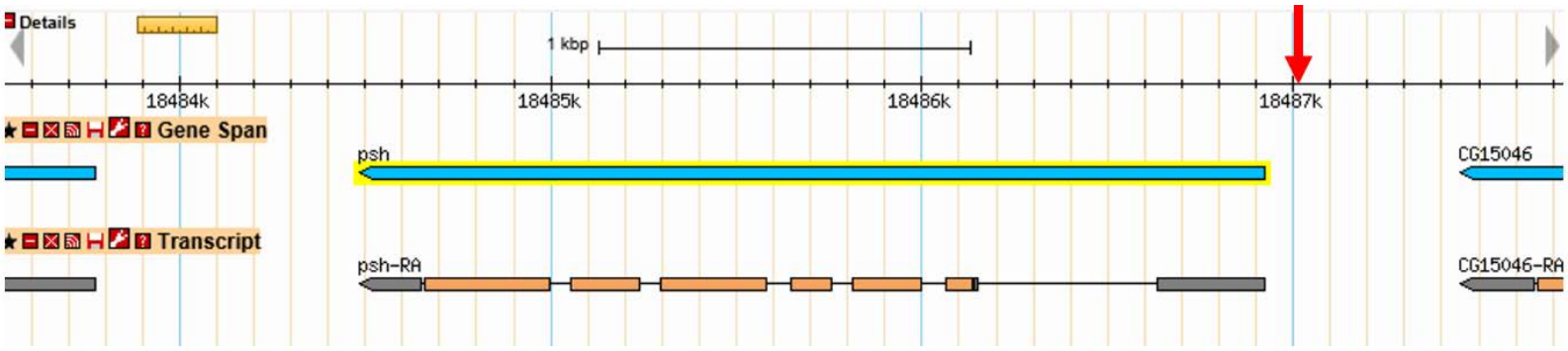

Figure 15: FlyBase view of *psh* location in *D. melanogaster* genome. Location of associated SNP identified with red arrow (FB2017_04).

