The Genetic Architecture of Larval Aggregation Behavior in Drosophila

Ross M McKinney[a] and Yehuda Ben-Shahar[a]

[a]*Department of Biology, Washington University in St. Louis, St. Louis, MO, USA*

Contact: Yehuda Ben-Shahar, benshahary@wustl.edu, Department of Biology, Washington University in St. Louis, St. Louis, MO 63130

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 SNP 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.

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. Yet, 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 et al., 2016; Louis & de Polavieja, 2017; Rooke et al., 2020; Steiger & Stokl, 2017; Symonds & Wertheim, 2005; Thibert et al., 2016). Specifically, in *Drosophila melanogaster*, at least two pheromones produced by larvae have been shown to act as chemoattractants (Mast et al., 2014). However, the downstream neural and genetic pathways that regulate larval aggregation behavior remain largely unexplored.

To optimize fitness, 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 et al., 2016) and genetics (Allen et al., 2017; Fitzpatrick et al., 2007; Kaun, Hendel, et al., 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.

Materials and Methods:

Animals

All fly lines were reared on standard corn syrup-soy food (Achron Scientific), and kept under a 12h:12h light:dark schedule at 25 ˚C and 60% humidity. Lines from the Drosophila Genetic Reference Panel (DGRP) (Mackay et al., 2012) used in this study are available from the Bloomington Drosophila Stock Center (BDSC, Bloomington, IN). UAS-RNAi lines and the *elav*- and *tubulin*-GAL4 lines were from either the Bloomington Drosophila Stock Center or the Vienna Drosophila Resource Center (VDRC) (Dietzl et al., 2007; Perkins et al., 2015). All fly lines used in this study, along with their stock numbers and genotypes, are listed in Table S1.

Larval aggregation assays

Larval aggregation was assayed as follows. Approximately 30, second/third instar larvae were collected from standard vials using a 15% sucrose solution (w/v). Larvae were placed onto the center of a 60mm petri dish containing 20% apple juice (v/v) and 1% agar (w/v) *en masse* and allowed to roam the plate freely for 15 minutes. Subsequently, a picture of the plate was taken (Figure 1A), and the fraction of aggregating larvae was calculated as described below. All behavioral assays were conducted at 25 ^{o}C and 70 % humidity.

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.

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 server (Grimm et al., 2017), with default parameters, to search for genotype by phenotype associations. A total of 2,370,987 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 PLINK (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-value 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-value 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 GO terms 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 qRT-PCR

mRNA was collected from groups of 10 whole larvae (n=3–4 replicates per line) using Trizol (ThermoFisher) and reverse transcribed to cDNA using SuperScript III Reverse Transcriptase (ThermoFisher). Sybr Green (ThermoFisher) was used to amplify and quantify expression levels for all genes containing significant SNPs 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 (Hill et al., 2017; Lu et al., 2012; Vernier et al., 2019). All qPCR primers used in this study are listed in 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) and placed into the pUCIDT-ampR plasmid (IDT). We subcloned this region into the pENTR-1A plasmid (ThermoFisher) using KpnI and XhoI restriction sites on either side of the promoter, and then used Gateway cloning (ThermoFisher) 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}. The *CG14205*-GAL4 line was crossed with *UAS-mCD8::GFP* (BDSC #32188) and imaged in third-instar larvae.

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 *Drosophila* Genetic Reference Panel (DGRP) (Mackay et al., 2012) for levels of aggregation in third instar larvae and subsequently performed a genome wide association study (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 ANOVAs to search for genetic variation (SNPs) associated with the mean fraction of aggregating larvae across lines (Shorter et al., 2015; Swarup et al., 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, 17 protein coding genes that fall within 500 bp of these SNPs were further considered as candidate genes that might be playing a role in larval aggregation decisions (Figure 2, Table S3).

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 gene ontology (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 analysed 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 1A). 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 1B). Together, these data suggest that at least some of the genetic variations we have identified impact

169 population level phenotypic variations in aggregation decisions via neuronal functions.

