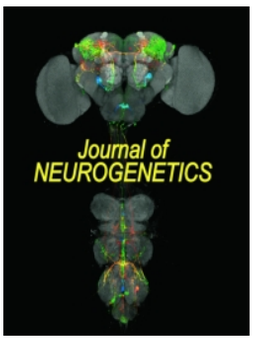


Journal of Neurogenetics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ineg20

The genetic architecture of larval aggregation behavior in *Drosophila*

Ross M. McKinney, Ryan Valdez & Yehuda Ben-Shahar

To cite this article: Ross M. McKinney, Ryan Valdez & Yehuda Ben-Shahar (2021): The genetic architecture of larval aggregation behavior in *Drosophila*, Journal of Neurogenetics, DOI: 10.1080/01677063.2021.1887174

To link to this article: https://doi.org/10.1080/01677063.2021.1887174

JOURNAL OF NEUROGENETICS
https://doi.org/10.1080/01677063.2021.1887174

Taylor & Francis
Taylor & Francis Group

ORIGINAL RESEARCH ARTICLE

The genetic architecture of larval aggregation behavior in *Drosophila*

Ross M. McKinney, Ryan Valdez and Yehuda Ben-Shahar

Department of Biology, Washington University in St. Louis, St. Louis, MO, USA

ABSTRACT

Many insect species exhibit basal social behaviors such as aggregation, which play important roles in their feeding and mating ecologies. However, the evolutionary, genetic, and physiological mechanisms that regulate insect aggregation remain unknown for most species. Here, we used natural populations of *Drosophila melanogaster* to identify the genetic architecture that drives larval aggregation feeding behavior. By using quantitative and reverse genetic approaches, we have identified a complex neurogenetic network that plays a role in regulating the decision of larvae to feed in either solitude or as a group. Results from single gene, RNAi-knockdown experiments show that several of the identified genes represent key nodes in the genetic network that determines the level of aggregation while feeding. Furthermore, we show that a single non-coding variant in the gene *CG14205*, a putative acyltransferase, is associated with both decreased mRNA expression and increased aggregate formation, which suggests that it has a specific role in inhibiting aggregation behavior. Our results identify, for the first time, the genetic components which interact to regulate naturally occurring levels of aggregation in *D. melanogaster* larvae.

ARTICLE HISTORY
Received 5 September 2020
Accepted 18 January 2021

KEYWORDS
Drosophila melanogaster; fruit fly; vinegar fly; foraging; sociality

Introduction

Group formation is one of the simplest forms of social interaction exhibited by individual animals. However, the genetic and physiological mechanisms underlying group formation are largely unknown for most species. *Drosophila melanogaster* larvae form simple cooperative group aggregates while feeding, which has been hypothesized to increase their fitness by providing defense against predation, as well as enabling individuals to communally digest food substrates more easily (Prokopy & Roitberg, 2001; Sokolowski, 2010; Wu *et al.*, 2003). Previous studies have suggested that in *Drosophila* and several other insect species, the formation and maintenance of larval aggregation is primarily regulated by the chemosensory detection of aggregation pheromones, as well as other sensory modalities (Leonhardt, Menzel, Nehring, & Schmitt, 2016; Louis & de Polavieja, 2017; Rooke, Rasool, Schneider, & Levine, 2020; Steiger & Stokl, 2017; Symonds & Wertheim, 2005; Thibert, Farine, Cortot, & Ferveur, 2016). Specifically, in *D.melanogaster*, at least two pheromones produced by larvae have been shown to act as chemoattractants (Mast, De Moraes, Alborn, Lavis, & Stern, 2014). However, the downstream neural and genetic pathways that regulate larval aggregation behavior remain largely unexplored.

The decision of individual larvae on whether to aggregate while feeding is likely regulated by the interplay between attractive and repulsive signals directly emitted by other conspecifics, or indirectly via feeding-related chemical changes of the consumed food. Indeed, it has been shown that food patch choice is influenced by the presence of other larvae, and the decision to choose one food patch over another is a function of group size (Durisko & Dukas, 2013; Lihoreau, Clarke, Buhl, Sumpter, & Simpson, 2016) and genetics (Allen, Anreiter, Neville, & Sokolowski, 2017; Fitzpatrick, Feder, Rowe, & Sokolowski, 2007; Kaun, Hendel, Gerber, & Sokolowski, 2007; Kaun, Riedl, et al., 2007). However, although some conserved peptidergic signaling pathways have been shown to regulate aggregation in *Drosophila* larvae (Wu *et al.*, 2003), most signals and downstream neuronal and genetic pathways that regulate group size via attractive and repulsive signals, remain unknown.

Understanding the genetic architecture that underlies insect aggregation is important not only for deciphering the biological principles that drive social decision making in general but would provide insight into means of offsetting the economic impact of insect pests. To address this important question, we used the *Drosophila* genetic reference panel (DGRP; Mackay *et al.*, 2012) to identify genetic variations associated with the extent of larval feeding aggregate size. By combining a genome-wide, quantitative genetics approach with single gene manipulations, we have identified several key genes that contribute to group size in natural populations of *Drosophila* larva. Our results highlight the utility of *D. melanogaster* for understanding the genetics of group formation and provide several genetic targets for further research on this topic.

CONTACT Yehuda Ben-Shahar benshahary@wustl.edu Department of Biology, Washington University in St. Louis, St. Louis 63130, MO, USA
Supplemental data for this article can be accessed here.

Materials and methods

Animals

All fly lines were reared on standard corn syrup-soy food (Achron Scientific), and kept under a 12 h:12h light:dark schedule at 25 °C and 60% humidity. Lines from the DGRP (Mackay *et al.,* 2012) used in this study are available from the Bloomington *Drosophila* Stock Center (BDSC, Bloomington, IN). The *elav*-GAL4 and *tubulin*-GAL4 lines were from the Bloomington *Drosophila* Stock Center. UAS-RNAi targeting specific candidate genes were from either the Vienna *Drosophila* Resource Center (VDRC) or the BDSC TRiP collection (Dietzl *et al.,* 2007; Perkins *et al.,* 2015). Because these two UAS-RNAi collections were generated by using different genetic backgrounds and expression vectors, the parental GAL4 and UAS lines were used as wild type controls for F1 VDRC crosses (Figure 4(A)), and the GAL4 > UAS-VALIUM10.GFP (BDSC # 35786) F1 controls for TRiP crosses. All fly lines used in this study, along with their stock numbers and genotypes, are listed in Supplemental Table S1.

Larval aggregation assays

Twenty-five to thirty adult flies, evenly distributed with males and females, were transferred to 25 ml vials containing standard corn syrup-soy food (Archon Scientific, Raleigh, NC), and females were allowed to lay eggs for 24 h. The adult flies were disposed of, and the vials were incubated at 25 °C for four days until healthy groups of larvae were visibly feeding. Larvae were extracted from the vials by pouring 15 ml of 15% sucrose solution (w/v) into each vial and gently agitating the food until the larvae floated to the surface and were able to be removed using a small spatula. Larvae were rinsed 2x with 1.5 ml of 15% sucrose solution to remove any remaining food from the larvae. Subsequently, 30 second/thirdinstar larvae were placed onto the center of a 60 mm petri dish containing 20% apple juice (v/v) and 1% agar (w/v) *en masse* and allowed to roam the plate freely for 15 min. A picture of the plate was then taken (Figure 1(A)), and the fraction of aggregating larvae was calculated as described below. All behavioral assays were conducted at 25 °C and 70% humidity. Although age can affect levels of larval foraging and aggregation behaviors (Durisko, Kemp, Mubasher, & Dukas, 2014; Sawin-McCormack, Sokolowski, & Campos, 1995), in our assay, mixed second and third instar larvae aggregate at similar levels to groups of only 3rd instar larvae (Figure 1 and Supplemental Figure S1).

Larval groups were defined as an 'aggregate' if two or more larvae were both (i) in physical contact with one another and (ii) burrowing into the agar plate. To calculate the fraction of larvae that were aggregating, we summed the number of larvae forming aggregates and divided it by the total number of larvae observable from the picture taken at the end of the test period. Any larvae that had roamed off the agar plate were excluded from the total number of larvae used for calculating their aggregation score.

Genome wide association study

A total of 4–9 behavioral assays were conducted for each DGRP line, and the mean proportion of aggregating larvae was used for comparison in a genome wide association study (GWAS). A linear regression model was run using the easyGWAS analysis package specifically designed for analyzing the DGRP GWAS data (Grimm *et al.,* 2017), with default statistical parameters, to search for genotype by phenotype associations. A total of 2,370,987 single nucleotide polymorphisms (SNPs) from each of 48 DGRP lines were included in the GWAS, after filtering out any SNPs that were of the same genotype across all lines. Linkage disequilibrium and minor allele frequencies (MAF) were calculated using the PLINK package (Purcell *et al.,* 2007).

Gene networks

GeneMANIA was used to predict a functional gene interaction network for all genes identified in the initial GWAS containing SNPs with a p values of less than $10-4.5$ (Warde-Farley *et al.,* 2010). A gene was said to contain a SNP if the SNP occurred within ±500 base pairs of its coding exons as annotated in the *Drosophila* reference genome (version 5.57, FB2014 03). Subsequently, co-expression, co-localization, shared protein domains, and protein-protein interactions were used to calculate the gene interaction network, and up to 20 genes that were not identified as significant in the GWAS were allowed to be added to the network. Genes added to the network were selected such they maximized the number of connections between genes already present in the network (Warde-Farley *et al.,* 2010).

Gene ontology analysis

Genes containing SNPs with a p values of less than $10-4.5$were screened for functionally-enriched gene ontologies using the bioprofiling.de servers ProfCom framework (Antonov et al., 2008). All genes included in the functional gene interaction network were also screened for functionally enriched gene ontologies using GeneMANIA (Warde-Farley, et al., 2010). The gene interaction network included 20 additional genes that did not contain significant SNPs; the gene ontology (GO) terms (The Gene Ontology, 2019) found to be associated with this network are therefore more general to a set of genes commonly found to interact with one another, rather than those specifically identified in the GWAS.