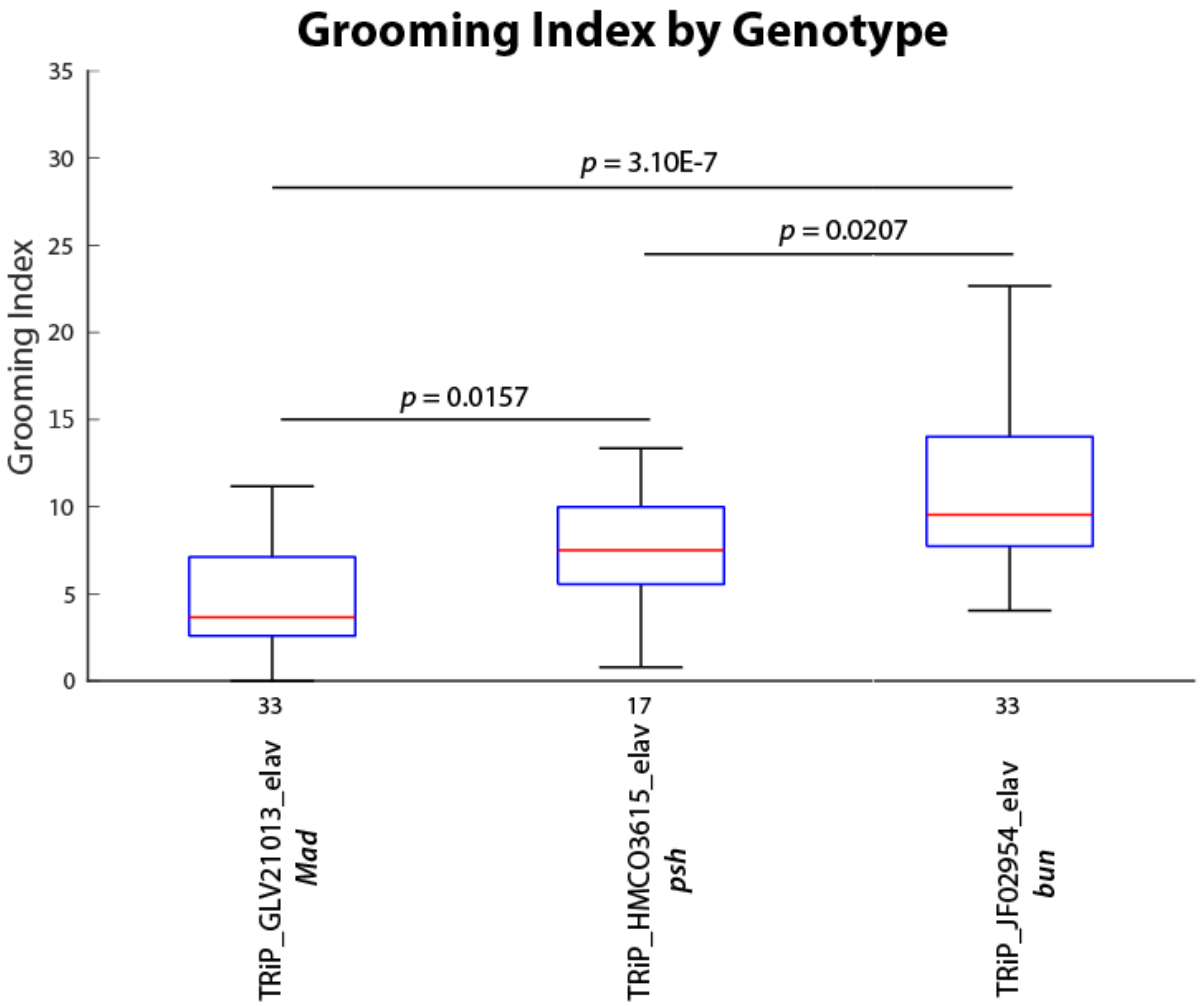

Figure 16: Grooming Index represents total time spent grooming. Organized based on median from lowest to highest. *n* values are displayed above line designations on the x-axis. Box-and-whisker plots show the IQR (25th-75th percentiles) in the box with the red line at the median and whiskers extending from the 10th to 90th percentiles. Wilcoxon Rank-Sum Test used to determine *p*-values.