170 ***Genetic variations associated with mRNA expression levels***

171 Single nucleotide polymorphisms falling within promoter and enhancer regions of a
172 protein coding gene often affect mRNA expression levels (Khurana et al., 2016; Nord &
173 West, 2020; Visel et al., 2009). Since most of the SNPs we have identified in our
174 GWAS are either intronic or fall upstream of their associated genes (37/46; Table S3),
175 we next tested the hypothesis that some of the identified SNPs affect gene action via
176 their effects on mRNA expression levels. To test this hypothesis, we compared the
177 mRNA expression levels of each of the 17 candidate genes identified in our initial
178 conservative screen between the three phenotypically highest ("High") and three lowest
179 ("Low") aggregating DGRP lines (Figure 3A, and B) by using real-time qRT-PCR
180 analyses. We found that at least one SNP (X:19488026) was significantly associated
181 with higher mRNA expression levels of its parent gene, *CG14205*, in all "low" lines
182 relative to all "high" lines (one-way ANOVA; $F(1,4) = 13.43$, $p = 0.02$) (Figure 3).
183 These results suggest that this specific SNP is playing a role in regulating the expression
184 or stability of the *CG14205* mRNAs. The location of this SNP immediately downstream
185 of a predicted splice donor site in the annotated intron 5 of *CG14205* (Figure 3C)
186 suggests that it may affect splicing and/ or stability of the pre-mRNA. Furthermore, we
187 found a significant interaction between *CG14205* expression level and SNP genotype on
188 the levels of aggregation between High and Low lines (two-way ANOVA; $F(2,3) =$
189 403.3, $p < 0.01$). As *CG14205* expression is significantly higher in Low lines than in
190 High lines, these data suggest that higher expression levels of *CG14205* may reduce
191 aggregation in *D. melanogaster* larvae.

Although the biological functions of *CG14205* are unknown, the protein is predicted to be 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 et al., 2006). However, the specific biochemical functions of these membrane-bound acyltransferases remain 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; Mengfang Du et al., 2015; M. Du et al., 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 knockdown of four of the genes (*CG8187*, *CG14502*, *CG32206*, and *rn*) lead to higher levels of aggregation relative to controls (Figure 4**A**, **B**, and **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. Given the strong association between the specific *CG14205* alleles, mRNA expression levels, and aggregation levels, we next tested whether genetic variation in this specific gene affect aggregation decision via its action in non-neuronal tissues by using the ubiquitous *tubulin*-GAL4 driver to knockdown *CG14205* in all tissues. 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, $p = 0.025$; Figure 4**D**). These results suggest that *CG14205* functions to suppress aggregation in *D. melanogaster* larvae via neuronal-independent signalling pathways in the larval midgut.

While we do not know yet how the midgut activity levels of *CG14205* might affect the decision of individual larvae to join a group, it is likely that this decision controlled by both external sensory stimuli and internal receptors which detect those stimuli. It is possible that the *CG14205* gene is responsible for the biosynthesis or release of a sensory stimulus which 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 larval midgut (Buchon et al., 2013; Celniker et al., 2009) (Figure 5A-B). This expression pattern was further confirmed by imaging the transgenic expression of GFP under the control of the *CG14205* promoter, which revealed strong expression in the most proximal and distal parts of the midgut (Figure 5C-H). Together, these results suggest that *CG14205* plays a

role in 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 depended 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 quantitative expression variations across different alleles of the *CG14205* gene, which encodes a putative acetyl transferase, are strongly associated with larval aggregation while feeding; DGRP lines that exhibit low levels of aggregation express higher levels of *CG14205* transcripts relative to those that display high levels of aggregation (Figure 3B). These data suggest that the activity of *CG14205* inhibits the formation of larval aggregates. 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 3C), which may affect mRNA splicing and/ or stability via posttranscriptional processes. How *CG14205* activity in the gut might regulate larval aggregation remains unknown. 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 et al., 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 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 et al., 2014; Wicker-Thomas et al., 2015; Zelle et al., 2019), our data indicate that gut derived metabolites can also possibley 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 signalling 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 et al., 2014; McKinney et al., 2015; Zelle et al., 2019). However, in our study, the knockdown of all identified candidate genes leads to increased levels of larval aggregation, which suggest that the primary contributions of these genes are to suppression of aggregation. One possible interpretation of these data is that in natural populations of *D. melanogaster*, it may be

that it is more beneficial for larvae to supress aggregation as a function of density to maximize larval fitness. 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.

Nevertheless, 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]

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. https://doi.org/10.1534/genetics.116.197939

Allison, J. D., & Cardé, R. T. (2016). *Pheromone communication in moths : evolution, behavior, and application* [still image]. University of California Press. https://doi.org/40026787697

Anreiter, I., Kramer, J., & Sokolowski, M. B. (2017). Epigenetic mechanisms modulate differences in Drosophila foraging behavior. *PNAS*.