Real time quantitative reverse transcription polymerase chain reaction (RT-qPCR)

mRNA was collected from groups of 10 whole larvae ($n = 3$–4 replicates per line) using Trizol (Thermo Fisher, Waltham, MA) and reverse transcribed to cDNA using SuperScript III Reverse Transcriptase (Thermo Fisher). Sybr Green (Thermo Fisher) was used to amplify and quantify expression levels for all genes containing significant SNPs

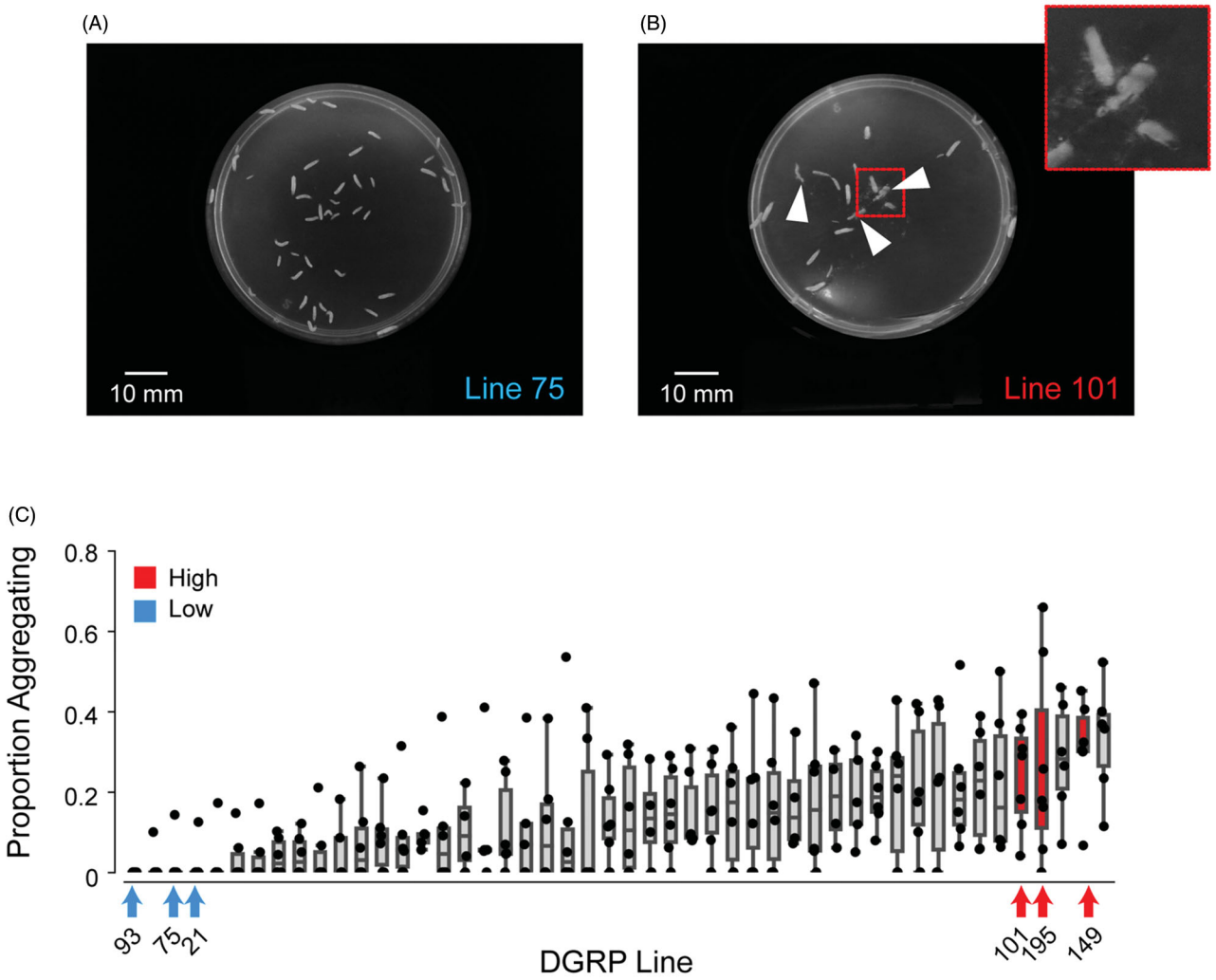


Figure 1. Variation in levels of aggregation between natural populations of *Drosophila*. (A) An image of a DGRP line (Line 75) that showed low levels of aggregation and (B) an image of a DGRP line (Line 101) that showed high levels of aggregation. White arrowheads point to groups of aggregating larvae. (C) Boxplots represent the median (horizontal line), with the box representing the 25 and 75th percentiles and the whiskers the 5 and 95th percentiles of the proportion of aggregating larvae for each of the 48 DGRP lines that were included in the GWAS ($n = 5$–9 replicates per line). Circles represent individual data points. DGRP lines with either low ('Low') or high ('High') levels of aggregation that were used in subsequent analyses are labeled by arrows and their DGRP line number.

identified in the GWAS. Expression values were calculated relative to the *rp49* control gene using the delta delta Ct method, as we have previously described (Ben-Shahar, Lu, Collier, Snyder, Schnizler, Welsh, 2010; Ben-Shahar,Nannapaneni, Casavant, Scheetz, & Welsh, 2007; Hill *et al.*, 2017; Hill & Ben-Shahar, 2018; Hill, Jain, Folan, & Ben-Shahar, 2019; Lu, LaMora, Sun, Welsh, & Ben-Shahar, 2012; Vernier *et al.*, 2020; Vernier, Krupp, Marcus, Hefetz, Levine, & Ben-Shahar, 2019). All RT-qPCR primers used in this study are listed in Supplemental Table S2.

CG14205-GAL4 transgenic flies

An approximately 3 kbp (X:19590171–19593107) region of the *CG14205* promoter was synthesized by Integrated DNA Technologies, Inc (IDT, Coralville, IA) and placed into the pUCIDT-ampR plasmid (IDT). We subcloned this region into the pENTR-1A plasmid (Thermo Fisher) using KpnI and XhoI restriction sites on either side of the promoter, and then used Gateway cloning (Thermo Fisher) to move the promoter into the pBPGAL4.2::p65 plasmid (Addgene #26229) (Pfeiffer *et al.*, 2010). This plasmid was subsequently injected into BDSC line #24483 (RainbowGene Inc.), and positive offspring were identified and back-crossed into w^{1118} following our standard protocols (Ben-Shahar *et al.*, 2010; Hill *et al.*, 2017; Hill, Jain, Folan, & Ben-Shahar, 2019; Lu *et al.*, 2012; Zheng, Valakh, DiAntonio, & Ben-Shahar 2014). The *CG14205*-GAL4 line was crossed with *UAS-mCD8::GFP* (BDSC #32188) and imaged in third-instar larvae (Ben-Shahar *et al.*, 2010; Ben-Shahar *et al.*, 2007; Hill *et al.*, 2017; Lu *et al.*, 2012; Sovik *et al.*, 2017; Sun *et al.*, 2009).

Results

Genetic variation underlying group formation

As *D. melanogaster* larvae develop, they exhibit a gradual increase in aggregation behavior (Wu *et al.*, 2003). However, the overall genetic architecture that drives the quantitative aspects of larval aggregation remains largely unknown.

Therefore, to better understand the genetics underlying aggregation, we screened 48 randomly chosen isogenic wild type lines from the DGRP (Mackay *et al.,* 2012) for levels of aggregation in third instar larvae and subsequently performed a GWAS to look for genetic variation associated with this phenotype.

We found that different lines varied significantly in the extent of aggregation, with some lines tending not to form any aggregates (termed 'Low' lines) and other lines containing as many as 40–60% of aggregating larvae (termed 'High' lines) (Figure 1). We then ran analysis of variances (ANOVAs) to search for genetic variation, in the form of SNPs, associated with the mean proportion of aggregating larvae across lines (Shorter *et al.,* 2015; Swarup, Huang, Mackay, & Anholt, 2013). A total of 2,370,987 ANOVAs were run for each unfiltered SNP in the 48 DGRP lines analyzed, which uncovered 58 significant SNPs ($p < 10-5$). Subsequently, SNPs that fell within 500 bp of the exons of protein coding genes identified 17 candidate genes that might be playing a role in larval aggregation decisions (Figure 2(A–C), Supplemental Table S3). As have been observed in previous studies that used the DGRP collection, many of the variants associated with our candidate genes showed high linkage (Figure 2(C); Morgante, Huang, Maltecca, & Mackay, 2018; Slattery *et al.,* 2014).

The neurogenetic network of larval aggregation behavior

To investigate whether specific genetic pathways might be playing a role in larval aggregation decisions, we next used GO analyses. Because our initial conservative $p < 10-5$significance threshold yielded only 17 protein-coding genes that might be causally associated with levels of aggregation, we used the less conservative threshold of $p < 10-4.5$, which increased the number of candidate genes to 68. This analysis indicated that this gene list is enriched for the GO terms 'axon guidance' (GO:0007411, $p = 0.01$) and 'plasma membrane' (GO:0005886, $p = 0.01$). To further expand the analyzed gene network, we next extended the empirically defined gene network by using the following edges: co-expression, co-localization, shared protein domains, and protein-protein interactions (Supplemental Figure S2(A)). GO analysis of the extended gene list was still enriched for 'axon guidance'; however, four out of the top six enriched GO terms are neural-tissue specific (Supplemental Figure S2(B)). Together, these data suggest that at least some of the genetic variations we have identified impact population level phenotypic variations in aggregation decisions via neuronal functions.

Genetic variations associated with mRNA expression levels

SNPs falling within promoter and enhancer regions of a protein coding gene often affect mRNA expression levels (Khurana *et al.,* 2016; Nord & West, 2020; Visel, Rubin, & Pennacchio 2009). Since most of the SNPs we have identified in our GWAS are either intronic or fall upstream of their associated genes (37/46; Supplemental Table S3), we next tested the hypothesis that some of the identified SNPs affect gene action via their effects on mRNA expression levels. To test this hypothesis, we compared the mRNA expression levels of each of the 17 candidate genes identified in our initial conservative screen between the three phenotypically highest ('High') and three lowest ('Low' aggregating DGRP lines (Figure 3(A,B)) by using real-time RT-qPCR analyses. We found that at least one SNP (X:19488026) was significantly associated with higher mRNA expression levels of its parent gene, *CG14205*, in all 'Low' lines relative to all 'High' lines (one-way analysis of variance (ANOVA); $F(1,4) = 13.43$, $p = 0.02$; Figure 3). Although we do not know yet how this specific SNP might affect *CG14205* mRNA expression levels, its location immediately downstream of a predicted splice donor site in intron 5 of *CG14205* (Figure 3(C)) suggests that the mRNAs transcribed by the 'Low' and 'High' alleles exhibit different splicing processivity and/or stability.

Although the biological functions of *CG14205* are unknown, the protein is predicted to be a membrane bound Acyltransferase 3 (IPR002656) that is related to the nose resistant-to-fluoxetine (NRF) protein family in *C. elegans* (Choy & Thomas, 1999). Since several family members have been found to be expressed in the gut epithelium of worms, it has been hypothesized that they may function as novel transporters of lipophilic molecules (Choy, Kemner, & Thomas 2006). However, the specific biochemical function of this protein family remains uncharacterized. Nevertheless, previous studies in the moth *Bombyx mori*, have shown that various acyltransferases are required for the synthesis of sex pheromones in moths and other insects (Ding *et al.*, 2016; Du *et al.,* 2015; Du, Zhang, Zhu, Yin, & An, 2012). Further, a quantitative trait locus (QTL) associated with intra- and interspecific variations in sex pheromones in noctuid moths has been mapped to the regulation of a gene containing a putative Acyltransferase 3 domain (Groot *et al.,* 2013). Therefore, it is possible that *CG14205* plays a direct role in the synthesis of larval aggregation pheromones in *D. melanogaster*.