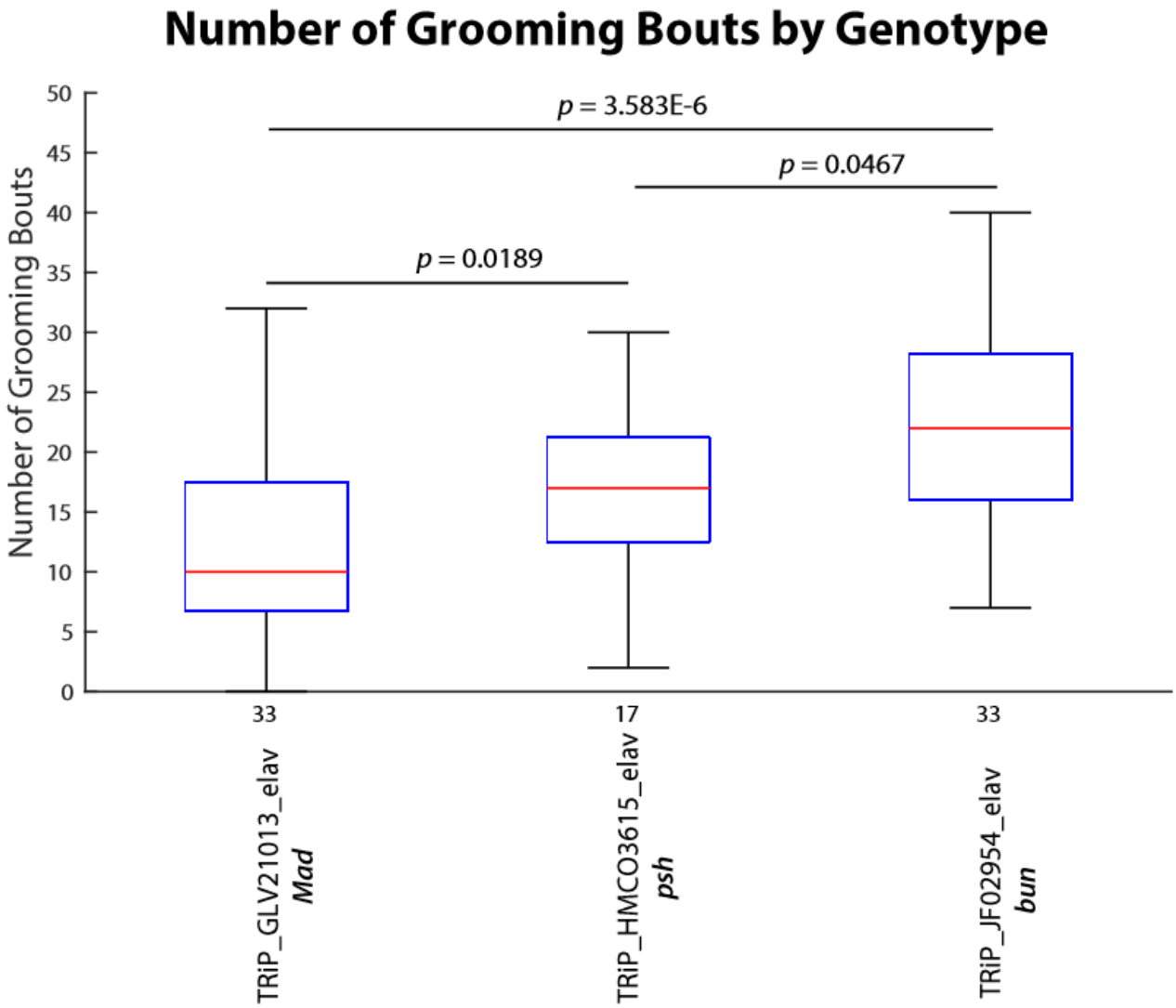

Figure 17: Number of separate grooming bouts, organized lowest to highest based on median. *n* values displayed above line designations on the x-axis. Box-and-whisker plots show the IQR (25th-75th percentiles) in the box with the red line at the median and whiskers extending from the 10th to 90th percentiles. Wilcoxon Rank-Sum Test used to determine *p*-values.