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 Res*, *36*(Web Server issue), W347-351. https://doi.org/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. https://doi.org/10.1371/journal.pone.0009395

Blomquist, G. J., Figueroa-Teran, R., Aw, M., Song, M., Gorzalski, A., Abbott, N. L., Chang, E., & Tittiger, C. (2010). Pheromone production in bark beetles. *Insect Biochem Mol Biol*, *40*(10), 699-712. https://doi.org/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*. 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 Rep*, *3*(5), 1725-1738. https://doi.org/10.1016/j.celrep.2013.04.001

Celniker, S. E., Dillon, L. A., Gerstein, M. B., Gunsalus, K. C., Henikoff, S., Karpen, G. H., Kellis, M., Lai, E. C., Lieb, J. D., MacAlpine, D. M., Micklem, G., Piano, F., Snyder, M., Stein, L., White, K. P., & Waterston, R. H. (2009). Unlocking the secrets of the genome. *Nature*, *459*(7249), 927-930. https://doi.org/459927a [pii] 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. *Sci Rep*, *9*(1), 1477. https://doi.org/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. https://doi.org/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. *Mol Cell*, *4*(2), 143-152. https://doi.org/10.1016/s1097-2765(00)80362-7

Dietzl, G., Chen, D., Schnorrer, F., Su, K. C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., Couto, A., Marra, V., Keleman, K., & Dickson, B. J. (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. *Nature*, *448*(7150), 151-156. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17625558

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. https://doi.org/10.1007/s11745-016-4122-4

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. *Sci Rep*, *5*(1), 8110. https://doi.org/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. *J Insect Physiol*, *58*(5), 699-703. https://doi.org/10.1016/j.jinsphys.2012.02.002

366 Durisko, Z., & Dukas, R. (2013). Attraction to and learning from social cues in fruitfly
367 larvae. *Proc Biol Sci*, *280*(1767), 20131398.
368 https://doi.org/10.1098/rspb.2013.1398
369 Fitzpatrick, M. J., Feder, E., Rowe, L., & Sokolowski, M. B. (2007). Maintaining a
370 behaviour polymorphism by frequency-dependent selection on a single gene.
371 *Nature*, *447*(7141), 210-212. https://doi.org/nature05764
372 [pii]10.1038/nature05764
373 Grimm, D. G., Roqueiro, D., Salome, P. A., Kleeberger, S., Greshake, B., Zhu, W., Liu,
374 C., Lippert, C., Stegle, O., Scholkopf, B., Weigel, D., & Borgwardt, K. M.
375 (2017). easyGWAS: A Cloud-Based Platform for Comparing the Results of
376 Genome-Wide Association Studies. *Plant Cell*, *29*(1), 5-19.
377 https://doi.org/10.1105/tpc.16.00551
378 Groot, A. T., Staudacher, H., Barthel, A., Inglis, O., Schofl, G., Santangelo, R. G.,
379 Gebauer-Jung, S., Vogel, H., Emerson, J., Schal, C., Heckel, D. G., & Gould, F.
380 (2013). One quantitative trait locus for intra- and interspecific variation in a sex
381 pheromone. *Mol Ecol*, *22*(4), 1065-1080. https://doi.org/10.1111/mec.12171
382 Hill, A., Zheng, X., Li, X., McKinney, R., Dickman, D., & Ben-Shahar, Y. (2017). The
383 Drosophila Postsynaptic DEG/ENaC Channel ppk29 Contributes to Excitatory
384 Neurotransmission. *J Neurosci*, *37*(12), 3171-3180.
385 https://doi.org/10.1523/JNEUROSCI.3850-16.2017
386 Hunt, D. W., & Borden, J. H. (1990). Conversion of verbenols to verbenone by yeasts
387 isolated fromDendroctonus ponderosae (Coleoptera: Scolytidae). *J Chem Ecol*,
388 *16*(4), 1385-1397. https://doi.org/10.1007/BF01021034
389 Kaun, K. R., Hendel, T., Gerber, B., & Sokolowski, M. B. (2007). Natural variation in
390 Drosophila larval reward learning and memory due to a cGMP-dependent
391 protein kinase. *Learn Mem*, *14*(5), 342-349. https://doi.org/14/5/342 [pii]
392 10.1101/lm.505807
393 Kaun, K. R., Riedl, C. A., Chakaborty-Chatterjee, M., Belay, A. T., Douglas, S. J.,
394 Gibbs, A. G., & Sokolowski, M. B. (2007). Natural variation in food acquisition
395 mediated via a Drosophila cGMP-dependent protein kinase. *J Exp Biol*, *210*(Pt
396 20), 3547-3558. https://doi.org/210/20/3547 [pii]
397 10.1242/jeb.006924
398 Khurana, E., Fu, Y., Chakravarty, D., Demichelis, F., Rubin, M. A., & Gerstein, M.
399 (2016). Role of non-coding sequence variants in cancer. *Nat Rev Genet*, *17*(2),
400 93-108. https://doi.org/10.1038/nrg.2015.17
401 Leonhardt, S. D., Menzel, F., Nehring, V., & Schmitt, T. (2016). Ecology and Evolution
402 of Communication in Social Insects. *Cell*, *164*(6), 1277-1287.
403 https://doi.org/10.1016/j.cell.2016.01.035
404 Lihoreau, M., Clarke, I. M., Buhl, J., Sumpter, D. J., & Simpson, S. J. (2016).
405 Collective selection of food patches in Drosophila. *J Exp Biol*, *219*(Pt 5), 668-
406 675. https://doi.org/10.1242/jeb.127431
407 Louis, M., & de Polavieja, G. (2017). Collective Behavior: Social Digging in
408 Drosophila Larvae. *Curr Biol*, *27*(18), R1010-R1012.
409 https://doi.org/10.1016/j.cub.2017.08.023
410 Lu, B., LaMora, A., Sun, Y., Welsh, M. J., & Ben-Shahar, Y. (2012). ppk23-Dependent
411 chemosensory functions contribute to courtship behavior in Drosophila
412 melanogaster. *PLoS Genet*, *8*(3), e1002587.
413 https://doi.org/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. *Biol Open*, *In Press*(2), 152-160. https://doi.org/10.1242/bio.20147369