Candidate gene knockdown leads to altered levels of aggregation

To further establish a causal role for the genes identified in our initial screen, we studied the effects of neuronal-specific RNAi knockdown of each gene by using the pan-neuronal *elav-GAL4* driver. However, neuronal knockdown of five of the 17 genes we examined (*Vha36-1*, *dsx-c73a*, *pros*, *cindr*, and *CG45002*) was lethal. Of the remaining 12 genes, neuronal knockdown of three (*CG8187*, *CG14502*, and *rn*) led to higher levels of aggregation relative to controls (Figure 4(A–C)). These results suggest that the activity of these four genes affects aggregation decisions in feeding larvae.

In contrast, knockdown of *CG14205* in neural tissues did not significantly alter aggregation levels, which suggested that the strong association between the specific *CG14205*

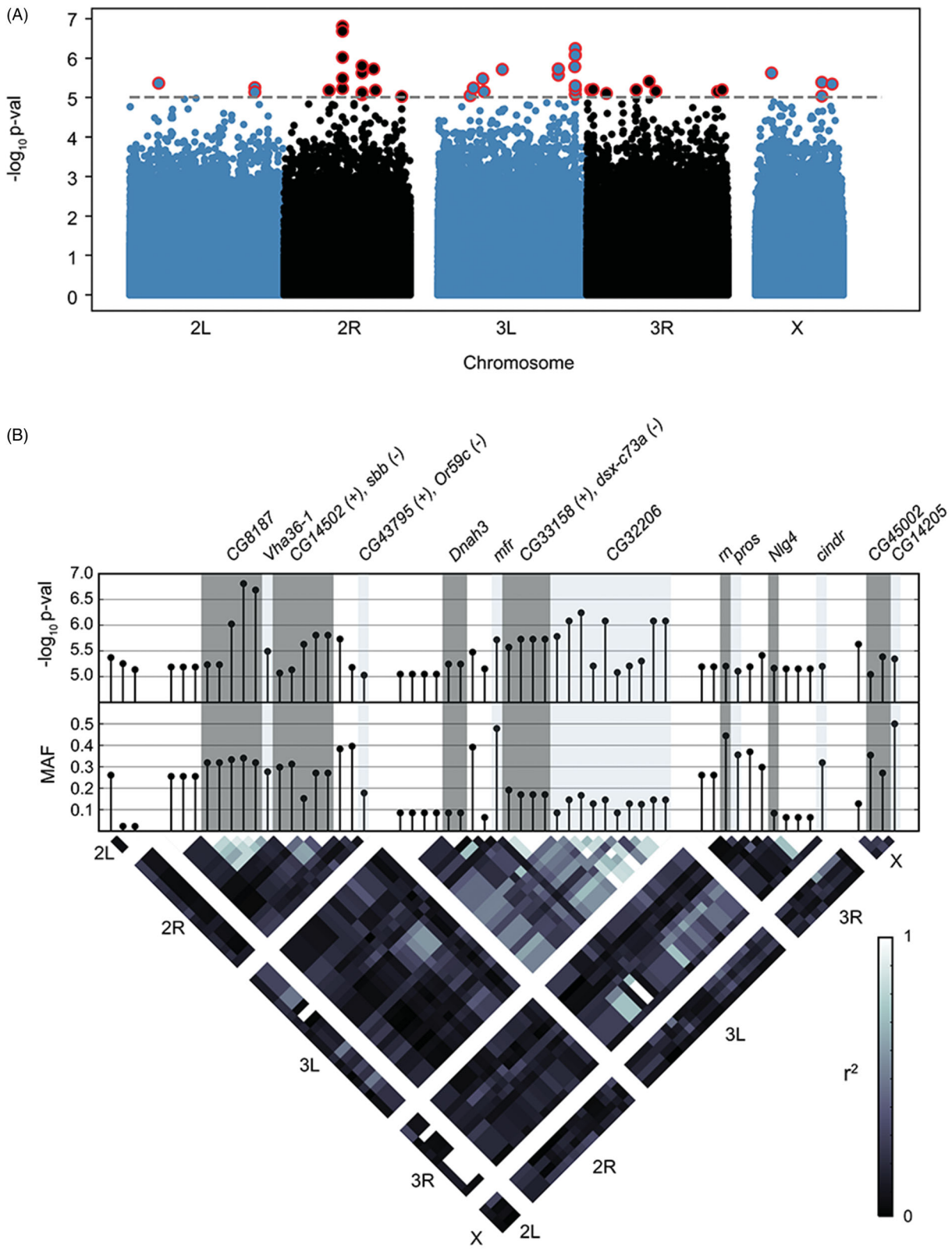


Figure 2. A genome-wide association study identified 58 SNPs that were associated with the extent of larval aggregation across DGRP lines. (A) Manhattan plot showing transformed *p* values for each of the SNPs included in the GWAS. SNPs with a *p* values less than 10^{-5} (shown by the dashed gray line) were retained for further analysis and are outlined in red. (B) Higher resolution view of SNPs highlighted in (A). (Top) Transformed *p* values and (Middle) minor allele frequencies (MAFs) for each of the retained SNPs. SNPs that fell within ±500 base pairs of the coding region of a gene are labeled and highlighted together. SNPs that fall within the coding region of genes with overlapping ORFs on the plus and minus DNA strands are labelled accordingly. (C) Linkage disequilibrium matrix between all identified SNPs. Note that some SNPs appear to be tightly linked because of, small MAF values due to the rarity of these SNPs in the DGRP collection.

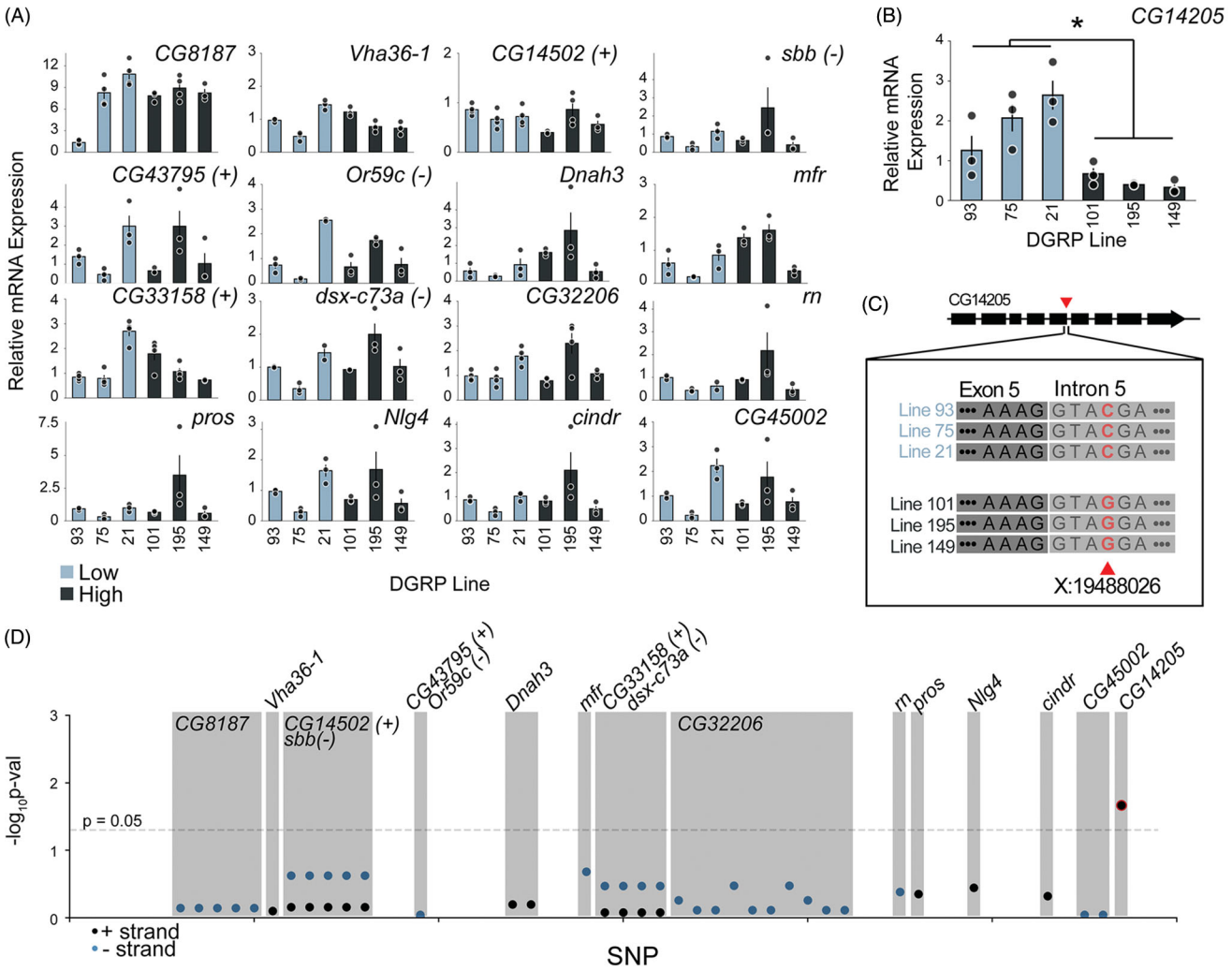


Figure 3. mRNA expression analysis of SNP-containing genes in lines with either low or high levels of aggregation. (A) Relative mRNA expression levels for each of the SNP-containing genes identified in the GWAS ($n = 3$–4 replicates per line). 'Low' aggregating lines are shown in light shade, and 'High' aggregating lines are shown in dark shade. (B) Relative mRNA expression levels for the *CG14205* gene. A significant association between SNP genotype and *CG14205* mRNA expression was identified ($p < 0.05$; one-way analysis of variance), whereby 'Low' aggregating lines had higher levels of expression than 'High' aggregating lines. Note that 'Low and High' lines segregated by genotype, as shown in (C). (C) Transformed p values for associations (ANOVAs) between specific SNP haplotypes and relative mRNA expression level of the gene associated with that SNP. SNPs falling within the same gene are labeled and highlighted together, and SNPs which were significantly associated ($p < 0.05$) with mRNA expression of its gene are outlined in red. (D) Genetic architecture of the *CG14205* gene and the DNA sequences surrounding the significantly associated SNP for each of the Low and High DGRP lines. Note that the SNP, X:19488026 (denoted by a arrow-head), falls just past the exon-intron boundary within intron 5 and is positioned to potentially effect mRNA splicing.

alleles, mRNA expression, and aggregation levels, is mediated via its action in non-neuronal tissues. However, because the specific non-neuronal cells that might mediate the effect of *CG14205* knockdown on larval aggregation are unknown, we used the ubiquitous *tubulin*-GAL4. As *CG14205* mRNA is expressed to a greater extent in low-aggregating lines, we hypothesized that knocking down *CG14205* should lead to increased levels of aggregation. Indeed, global *CG14205* knockdown resulted in an increase in the fraction of larvae aggregating (one-tailed, Student's *t*-test, $p = 0.025$; $N = 11$–12 per group; Figure 4(D)). These results suggest that *CG14205* functions to suppress aggregation in *D. melanogaster* larvae via neuronal-independent pathways in the larval midgut.