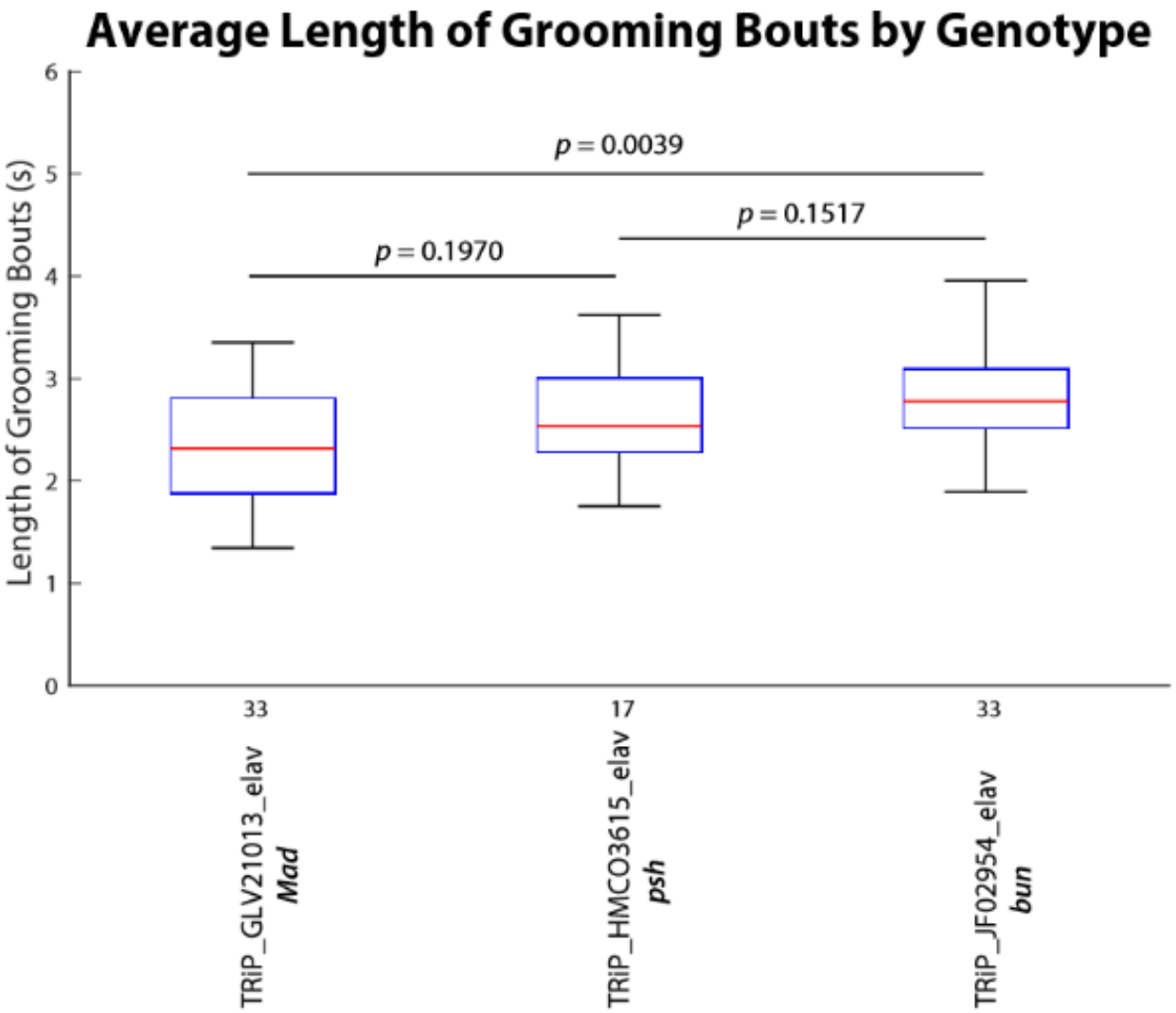

Figure 18: Average length of grooming bouts, organized lowest to highest based on median. *n* values displayed above line designations on the x-axis. Box-and-whisker plots show the IQR (25th-75th percentiles) in the box with the red line at the median and whiskers extending from the 10th to 90th percentiles. Wilcoxon Rank-Sum Test used to determine *p*-values.