Mackay, T. F. C., Richards, S., Stone, E. A., Barbadilla, A., Ayroles, J. F., Zhu, D., Casillas, S., Han, Y., Magwire, M. M., & Cridland, J. M. (2012). The Drosophila melanogaster Genetic Reference Panel. *Nature*, *482*(7384), 173-178.

Makki, R., Cinnamon, E., & Gould, A. P. (2014). The development and functions of oenocytes. *Annu Rev Entomol*, *59*, 405-425. https://doi.org/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. https://doi.org/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. https://doi.org/http://dx.doi.org/10.1016/j.cois.2015.09.010

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

Perkins, L. A., Holderbaum, L., Tao, R., Hu, Y., Sopko, R., McCall, K., Yang-Zhou, D., Flockhart, I., Binari, R., Shim, H.-S., Miller, A., Housden, A., Foos, M., Randkelv, S., Kelley, C., Namgyal, P., Villalta, C., Liu, L.-P., Jiang, X., Huan-Huan, Q., Wang, X., Fujiyama, A., Toyoda, A., Ayers, K., Blum, A., Czech, B., Neumuller, R., Yan, D., Cavallaro, A., Hibbard, K., Hall, D., Cooley, L., Hannon, G. J., Lehmann, R., Parks, A., Mohr, S. E., Ueda, R., Kondo, S., Ni, J.-Q., & Perrimon, N. (2015). The Transgenic RNAi Project at Harvard Medical School: Resources and Validation. *Genetics*, *201*(3), 843-852. https://doi.org/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. https://doi.org/genetics.110.119917 [pii]
10.1534/genetics.110.119917

Prokopy, R. J., & Roitberg, B. D. (2001). Joining and avoidance behavior in nonsocial insects. *Annu Rev Entomol*, *46*, 631-665. https://doi.org/10.1146/annurev.ento.46.1.631
46/1/631 [pii]

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, *81*(3), 559-575. https://doi.org/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. *Commun Biol*, *3*(1), 304. https://doi.org/10.1038/s42003-020-1024-z

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. *Proc Natl Acad Sci U S A*, *112*(27), E3555-3563. https://doi.org/10.1073/pnas.1510104112