While we do not know yet if or how the midgut activity levels of *CG14205* might affect the decision of individual larvae to join a group, this decision is likely controlled by both external sensory stimuli and internal receptors which detect those stimuli. One possible interpretation of these data is that the *CG14205* gene is responsible for the biosynthesis or release of a sensory stimulus that inhibits larvae from interacting with one another and forming groups. This hypothesis is consistent with the fact that *CG14205* is required in non-neuronal cells for maintaining normal levels of larval aggregation (compared to controls). Further, mining the FlyExpress and Flygut databases revealed that the expression of *CG14205* is enriched in enterocytes in the R2, R4, and R5 regions of the larval midgut (Buchon *et al.*, 2013; Celniker *et al.*, 2009; Figure 5(A–C)). This expression pattern was further confirmed by generating a transgenic reporter of the putative *CG14205* promoter, which revealed strong expression in two distinct regions of the midgut (Figure 5(D–F)). Together, these results suggest that *CG14205* plays a role in

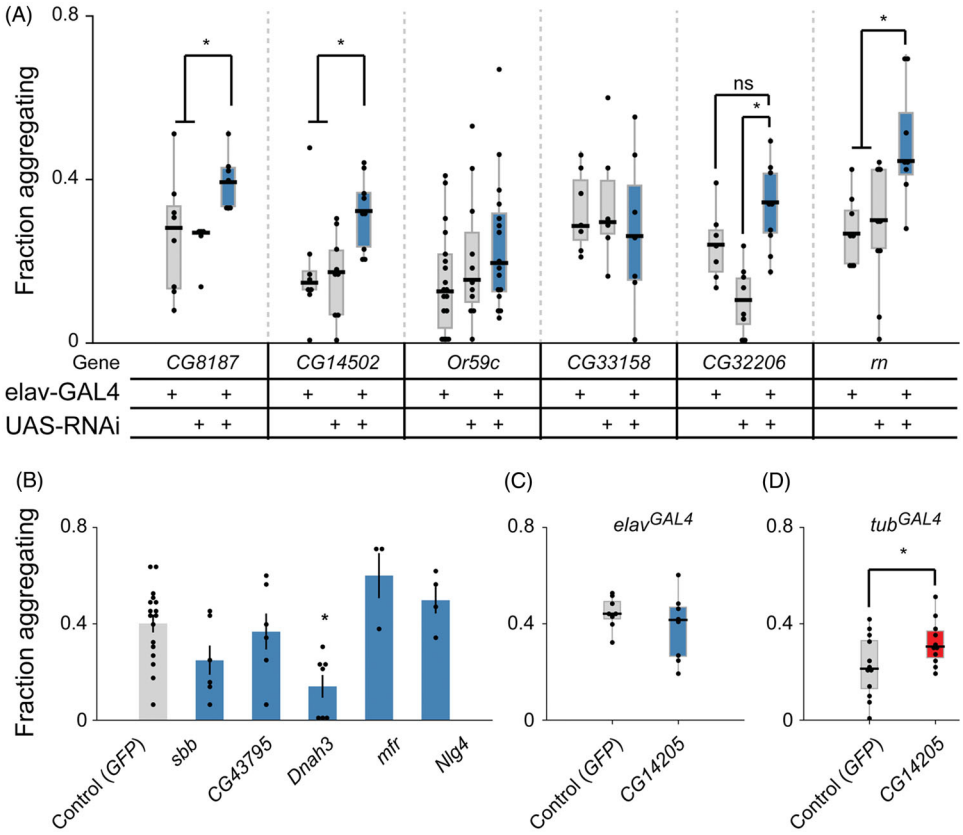


Figure 4. Neuronal knockdown of some candidate genes leads to altered aggregation behavior. (A) Pan-neuronal RNAi-mediated knockdown of SNP-associated genes (UAS-RNAi lines from the Vienna *Drosophila* resource center). Knockdown of *CG8187* ($n = 5–8$, $p < 0.05$), *CG14502* ($n = 8–9$, $p < 0.05$), and *rn* ($n = 8–9$, $p < 0.01$) using *elav*-GAL4 lead to increased levels of aggregation when compared to parental controls ($n = 6–19$, for all other lines). All statistical comparisons used one-way ANOVA followed by a Tukey's HSD *post-hoc* test. (B) Pan-neuronal RNAi-mediated knockdown of SNP-associated genes (UAS-RNAi lines from the Bloomington TRiP collection). Knockdown of *Dnah3* using *elav*-GAL4 lead to a decrease in fraction of larvae aggregating ($n = 7–17$, $p < 0.01$), whereas no other gene knockdowns were significantly different from control ($n = 4–17$). Pairwise Student's t-tests were run between each gene knockdown and control to look for statistical significance, and *p* values were adjusted for multiple comparisons using a Bonferroni correction. (C) TRiP-RNAi-mediated knockdown of *CG14205* in neural tissues, using the *elav*-GAL4 driver, did not lead to altered aggregation ($n = 8$ per group, $p > 0.05$; one-tailed Student's t-test). (D) TRiP-RNAi-mediated knockdown of *CG14205* in all tissues, using the *tubulin*-GAL4 driver, led to a significant increase in the fraction of larvae aggregating compared to control ($n = 11–12$, $p = 0.025$; one-tailed Student's t-test).

the synthesis or release, rather than detection, of an inhibitory molecule regulating aggregation.

Discussion

It is often assumed that group and social behaviors arise via complex interactions between many genes. Here, we have used an unbiased behavioral quantitative genetic screen to identify population-level natural genetic variations that underlie aggregation in *D. melanogaster* larvae. As expected, our analysis revealed that the decision of individuals on whether to aggregate with other conspecifics is likely dependent on a complex genetic network that acts in both neuronal and non-neuronal tissues. Furthermore, by using *in vivo* genetic manipulations, we show that at the population level, both qualitative and quantitative variations could be causally associated with the overall observed behavioral variations between individuals. However, whether the specific identified genes exert their impact on aggregation via a common pathway, and the exact cellular and physiological processes affected by these genes, remain unknown.

Specifically, we found that mRNA expression level of *CG14205*, which encodes a putative acetyl transferase, is higher in DGRP lines that exhibit low levels of aggregation relative to its expression in lines the exhibit high levels of aggregation (Figure 3(B)). While the mechanism regulating this variation in transcript levels is not known, the SNP identified in our initial GWAS screen is adjacent to a predicted intronic splice donor site (Figure 3(C)), which can affect gene function via differential mRNA splicing and/or stability via posttranscriptional processes (Witte, 2010; Xiao, Chang, & Li, 2017). Regardless of the molecular mechanism, the genetic and behavioral data we present here suggest that higher activity of *CG14205*, possibly in the midgut, inhibits larval aggregation by an unknown physiological process. Although our RNAi knockdown studies indicate that *CG14205* is not specifically required in neurons, it remains a possibility that it influences larval behavior via its action in glia or the endocrine system. Alternatively, this gene could be required for the production of a chemical signal that modulates larval aggregation decisions via the enzymatic modification of gut metabolites (Blomquist *et al.*, 2010; Chiu, Keeling, & Bohlmann, 2019; Hunt & Borden, 1990).

Recent studies have identified both specific chemical cues—pheromones—and receptors to be required for directing aggregation behaviors in *D. melanogaster* larvae (Mast

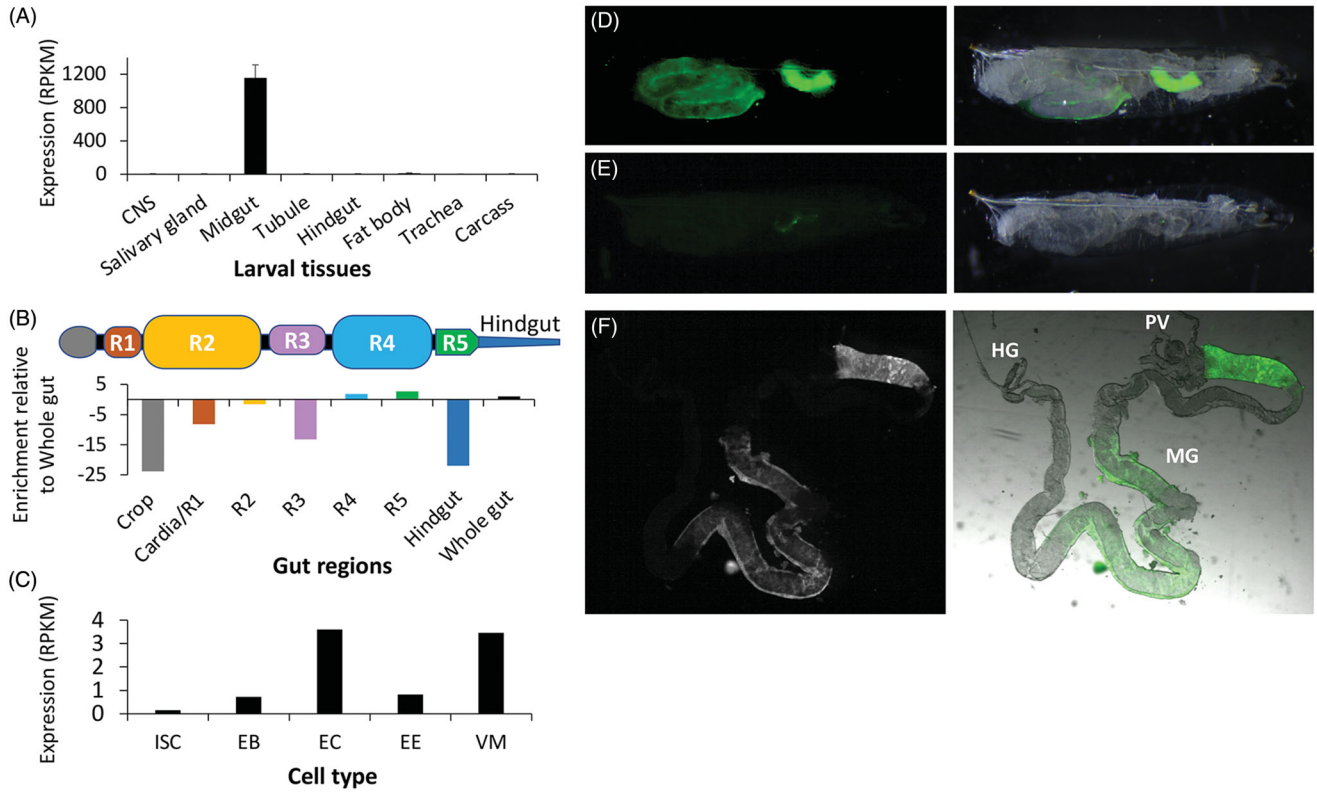


Figure 5. CG14205 expression in larvae. (A) Expression levels of *CG14205* across larval tissues. Data were extracted from the FlyAtlas database. (B) Enrichment of CG14205 mRNA, relative to the whole gut. (C) Expression levels of *CG14205* across midgut cell types. ISC: intestinal stem cells; EB: enteroblasts; EC: enterocytes; EE: enteroendocrine cells; VM: visceral muscle. (D) *CG14205* expression is restricted to the larval midgut. Image of an intact larva expressing GFP under the control of the *CG14205* GAL4. (E) Image of a control *CG14205-GAL4* larva. (F) Image of third instar dissected gut: (Left) GFP signal, (Right) Overlay of GFP signal with a DIC image.