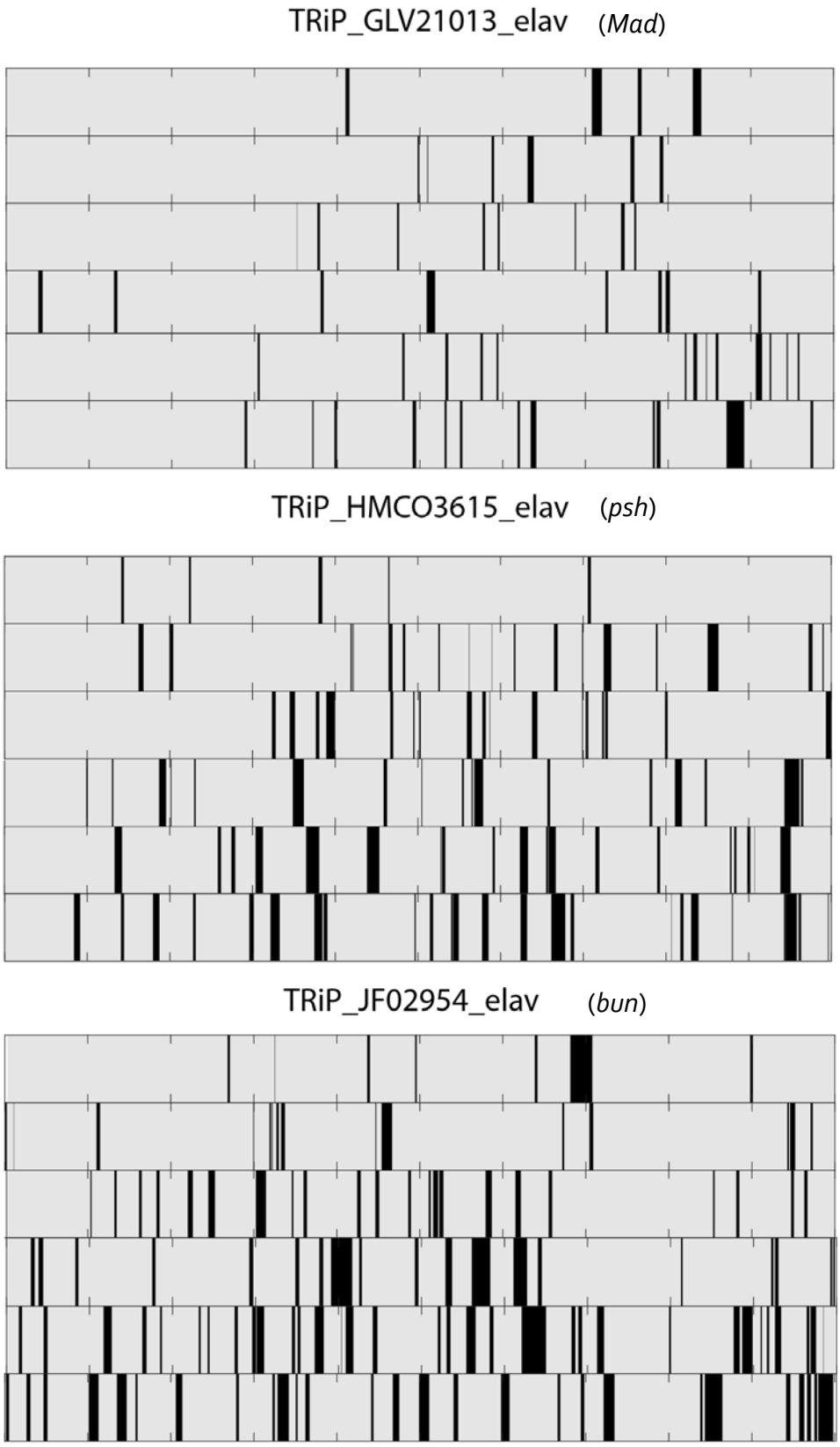

Figure 19: Ethograms from 6 individual flies from each completed RNAi line. Each row represents an individual fly's behavior (Black: Grooming; Grey: Other) during the 10-minute video.

References

American Psychiatric Association (2013) Neurodevelopmental Disorders. *Diagnostic and statistical manual of mental disorders* (5th ed.) Washington, DC, pp. 50-64.

America's Children and the Environment (2015) Neurodevelopmental Disorders. *EPA*.

Bagni, C., & Greenough, W.T. (2005) From mRNP trafficking to spine dysmorphogenesis: the roots of fragile x syndrome. *Nat Rev Neurosci* **6**, 376-387.

Buchon, N., Poidevin, M., Kwon, H.M., Guillou, A., Sottas, V., Lee, B.L., et al. (2009) A single modular serine protease integrates signals from pattern-recognition receptors upstream of the *Drosophila* Toll pathway. *Proc Natl Acad Sci* **106**, 12442-12447.

Dobens, L.L., Hsu, T., Twombly, V., Gelbart, W.M., Raftery, L.A., & Kafators, F.C. (1997) The *Drosophila* bunched gene is a homologue of the growth factor stimulated mammalian TSC-22 sequence and is required during oogenesis. *Mech Dev* **65**, 197-208.

Dobens, L.L., Peterson, J.S., Treisman, J., & Raftery, L.A. (2000) *Drosophila* bunched integrated opposing DPP and EGF signals to set the operculum boundary. *Development* **127**, 745-754.