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

Steiger, S., & Stokl, J. (2017). Pheromones involved in insect parental care and family life. *Curr Opin Insect Sci*, *24*, 89-95. https://doi.org/10.1016/j.cois.2017.09.006

464 Swarup, S., Huang, W., Mackay, T. F., & Anholt, R. R. (2013). Analysis of natural
465 variation reveals neurogenetic networks for Drosophila olfactory behavior. *Proc*
466 *Natl Acad Sci U S A*, *110*(3), 1017-1022.
467 https://doi.org/10.1073/pnas.1220168110
468 Symonds, M. R., & Wertheim, B. (2005). The mode of evolution of aggregation
469 pheromones in Drosophila species. *J Evol Biol*, *18*(5), 1253-1263.
470 https://doi.org/10.1111/j.1420-9101.2005.00971.x
471 Thibert, J., Farine, J. P., Cortot, J., & Ferveur, J. F. (2016). Drosophila Food-Associated
472 Pheromones: Effect of Experience, Genotype and Antibiotics on Larval
473 Behavior. *PLoS One*, *11*(3), e0151451.
474 https://doi.org/10.1371/journal.pone.0151451
475 Vernier, C. L., Krupp, J. J., Marcus, K., Hefetz, A., Levine, J. D., & Ben-Shahar, Y.
476 (2019). The cuticular hydrocarbon profiles of honey bee workers develop via a
477 socially-modulated innate process. *Elife*, *8*. https://doi.org/10.7554/eLife.41855
478 Visel, A., Rubin, E. M., & Pennacchio, L. A. (2009). Genomic views of distant-acting
479 enhancers. *Nature*, *461*(7261), 199-205. https://doi.org/10.1038/nature08451
480 Warde-Farley, D., Donaldson, S. L., Comes, O., Zuberi, K., Badrawi, R., Chao, P.,
481 Franz, M., Grouios, C., Kazi, F., Lopes, C. T., Maitland, A., Mostafavi, S.,
482 Montojo, J., Shao, Q., Wright, G., Bader, G. D., & Morris, Q. (2010). The
483 GeneMANIA prediction server: biological network integration for gene
484 prioritization and predicting gene function. *Nucleic Acids Res*, *38*(Web Server
485 issue), W214-220. https://doi.org/10.1093/nar/gkq537
486 Wicker-Thomas, C., Garrido, D., Bontonou, G., Napal, L., Mazuras, N., Denis, B.,
487 Rubin, T., Parvy, J. P., & Montagne, J. (2015). Flexible origin of
488 hydrocarbon/pheromone precursors in Drosophila melanogaster. *J Lipid Res*,
489 *56*(11), 2094-2101. https://doi.org/10.1194/jlr.M060368
490 Wu, Q., Wen, T., Lee, G., Park, J. H., Cai, H. N., & Shen, P. (2003). Developmental
491 control of foraging and social behavior by the Drosophila neuropeptide Y-like
492 system. *Neuron*, *39*(1), 147-161. https://doi.org/S0896627303003969 [pii]
493 Zelle, K., Vernier, C., Liang, X., Halloran, S., Millar, J., & Ben-Shahar, Y. (2019). A
494 pleiotropic chemoreceptor facilitates the coupling of pheromonal signal
495 perception and production. *bioRxiv*, 124305. https://doi.org/10.1101/124305
496 Zhang, Y.-N., Zhang, L.-W., Chen, D.-S., Sun, L., Li, Z.-Q., Ye, Z.-F., Zheng, M.-Y.,
497 Li, J.-B., & Zhu, X.-Y. (2017). Molecular identification of differential
498 expression genes associated with sex pheromone biosynthesis in Spodoptera
499 exigua. *Molecular Genetics and Genomics*, *292*(4), 795-809.
500 https://doi.org/10.1007/s00438-017-1307-3

501

502

Figures

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 showing the fraction of aggregating larvae for each of the 48 DGRP lines that were included in the GWAS (n=5–9 replicates per line); outliers are shown with open circles. DGRP lines with either low (Low) or high (High) levels of aggregation that were used in subsequent analyses are labeled and shown in either light or dark blue, respectively.

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-value less than 10−5 (shown by the dashed gray line) were retained for further analysis and are outlined in red. **(B)** A 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. Some SNPs fell within the coding region of genes on both the plus and minus strand of DNA and are labelled accordingly. (Bottom) Linkage disequilibrium matrix between all of the retained SNPs.