et al., 2014). Although most of what is known about pheromone synthesis in *Drosophila* and other insects relates to cuticular hydrocarbons production by fat-body cells and the oenocytes (Makki, Cinnamon, & Gould, 2014; Wicker-Thomas *et al.,* 2015; Zelle *et al.,* 2019), our data indicate that gut-derived metabolites can also possibly act as pheromones in *Drosophila*. The possible contribution of *CG14205* to pheromone synthesis is further supported by previous findings about the contribution of acyltransferases to pheromonal signaling in other insect species (Ding *et al.,* 2016; Zhang *et al.,* 2017). Therefore, it is possible that this enzyme functions in the production of some inhibitory chemical cues that *Drosophila* larvae are responsive to during feeding.

Previous studies by us and others have shown that pheromone-driven social interactions in *Drosophila* and other insects often require the balancing action of both attractive and repulsive cues (Allison & Cardé, 2016; Ben-Shahar *et al.,* 2010; Blomquist & Vogt, 2003; Lu *et al.,* 2012; Lu, Zelle, Seltzer, Hefetz, & Ben-Shahar, 2014; McKinney, Vernier, & Ben-Shahar 2015; Zelle *et al.,* 2019). However, in our study, the knockdown of majority of identified candidate genes leads to increased levels of larval aggregation, which suggest that the primary contributions of these genes are to the suppression of aggregation. We do not know yet whether the behavioral variations in larval aggregation across the different DGRP haplotypes that comprise the DGRP collection used here represents is directly related to natural population-level adaptive variations. Nevertheless, previous empirical and theoretical studies have indicated that the Allee effect drives *Drosophila* larval aggregation, which can vary in magnitude across species, genetic backgrounds, and niches (Douglas, Dawson-Scully, & Sokolowski, 2005; Durisko *et al.,* 2014; Fitzpatrick *et al.,* 2007; Louis & de Polavieja, 2017; Sokolowski, 2010; Takahashi, 2006; Wertheim, Marchais, Vet, & Dicke 2002). Another non-mutually exclusive explanation might be that our lab assay conditions, and the specific behavioral paradigm used, biased our screen towards the identification of genes whose role contributes specifically to the suppression of larval aggregation.

With the exception of the gene *scribbler* (*sbb*), which has been previously shown to affect larval foraging (Suster, Karunanithi, Atwood, & Sokolowski, 2004; Yang, Shaver, Hilliker, & Sokolowski, 2000), our study has uncovered several novel genes involved in directing social aggregation while feeding in *Drosophila* larvae. Although we do not know yet the specific molecular and cellular mechanisms by which any of these genes affect larval feeding behaviors, our data further indicate that natural genetic polymorphisms affect larval social feeding behaviors via both neuronal and non-neuronal pathways (Allen *et al.,* 2017; Anreiter *et al.,* 2017; Sokolowski, 2010).

Disclosure statement

The authors declare there are no conflicts of interest.

Funding

This work was supported by grants from the National Science Foundation to YB-S [NSF-IOS 1322783, NSF-IOS 1754264, and NSF-DBI 1707221].

ORCID

Yehuda Ben-Shahar http://orcid.org/0000-0002-2956-2926

References

Allen, A.M., Anreiter, I., Neville, M.C., & Sokolowski, M.B. (2017). Feeding-related traits are affected by dosage of the foraging gene in *Drosophila melanogaster*. *Genetics*, *205*(2), 761–773. doi:10.1534/genetics.116.197939

Allison, J.D., & Cardé, R.T. (2016). *Pheromone communication in moths: Evolution, behavior, and application [still image]*. Oakland, California: University of California Press.

Anreiter, I., Kramer, J., & Sokolowski, M.B. (2017). Epigenetic mechanisms modulate differences in Drosophila foraging behavior. *Proceedings of the National Academy of Sciences of the USA*, *114*(47), 12518–12523. doi:10.1073/pnas.1710770114

Antonov, A.V., Schmidt, T., Wang, Y., & Mewes, H.W. (2008). ProfCom: A web tool for profiling the complex functionality of gene groups identified from high-throughput data. *Nucleic Acids Research*, *36*, W347–351. doi:10.1093/nar/gkn239

Ben-Shahar, Y., Lu, B., Collier, D.M., Snyder, P.M., Schnizler, M., & Welsh, M.J. (2010). The *Drosophila* gene CheB42a is a novel modifier of Deg/ENaC channel function. *PLoS One*, *5*(2), e9395. doi:10.1371/journal.pone.0009395

Ben-Shahar, Y., Nannapaneni, K., Casavant, T.L., Scheetz, T.E., & Welsh, M.J. (2007). Eukaryotic operon-like transcription of functionally related genes in *Drosophila*. *Proceedings of the National Academy of Sciences of the USA*, *104*(1), 222–227. doi:10.1073/pnas.0609683104

Blomquist, G.J., Figueroa-Teran, R., Aw, M., Song, M., Gorzalski, A., Abbott, N.L., … Tittiger, C. (2010). Pheromone production in bark beetles. *Insect Biochemistry and Molecular Biology*, *40*(10), 699–712. doi:10.1016/j.ibmb.2010.07.013

Blomquist, G.J., & Vogt, R. (2003). *Insect pheromone biochemistry and molecular biology the biosynthesis and detection of pheromones and plant volatiles*. Amsterdam, Boston: Elsevier/Academic Press.

Buchon, N., Osman, D., David, F.P., Fang, H.Y., Boquete, J.P., Deplancke, B., & Lemaitre, B. (2013). Morphological and molecular characterization of adult midgut compartmentalization in *Drosophila*. *Cell Reports*, *3*(5), 1725–1738. doi:10.1016/j.celrep.2013.04.001

Celniker, S.E., Dillon, L.A., Gerstein, M.B., Gunsalus, K.C., Henikoff, S., Karpen, G.H., … Waterston, R.H(2009). Unlocking the secrets of the genome. *Nature*, *459*(7249), 927–930. doi:10.1038/459927a

Chiu, C.C., Keeling, C.I., & Bohlmann, J. (2019). The cytochrome P450 CYP6DE1 catalyzes the conversion of α-pinene into the mountain pine beetle aggregation pheromone trans-verbenol. *Scientific Reports*, *9*(1), 1477. doi:10.1038/s41598-018-38047-8

Choy, R.K., Kemner, J.M., & Thomas, J.H. (2006). Fluoxetine-resistance genes in caenorhabditis elegans function in the intestine and may act in drug transport. *Genetics*, *172*(2), 885–892. doi:10.1534/genetics.103.024869

Choy, R.K., & Thomas, J.H. (1999). Fluoxetine-resistant mutants in C. elegans define a novel family of transmembrane proteins. *Molecular Cell*, *4*(2), 143–152. doi:10.1016/S1097-2765(00)80362-7

Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., … Dickson, B.J. (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature*, *448*(7150), 151–156. doi:10.1038/nature05954

Ding, B.J., Lager, I., Bansal, S., Durrett, T.P., Stymne, S., & Lofstedt, C. (2016). The yeast ATF1 acetyltransferase efficiently acetylates insect pheromone alcohols: Implications for the biological production of moth pheromones. *Lipids*, *51*(4), 469–475. doi:10.1007/s11745-016-4122-4

Douglas, S.J., Dawson-Scully, K., & Sokolowski, M.B. (2005). The neurogenetics and evolution of food-related behaviour. *Trends in Neurosciences*, *28*(12), 644–652. doi:10.1016/j.tins.2005.09.006

Du, M., Liu, X., Liu, X., Yin, X., Han, S., Song, Q., & An, S. (2015). Glycerol-3-phosphate O-acyltransferase is required for PBAN-induced sex pheromone biosynthesis in *Bombyx mori*. *Scientific Reports*, *5*(1), 8110. doi:10.1038/srep08110

Du, M., Zhang, S., Zhu, B., Yin, X., & An, S. (2012). Identification of a diacylglycerol acyltransferase 2 gene involved in pheromone biosynthesis activating neuropeptide stimulated pheromone production in Bombyx mori. *Journal of Insect Physiology*, *58*(5), 699–703. doi:10.1016/j.jinsphys.2012.02.002

Durisko, Z., & Dukas, R. (2013). Attraction to and learning from social cues in fruitfly larvae. *Proceedings. Biological Sciences*, *280*(1767), 20131398. doi:10.1098/rspb.2013.1398

Durisko, Z., Kemp, R., Mubasher, R., & Dukas, R. (2014). Dynamics of social behavior in fruit fly larvae. *PLoS One*, *9*(4), e95495. doi:10.1371/journal.pone.0095495

Fitzpatrick, M.J., Feder, E., Rowe, L., & Sokolowski, M.B. (2007). Maintaining a behaviour polymorphism by frequency-dependent selection on a single gene. *Nature*, *447*(7141), 210–212. doi:10.1038/nature05764

Grimm, D.G., Roqueiro, D., Salome, P.A., Kleeberger, S., Greshake, B., Zhu, W., … Borgwardt, K.M. (2017). easyGWAS: A cloud-based platform for comparing the results of genome-wide association studies. *Plant Cell*, *29*(1), 5–19. doi:10.1105/tpc.16.00551

Groot, A.T., Staudacher, H., Barthel, A., Inglis, O., Schofl, G., Santangelo, R.G., … Gould, F. (2013). One quantitative trait locus for intra- and interspecific variation in a sex pheromone. *Molecular Ecology*, *22*(4), 1065–1080. doi:10.1111/mec.12171

Hill, A., Zheng, X., Li, X., McKinney, R., Dickman, D., & Ben-Shahar, Y. (2017). The *Drosophila* postsynaptic DEG/ENaC channel ppk29 contributes to excitatory neurotransmission. *Journal of Neuroscience*, *37*(12), 3171–3180. doi:10.1523/JNEUROSCI.3850-16.2017