Doudna, J.A., & Charpentier, E. (2014) The new frontier of genome engineering with CRISPRCas9. *Science*. DOI: 10.1126/science.1258096.

Dwarkin, I., & Gibson, G. (2006) Epidermal growth factor receptor and transforming growth factor signaling contributes to variation for wing shape in *Drosophila melanogaster*. *Genetics* **173**, 1417-1431.

El Chamy, L., Leclerc, V., Caldelari, I., & Reichhart, J.M. (2008) Sensing of 'danger signals' and pathogen-associated molecular patterns defines binary signaling pathways 'upstream' of Toll. *Nat Immunol* **9**, 1165-1170.

Fisherbrand®. Thermo Fisher Scientific.

Friedman, S.S., Dani, N., Rushton, E. & Broadie, K. (2013) Fragile X mental retardation protein regulates trans-synaptic signaling in *Drosophila*. *Disease Models & Mechanisms* **6**, 1400-1413.

Gao, F-B. (2002) Understanding Fragile X Syndrome: Insights from retarded flies. *Neuron* **34**, 859-862.

Garlapow, M.E., Huang, W., Yarboro, M.T., Peterson, K.R. & Mackay T.F.C. (2015) Quantitative genetics of food intake in *Drosophila melanogaster*. *PLoS one* **10**, 1-25.

Genesee Scientific. San Francisco, CA.

Gottlieb, D.J., DeStefano, A.L., Foley, D.J., Mignot, E., Redline, S., Givelber, R.J., et al. (2004) APOE epsilon4 is associated with obstructive sleep apnea/hypopnea: the sleep heart health study. *Neurology* **63**, 664-668.

Gramates, L.S., Marygold, S.J., Santos, G., Urbano, J.M, Antonazzo, G. Matthews, B.B. *et al*. (2017) The FlyBase Consortium.

Greenspan, R.J. (2004) *Fly Pushing* (2nd ed.) Cold Spring Harbor, NY, pp. 18-21.

Hagedorn, J., Hailpern, J., & Karahalios, K.G. (2008) VCode and VData: Illustrating a new framework for supporting the video annotation workflow. *AVI Extend Abst*.

Harbison, H.T., McCoy, L.J. & Mackay, T.F.C. (2013) Genome-wide association study of sleep in *Drosophila melanogaster*. *BMC Genomics* **14**, 1-18.

Heitzer, A.M., Roth, A.K., Nawrocki, L., Wrenn, C.C. & Valdovinos, M.G. (2013) Brief report: Altered social behavior in isolation-reared Fmr1 knockout mice. *J Autism Dev Disord* **43**, 1452-1458.

Huang, W., Richards, S., Carbone, M.A., Zhu, D., Anholt, R.R.H., Ayroles, J.F., et al. (2012) Epistasis dominates the genetic architecture of *Drosophila* quantitative traits. *Proc Natl Acad Sci USA* **109**, 15553-15559.

Huang, W., Massouras, A., Inoue, Y., Peiffer, J., Ramia, M., Tarone, A. M., et al. (2016) Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res* **24**, 1193-1208.

Hu, Y., Comjean A., Roesel C., Vinayagam A., Flockhart I., Zirin J., et al. (2016) FlyRNAi.org—the database of the *Drosophila* RNAi screening center and transgenic RNAi project: 2017 update. *Nucleic Acids Res* **45**, 672-678.

Ito, K., Awano, W., Suzuki, K., Hiromi, Y., & Yamamoto, D. (1997) The *Drosophila* mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurons and glial cells. *Development* **124**, 161-177.

Jang, I-H., Chosa, N., Kim, S-H., Nam, H-J., Lemaitre, B., Ochiai, M., et al. (2005) A spätzle-processing enzyme required for Toll signaling activation in *Drosophila* innate immunity. *Dev Cell* **10**, 45-55.

Jin, P., Zarnescu, D.C., Ceman, S., Nakamoto, M., Mowrey, J. & Jongens, T.A., et al. (2004) Biochemical and genetic interaction between the fragile X mental retardation protein and the microRNA pathway. *Nat Neurosci* **7**, 113-117.

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

Kays, I., Cvetkovska, V. & Chen, B.E. (2014) Structural and functional analysis of single neurons to correlate synaptic connectivity with grooming behavior. *Nat Protoc* **9**, 1-10.

Kim, J., Lee, S., Hwang, M., Ko, S., Min, C., & Kim-Ha, J. (2009) *bunched* specifically regulated a/b mushroom body neuronal cell proliferation during metamorphosis. *Neuroscience* **161**, 46-52.

Kornum, B.R, Kawashima, M., Faraco, J., Lin, L., Rico, T.J., Hesselson, S., et al. (2011) Common variants in P2RY11 are associated with narcolepsy. *Nat Genet* **43**, 66-71.

Liu, Z., Huang, Y., Hu, W., Huang, S., Wang Q., Han, J., et al. dAcsl, the *Drosophila* ortholog of Acyl-CoA synthetase long-chain family member 3 and 4, inhibits synapse growth by attenuating bone morphogenetic protein signaling via endocytic recycling. *J Neurosci* **34**, 2785-2796.

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

Marquez, R.M., Singer, M.A., Takaesu, N.T., Waldrip, W.R., Kraytsberg, Y., & Newfeld, S.J. (2001) Transgenic analysis of the smad family of TGF-B signal transducers in *Drosophila melanogaster* suggests new roles and new interactions between family members. *Genetics* **157**, 1639-1648.

McNaughton, C.H., Moon, J., & Strawderman, M.S. (2008) Evidence for social anxiety and impaired social cognition in a mouse model of Fragile X Syndrome*. Behav Neurosci* **122**, 293-300.

Mines, M.A., Yuskaitis, C.J., King, M.K., Beurel, E., & Jope, R.S. (2010) GSK3 Influences social preference and anxiety-related behaviors during social interaction in a mouse model of fragile X syndrome and autism*. PLoS one* **5**, e9706.

Noldus, L.P.J.J., Spink, A.K., & Tegelenbosch, R.A.J. (2001). EthoVision: A versatile video tracking system for automation of behavioral experiments. *Behavior Research Methods, Instruments, & Computers* **33**, 398-414.

Nutri-Fly™ BF Food Packages. Genesee Scientific. San Francisco, CA.

Seeds, A.M., Ravbar, P., Chung, P., Hampel, S., Midgley, F.M. & Mensh, B.D. (2014) A suppression hierarchy among competing motor programs drives sequential grooming in *Drosophila*. *eLife* **3**, 1-23.

Tauber, J.M., Vanlandingham, P.A. & Zhang, B. (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.

Treisman, J.E., Lai, Z-C., & Rubin, G.M. (1995) *shortsighted* acts in the *decapentaplegic* pathway in *Drosophila* eye development and has homology to a mouse TGF-B-responsive gene. *Development* **121**, 2835-2845.

Wan, L., Dockendorff, T.C., Jongens, T.A. & Dreyfuss, G. (2000) Characterization of dFMR1, a *Drosophila melanogaster* homolog of the Fragile X mental retardation protein. *Mol Cell Biol* **20**, 8536-8547.

Xie, T. & Spralding, A.C. (1998) *decapentaplegic* is essential for the maintenance and division of germline stem cells in the *Drosophila* ovary. *Cell* **94**, 251-260.

Yamamoto-Hino, M. & Goto, S. (2013) In vivo RNAi-based screens: Studies in model organisms. *Genes* **4**, 646-665.

Yamamoto-Hino, M. & Goto, S. (2016) Spätzle-processing enzyme- independent activation of the Toll pathway in *Drosophila* Innate Immunity*. Cell Struct Funct* **41**, 55-60.

Zhang, Y.Q., Bailey, A.M., Matthies, H.J.G., Renden, R.B., Smith, M.A., Speese, S.D., et al. (2001) *Drosophila* Fragile X-Related gene regulates the MAP1B homolog Futsch to control synaptic structure and function. *Cell* **107**, 591-603.