Hill, A.S., & Ben-Shahar, Y. (2018). The synaptic action of degenerin/epithelial sodium channels. *Channels*, *12*(1), 262–275. doi:10.1080/19336950.2018.1495006

Hill, A.S., Jain, P., Folan, N.E., & Ben-Shahar, Y. (2019). The *Drosophila* ERG channel seizure plays a role in the neuronal homeostatic stress response. *PLoS Genetics*, *15*(8), e1008288. doi:10.1371/journal.pgen.1008288

Hunt, D.W., & Borden, J.H. (1990). Conversion of verbenols to verbenone by yeasts isolated fromDendroctonus ponderosae (Coleoptera: Scolytidae). *Journal of Chemical Ecology*, *16*(4), 1385–1397. doi:10.1007/BF01021034

Kaun, K.R., Hendel, T., Gerber, B., & Sokolowski, M.B. (2007). Natural variation in *Drosophila* larval reward learning and memory due to a cGMP-dependent protein kinase. *Learning & Memory*,*14*(5), 342–349. doi:10.1101/lm.505807

Kaun, K.R., Riedl, C.A., Chakaborty-Chatterjee, M., Belay, A.T., Douglas, S.J., Gibbs, A.G., & Sokolowski, M.B. (2007). Natural variation in food acquisition mediated via a *Drosophila* cGMP-dependent protein kinase. *Journal of Experimental Biology*, *210*, 3547–3558. doi:10.1242/jeb.006924

Khurana, E., Fu, Y., Chakravarty, D., Demichelis, F., Rubin, M.A., & Gerstein, M. (2016). Role of non-coding sequence variants in cancer. *Nature Reviews Genetics*, *17*(2), 93–108. doi:10.1038/nrg.2015.17

Leonhardt, S.D., Menzel, F., Nehring, V., & Schmitt, T. (2016). Ecology and evolution of communication in social insects. *Cell*, *164*(6), 1277–1287. doi:10.1016/j.cell.2016.01.035

Lihoreau, M., Clarke, I.M., Buhl, J., Sumpter, D.J., & Simpson, S.J. (2016). Collective selection of food patches in *Drosophila*. *Journal of Experimental Biology*, *219*, 668–675. doi:10.1242/jeb.127431

Louis, M., & de Polavieja, G. (2017). Collective behavior: Social digging in *Drosophila* larvae. *Current Biology*, *27*(18), R1010–R1012. doi:10.1016/j.cub.2017.08.023

Lu, B., LaMora, A., Sun, Y., Welsh, M.J., & Ben-Shahar, Y. (2012). ppk23-Dependent chemosensory functions contribute to courtship behavior in *Drosophila* melanogaster. *PLoS Genetics*, *8*(3), e1002587. doi:10.1371/journal.pgen.1002587

Lu, B., Zelle, K.M., Seltzer, R., Hefetz, A., & Ben-Shahar, Y. (2014). Feminization of pheromone-sensing neurons affects mating decisions in *Drosophila* males. *Biology Open*, *3*(2), 152–160. doi:10.1242/bio.20147369

Mackay, T.F.C., Richards, S., Stone, E.A., Barbadilla, A., Ayroles, J.F., Zhu, D., … Cridland, J.M. (2012). The *Drosophila* melanogaster genetic reference panel. *Nature*, *482*(7384), 173–178. doi:10.1038/nature10811

Makki, R., Cinnamon, E., & Gould, A.P. (2014). The development and functions of oenocytes. *Annual Review of Entomology*, *59*, 405–425. doi:10.1146/annurev-ento-011613-162056

Mast, J.D., De Moraes, C.M., Alborn, H.T., Lavis, L.D., & Stern, D.L. (2014). Evolved differences in larval social behavior mediated by novel pheromones. *eLife*, *3*, e04205. doi:10.7554/eLife.04205

McKinney, R.M., Vernier, C., & Ben-Shahar, Y. (2015). The neural basis for insect pheromonal communication. *Current Opinion in Insect Science*, *12*, 86–92. doi:10.1016/j.cois.2015.09.010

Morgante, F., Huang, W., Maltecca, C., & Mackay, T.F.C. (2018). Effect of genetic architecture on the prediction accuracy of quantitative traits in samples of unrelated individuals. *Heredity*, *120*(6), 500–514. doi:10.1038/s41437-017-0043-0

Nord, A.S., & West, A.E. (2020). Neurobiological functions of transcriptional enhancers. *Nature Neuroscience*, *23*(1), 5–14. doi:10.1038/s41593-019-0538-5

Perkins, L.A., Holderbaum, L., Tao, R., Hu, Y., Sopko, R., McCall, K., … Perrimon, N. (2015). The Transgenic RNAi project at Harvard medical school: Resources and validation. *Genetics*, *201*(3), 843–852. doi:10.1534/genetics.115.180208

Pfeiffer, B.D., Ngo, T.T., Hibbard, K.L., Murphy, C., Jenett, A., Truman, J.W., & Rubin, G.M. (2010). Refinement of tools for targeted gene expression in *Drosophila*. *Genetics*, *186*(2), 735–755. doi:10.1534/genetics.110.119917

Prokopy, R.J., & Roitberg, B.D. (2001). Joining and avoidance behavior in nonsocial insects. *Annual Review of Entomology*, *46*, 631–665. . 46] doi:10.1146/annurev.ento.46.1.631

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., … Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. doi:10.1086/519795

Rooke, R., Rasool, A., Schneider, J., & Levine, J.D. (2020). *Drosophila* melanogaster behaviour changes in different social environments based on group size and density. *Communications Biology*, *3*(1), 304. doi:10.1038/s42003-020-1024-z

Sawin-McCormack, E.P., Sokolowski, M.B., & Campos, A.R. (1995). Characterization and genetic analysis of *Drosophila* melanogaster photobehavior during larval development. *Journal of Neurogenetics*, *10*(2), 119–135. doi:10.3109/01677069509083459

Shorter, J., Couch, C., Huang, W., Carbone, M.A., Peiffer, J., Anholt, R.R., & Mackay, T.F. (2015). Genetic architecture of natural variation in *Drosophila* melanogaster aggressive behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(27), E3555–3563. doi:10.1073/pnas.1510104112

Slattery, M., Ma, L., Spokony, R.F., Arthur, R.K., Kheradpour, P., Kundaje, A., … White, K.P. (2014). Diverse patterns of genomic targeting by transcriptional regulators in *Drosophila* melanogaster. *Genome Research*, *24*(7), 1224–1235. doi:10.1101/gr.168807.113

Sokolowski, M.B. (2010). Social interactions in "simple" model systems. *Neuron*, *65*(6), 780–794. doi:10.1016/j.neuron.2010.03.007

Sovik, E., LaMora, A., Seehra, G., Barron, A.B., Duncan, J.G., & Ben-Shahar, Y. (2017). *Drosophila* divalent metal ion transporter Malvolio is required in dopaminergic neurons for feeding decisions. *Genes, Brain, and Behavior*, *16*(5), 506–514. doi:10.1111/gbb.12375

Steiger, S., & Stokl, J. (2017). Pheromones involved in insect parental care and family life. *Current Opinion in Insect Science*, *24*, 89–95. doi:10.1016/j.cois.2017.09.006

Sun, Y., Liu, L., Ben-Shahar, Y., Jacobs, J.S., Eberl, D.F., & Welsh, M.J. (2009). TRPA channels distinguish gravity sensing from hearing in Johnston's organ. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(32), 13606–13611. doi:10.1073/pnas.0906377106

Suster, M.L., Karunanithi, S., Atwood, H.L., & Sokolowski, M.B. (2004). Turning behavior in *Drosophila* larvae: a role for the small scribbler transcript. *Genes, Brain, and Behavior*, *3*(5), 273–286. GBB082 [pii] doi:10.1111/j.1601-183X.2004.00082.x

Swarup, S., Huang, W., Mackay, T.F., & Anholt, R.R. (2013). Analysis of natural variation reveals neurogenetic networks for *Drosophila* olfactory behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(3), 1017–1022. doi:10.1073/pnas.1220168110

Symonds, M.R., & Wertheim, B. (2005). The mode of evolution of aggregation pheromones in *Drosophila* species. *Journal of Evolutionary Biology*, *18*(5), 1253–1263. doi:10.1111/j.1420-9101.2005.00971.x

Takahashi, K.H. (2006). Spatial aggregation and association in different resource-patch distributions: Experimental analysis with *Drosophila*. *The Journal of Animal Ecology*, *75*(1), 266–273. doi:10.1111/j.1365-2656.2006.01043.x

The Gene Ontology (2019). The gene ontology resource: 20 years and still going strong. *Nucleic Acids Research.*, *47*(D1), D330–D338. doi:10.1093/nar/gky1055

Thibert, J., Farine, J.P., Cortot, J., & Ferveur, J.F. (2016). *Drosophila* food-associated pheromones: Effect of experience, genotype and antibiotics on larval behavior. *PLoS One*, *11*(3), e0151451. doi:10.1371/journal.pone.0151451

Vernier, C.L., Chin, I.M., Adu-Oppong, B., Krupp, J.J., Levine, J., Dantas, G., & Ben-Shahar, Y. (2020). The gut microbiome defines social group membership in honey-bee colonies. *Science Advances*, *6*(42), eabd3431. doi:10.1126/sciadv.abd3431

Vernier, C.L., Krupp, J.J., Marcus, K., Hefetz, A., Levine, J.D., & Ben-Shahar, Y. (2019). The cuticular hydrocarbon profiles of honey bee workers develop via a socially-modulated innate process. *eLife*, *8*, e41855. doi:10.7554/eLife.41855

Visel, A., Rubin, E.M., & Pennacchio, L.A. (2009). Genomic views of distant-acting enhancers. *Nature*, *461*(7261), 199–205. doi:10.1038/nature08451

Warde-Farley, D., Donaldson, S.L., Comes, O., Zuberi, K., Badrawi, R., Chao, P., … Morris, Q. (2010). The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res*, *38*, W214–220. doi:10.1093/nar/gkq537

Wertheim, B., Marchais, J., Vet, L.E.M., & Dicke, M. (2002). Allee effect in larval resource exploitation in *Drosophila*: An interaction among density of adults, larvae, and micro-organisms. *Ecological Entomology*, *27*(5), 608–617. doi:10.1046/j.1365-2311.2002.00449.x

Wicker-Thomas, C., Garrido, D., Bontonou, G., Napal, L., Mazuras, N., Denis, B., … Montagne, J. (2015). Flexible origin of hydrocarbon/pheromone precursors in *Drosophila* melanogaster. *Journal of Lipid Research*, *56*(11), 2094–2101. doi:10.1194/jlr.M060368

Witte, J.S. (2010). Genome-wide association studies and beyond. *Annual Review of Public Health*, *31*(1), 9–20. doi:10.1146/annurev.publhealth.012809.103723

Wu, Q., Wen, T., Lee, G., Park, J.H., Cai, H.N., & Shen, P. (2003). Developmental control of foraging and social behavior by the *Drosophila* neuropeptide Y-like system. *Neuron*, *39*(1), 147–161. doi:10.1016/S0896-6273(03)00396-9

Xiao, X., Chang, H., & Li, M. (2017). Molecular mechanisms underlying noncoding risk variations in psychiatric genetic studies. *Molecular Psychiatry*, *22*(4), 497–511. doi:10.1038/mp.2016.241

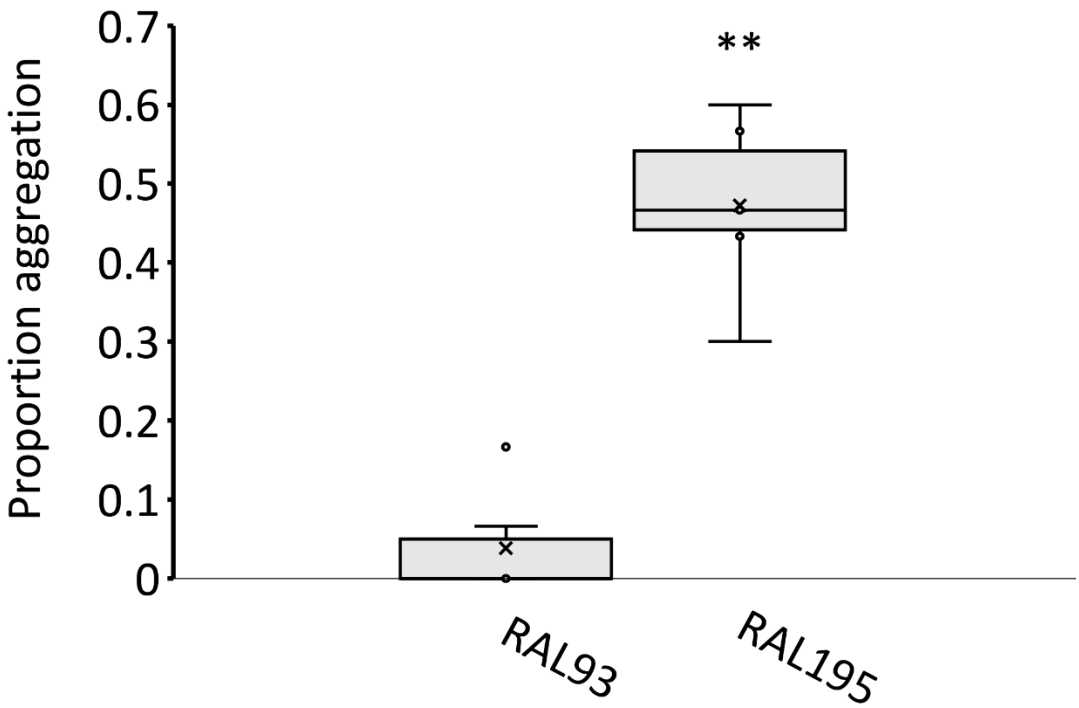
Yang, P., Shaver, S.A., Hilliker, A.J., & Sokolowski, M.B. (2000). Abnormal turning behavior in *Drosophila* larvae. Identification and molecular analysis of scribbler (sbb). *Genetics*, *155*(3), 1161–1174. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=10880478

Zelle, K., Vernier, C., Liang, X., Halloran, S., Millar, J., & Ben-Shahar, Y. (2019). A pleiotropic chemoreceptor facilitates the coupling of pheromonal signal perception and production. *bioRxiv*, 124305. doi:10.1101/124305

Zhang, Y.-N., Zhang, L.-W., Chen, D.-S., Sun, L., Li, Z.-Q., Ye, Z.-F., … Zhu, X.-Y. (2017). Molecular identification of differential expression genes associated with sex pheromone biosynthesis in Spodoptera exigua. *Molecular Genetics and Genomics : MGG*, *292*(4), 795–809. doi:10.1007/s00438-017-1307-3

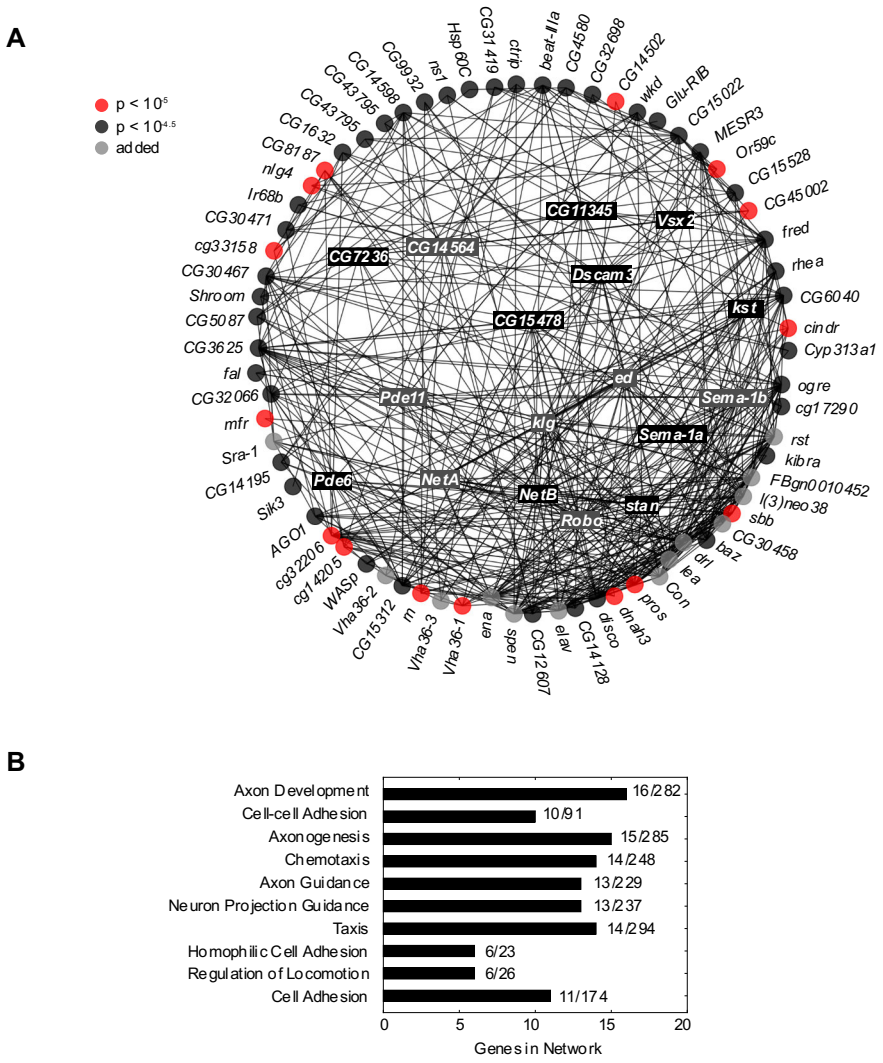
Zheng, X., Valakh, V., DiAntonio, A., & Ben-Shahar, Y. (2014). Natural antisense transcripts regulate the neuronal stress response and excitability. *eLife*, *3*, e01849. doi:10.7554/eLife.01849

Supplemental figures:



Supplemental Figure 1: Boxplots showing the proportion of aggregating 3rd instar larvae for the "Low" aggregating haplotype RAL93 and High aggregating haplotype RAL195 (N=5 plates per genotype, 30 larvae per plate). **, $p<0.01$; Mann-Whitney U Test.

,



Supplemental Figure 2: Interaction network of SNP-containing genes. **(A)** Genes containing SNPs identified in the GWAS at two significance levels ($p < 10^{-4.5}$ [black rectangles/circles] and $p < 10^{-5}$ [red circles]) were used to create a gene interaction network. Red and black circles/rectangles (nodes) in the network represent SNP-containing genes, whereas gray nodes represent genes which were allowed to be added to the network based on the extent of connectivity with SNP-containing genes (see Methods). Lines (edges) connecting each of the nodes represent any of the following interactions: (1) co-expression, (2) co-localization, (3) shared protein domains, (4) and physical protein-protein interactions. **(B)** Enriched functional gene ontologies for all genes present in the network shown in (A).

Supplemental tables:

Table S1. The *Drosophila* lines used in the study.

Table S2. The ssDNA oligos used for RT-qPCRs**.**

Table S3. The list of all SNPs identified in the GWAS.

TABLE S1

Gene Target	Stock Center	Stock ID	Genotype	Associated Figure
DGRP Line	BDSC	28122	RAL21	Figures 1-3
DGRP Line	BDSC	28123	RAL26	Figures 1-3
DGRP Line	BDSC	28128	RAL45	Figures 1-3
DGRP Line	BDSC	28129	RAL59	Figures 1-3
DGRP Line	BDSC	28132	RAL75	Figures 1-3
DGRP Line	BDSC	28274	RAL85	Figures 1-3
DGRP Line	BDSC	28136	RAL91	Figures 1-3
DGRP Line	BDSC	28137	RAL93	Figures 1-3
DGRP Line	BDSC	28138	RAL101	Figures 1-3
DGRP Line	BDSC	28142	RAL136	Figures 1-3
DGRP Line	BDSC	28145	RAL149	Figures 1-3
DGRP Line	BDSC	28148	RAL161	Figures 1-3
DGRP Line	BDSC	28153	RAL195	Figures 1-3
DGRP Line	BDSC	25174	RAL208	Figures 1-3
DGRP Line	BDSC	28157	RAL228	Figures 1-3
DGRP Line	BDSC	28160	RAL237	Figures 1-3
DGRP Line	BDSC	NA	RAL272	Figures 1-3
DGRP Line	BDSC	25175	RAL301	Figures 1-3
DGRP Line	BDSC	25177	RAL304	Figures 1-3
DGRP Line	BDSC	25180	RAL313	Figures 1-3
DGRP Line	BDSC	28168	RAL318	Figures 1-3
DGRP Line	BDSC	25182	RAL324	Figures 1-3
DGRP Line	BDSC	28179	RAL359	Figures 1-3
DGRP Line	BDSC	25186	RAL360	Figures 1-3
DGRP Line	BDSC	28180	RAL361	Figures 1-3
DGRP Line	BDSC	25445	RAL365	Figures 1-3
DGRP Line	BDSC	NA	RAL378	Figures 1-3
DGRP Line	BDSC	25189	RAL379	Figures 1-3
DGRP Line	BDSC	25190	RAL380	Figures 1-3
DGRP Line	BDSC	NA	RAL387	Figures 1-3
DGRP Line	BDSC	28194	RAL392	Figures 1-3
DGRP Line	BDSC	25192	RAL399	Figures 1-3
DGRP Line	BDSC	28196	RAL426	Figures 1-3
DGRP Line	BDSC	25193	RAL427	Figures 1-3
DGRP Line	BDSC	25194	RAL437	Figures 1-3
DGRP Line	BDSC	28198	RAL441	Figures 1-3
DGRP Line	BDSC	28199	RAL443	Figures 1-3
DGRP Line	BDSC	28204	RAL502	Figures 1-3
DGRP Line	BDSC	28208	RAL535	Figures 1-3
DGRP Line	BDSC	25198	RAL555	Figures 1-3
DGRP Line	BDSC	28212	RAL584	Figures 1-3
DGRP Line	BDSC	28218	RAL703	Figures 1-3
DGRP Line	BDSC	25744	RAL705	Figures 1-3
DGRP Line	BDSC	25201	RAL712	Figures 1-3
DGRP Line	BDSC	28219	RAL716	Figures 1-3
DGRP Line	BDSC	28222	RAL737	Figures 1-3
DGRP Line	BDSC	28223	RAL738	Figures 1-3
DGRP Line	BDSC	NA	RAL744	Figures 1-3
DGRP Line	BDSC	25206	RAL786	Figures 1-3
DGRP Line	BDSC	25208	RAL820	Figures 1-3

DGRP Line	BDSC	25209	RAL852	Figures 1-3
DGRP Line	BDSC	28250	RAL853	Figures 1-3
DGRP Line	BDSC	28253	RAL861	Figures 1-3
DGRP Line	BDSC	NA	RAL874	Figures 1-3
elav-GAL4	BDSC	25750	elav-GAL4;UAS-DCR2	Figure 4
Tub-GAL4	NA	NA	Gift from Duncan lab Wash U	Figure 4
CG8187	VDRC	v24230	P{GD13832}	Figure 4
CG14502	VDRC	v105030	P{KK113036}	Figure 4
Or59c	VDRC	v33122	P{GD626}	Figure 4
CG33158	VDRC	v23362	P{GD13424}	Figure 4
CG32206	VDRC	v102647	P{KK103727}	Figure 4
rn	VDRC	v109848	P{KK110545}	Figure 4
GFP Control	BDSC	35786	UAS-GFP.VALIUM10	Figure 4
sbb	BDSC	27049	TRiP.JF0237	Figure 4
CG43795	BDSC	51472	TRiP.HMC03214	Figure 4
Dnah3	BDSC	65239	TRiP.HMC05922	Figure 4
mfr	BDSC	61769	TRiP.HMC03324	Figure 4
Nlg4	BDSC	58119	TRiP.HMJ22056	Figure 4
CG14205	BDSC	64579	TRiP.HMC05598	Figure 4

Table S2

Gene	Forward Primer (5'-->3')	Reverse Primer (5'-->3')
CG8187	GGTGGAACAATTCAAGCTGTC	AGGCCACCCATATAAACTTGTAC
Vha36-1	ATCAAGATCACCAACCGTCG	GGCCAAAGTCCTATCGATTCG
CG14502	TGCTACCACGTTTGCTACTG	AATCTTCAGCTTAGTCTCATCCG
sbb	TCGAACTCCAAAGGGACGCG	CACCTGGGCGTGTGAGAAGT
CG43795	TCAACGTCTCCATGCTCTTC	TGCGTAGCACGATGTACAC
Or59c	CCAAATGGTGTTCGGAAAGAAG	GTTTCAGGTGCTCCCAGATTA
Dnah3	TGAGGACGACGCTGATCCTT	AGTTCAAACTGAGCCACATCATCA
mfr	GCAGAGGTTTTGATGTCAGCCG	TGGCCATTGGAATGCCCGTT
CG33158	TCCTCACTGTTCTCACTAGACC	AAGGTGTCAGCATGTGGTC
dsx-c73a	CCTCGAGCCATTTCCTTCTT	CACCTCGACAAGACACCTTT
CG32206	TTCGTAACCATCCACATCCG	AATTGCCCTTCCTCTGTCAC
rn	GCCAATGAGTTCGTGCATG	TCCGTAACCGTGTCCTTTTG
pros	GGCTTCGCAGATCCTCGACC	GTTTTGGGCGCCTGGAACATG
Nlg4	CAGTGTCCTGGCGAGCTACG	ACTCTGGCATGGGCATGTGG
cindr	GTGGGCGTCTTTCCCGACAA	GCGGTGTTGCTGACCTTCCT
CG45002	TGCGACAACACGGAGTCCAC	GGCAATCGTCTGGCTCCAGG
CG14205	ACGATTTGAGGTGCCTGGCA	GCCGTACAGGATGCCCGAAG
rp49	CACCAAGCACTTCATCCG	TCGATCCGTAACCGATGT

Table S3

SNP	Chr	Location	Major Allele	Minor Allele	-LogPVal	MAF	Assoc. Gene	SNP Gene Location
107596	chr2L	3962171	A	G	5.370784569	0.2609	NA	NA
467405	chr2L	16488108	G	C	5.253456939	0.02273	NA	NA
467406	chr2L	16488111	A	C	5.136794735	0.02222	NA	NA
748617	chr2R	9463144	A	T	5.18763654	0.2553	NA	NA
748619	chr2R	9463164	A	C	5.18763654	0.2553	NA	NA
748653	chr2R	9463447	T	G	5.18763654	0.2553	NA	NA
799710	chr2R	11427595	G	T	5.232345742	0.3191	cg8187	Missense E to D
799711	chr2R	11427619	T	C	5.232345742	0.3191	cg8187	Synonymous
799712	chr2R	11427715	T	C	6.023327692	0.3333	cg8187	Synonymous
799737	chr2R	11428706	A	G	6.80698681	0.3404	cg8187	Missense D to G
799741	chr2R	11428807	A	T	6.683252763	0.3191	cg8187	Synonymous
799747	chr2R	11429074	C	G	5.492773834	0.2766	vha36	5prime UTR
872755	chr2R	14179864	T	C	5.070107064	0.2979	cg14502	5prime UTR
872757	chr2R	14179888	A	C	5.133368991	0.3125	cg14502	5prime UTR
872766	chr2R	14180184	C	T	5.629316744	0.1522	cg14502	5prime UTR
872791	chr2R	14181064	G	T	5.805424167	0.2708	cg14502	5prime UTR
872800	chr2R	14181157	C	T	5.805424167	0.2708	cg14502	5prime UTR
872755	chr2R	14179864	T	C	5.070107064	0.2979	sbb	5prime UTR
872757	chr2R	14179888	A	C	5.133368991	0.3125	sbb	5prime UTR
872766	chr2R	14180184	C	T	5.629316744	0.1522	sbb	5prime UTR
872791	chr2R	14181064	G	T	5.805424167	0.2708	sbb	5prime UTR
872800	chr2R	14181157	C	T	5.805424167	0.2708	sbb	5prime UTR
916242	chr2R	15864115	T	A	5.730712698	0.383	NA	NA
922405	chr2R	16114803	G	A	5.179710513	0.3958	NA	NA
1021666	chr2R	19360584	T	A	5.02887411	0.1778	cg43795	Intron
1021666	chr2R	19360584	T	A	5.02887411	0.1778	or59c	Missense F to L
1175091	chr3L	4448433	A	G	5.048749197	0.08511	NA	NA
1175121	chr3L	4457161	T	G	5.048749197	0.08511	NA	NA
1175166	chr3L	4458652	C	A	5.048749197	0.08511	NA	NA
1175173	chr3L	4458904	A	G	5.048749197	0.08511	NA	NA
1187085	chr3L	4894924	A	T	5.243940627	0.08511	dnah3	Intron/3prime UTR
1187086	chr3L	4894928	T	C	5.243940627	0.08511	dnah3	Intron/3prime UTR
1223179	chr3L	6276454	T	C	5.475742001	0.3913	NA	NA
1227224	chr3L	6458617	A	G	5.152110183	0.06383	NA	NA
1295701	chr3L	8902912	T	C	5.716209361	0.4783	mfr	Synonymous
1506307	chr3L	16435860	C	A	5.571593164	0.1915	cg33158	Intron
1506309	chr3L	16435894	A	T	5.728142431	0.1702	cg33158	Intron
1506310	chr3L	16435901	C	T	5.728142431	0.1702	cg33158	Intron
1506312	chr3L	16435910	C	T	5.728142431	0.1702	cg33158	Intron
1506307	chr3L	16435860	C	A	5.571593164	0.1915	dsx-c73a	Intron
1506309	chr3L	16435894	A	T	5.728142431	0.1702	dsx-c73a	Intron
1506310	chr3L	16435901	C	T	5.728142431	0.1702	dsx-c73a	Intron
1506312	chr3L	16435910	C	T	5.728142431	0.1702	dsx-c73a	Intron
1569516	chr3L	19335533	T	A	5.781562504	0.08511	cg32206	3prime UTR
1569661	chr3L	19343297	C	T	6.081256409	0.1458	cg32206	Intron
1569670	chr3L	19343994	A	G	6.241253379	0.1667	cg32206	Intron

1569694	chr3L	19344794	T	C	5.206736999	0.1277	cg32206	Intron
1569713	chr3L	19345400	G	T	6.081256409	0.1458	cg32206	Intron
1569738	chr3L	19345970	G	A	5.081434311	0.08511	cg32206	Intron
1569758	chr3L	19346948	C	T	5.206736999	0.1277	cg32206	Intron
1569784	chr3L	19348702	A	G	5.302563203	0.125	cg32206	Intron
1569796	chr3L	19349135	T	G	6.081256409	0.1458	cg32206	Intron
1569797	chr3L	19349136	A	T	6.081256409	0.1458	cg32206	Intron
1629657	chr3R	2288310	C	T	5.192298662	0.2609	NA	NA
1629658	chr3R	2288311	T	G	5.192298662	0.2609	NA	NA
1638625	chr3R	3117993	A	T	5.203421611	0.4444	rn	5prime UTR/Intron
1688186	chr3R	7216825	A	C	5.105296094	0.3556	pros	Intron
1800140	chr3R	13111505	T	A	5.192725136	0.3696	NA	NA
1848321	chr3R	14926325	T	C	5.413535243	0.2979	NA	NA
1873218	chr3R	16040523	A	T	5.16486728	0.08333	nlg4	Intron
2106576	chr3R	26074206	T	C	5.152110183	0.06383	NA	NA
2106577	chr3R	26074213	C	G	5.152110183	0.06383	NA	NA
2106579	chr3R	26074224	G	A	5.152110183	0.06383	NA	NA
2124426	chr3R	26649080	G	A	5.200048167	0.3191	cindr	5prime UTR/Intron
2206010	chrX	4879512	A	T	5.630389811	0.1277	NA	NA
2395384	chrX	16892876	G	T	5.04331991	0.3542	cg45002	5prime UTR/Intron
2395451	chrX	16893967	A	G	5.385341465	0.2708	cg45002	5prime UTR/Intron
2435484	chrX	19488026	C	G	5.34560268	0.5	cg14205	Intron