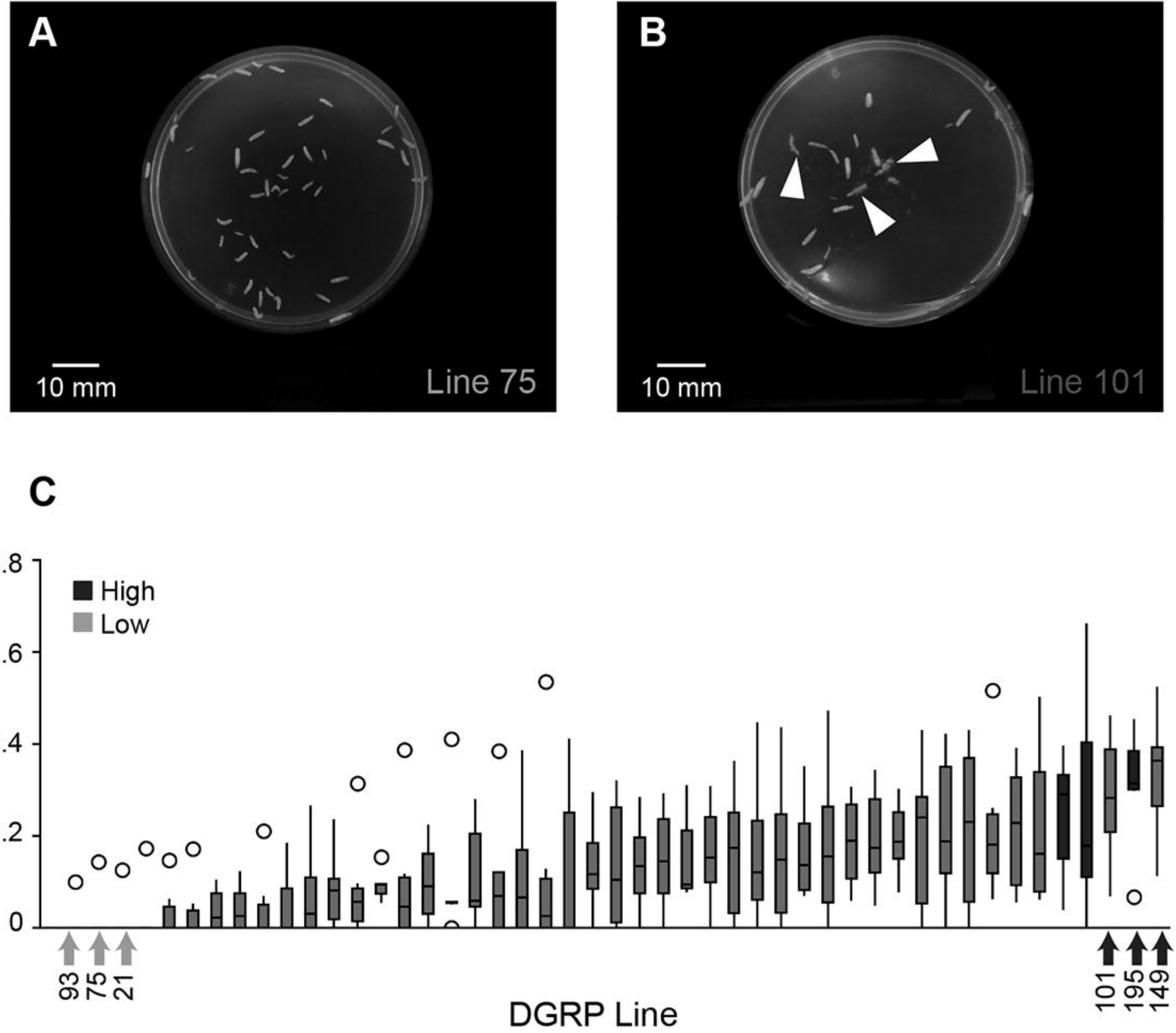
Figure 3: mRNA expression analysis of SNP-containing genes in lines with either low or high levels of aggregation. (A) Relative mRNA expression levels for each of the SNP-containing genes identified in the GWAS (n=3–4 replicates per line). Low aggregating lines are shown in light blue, and High aggregating lines are shown in dark blue. **(B)** Relative mRNA expression levels for the *CG14205* gene. A significant

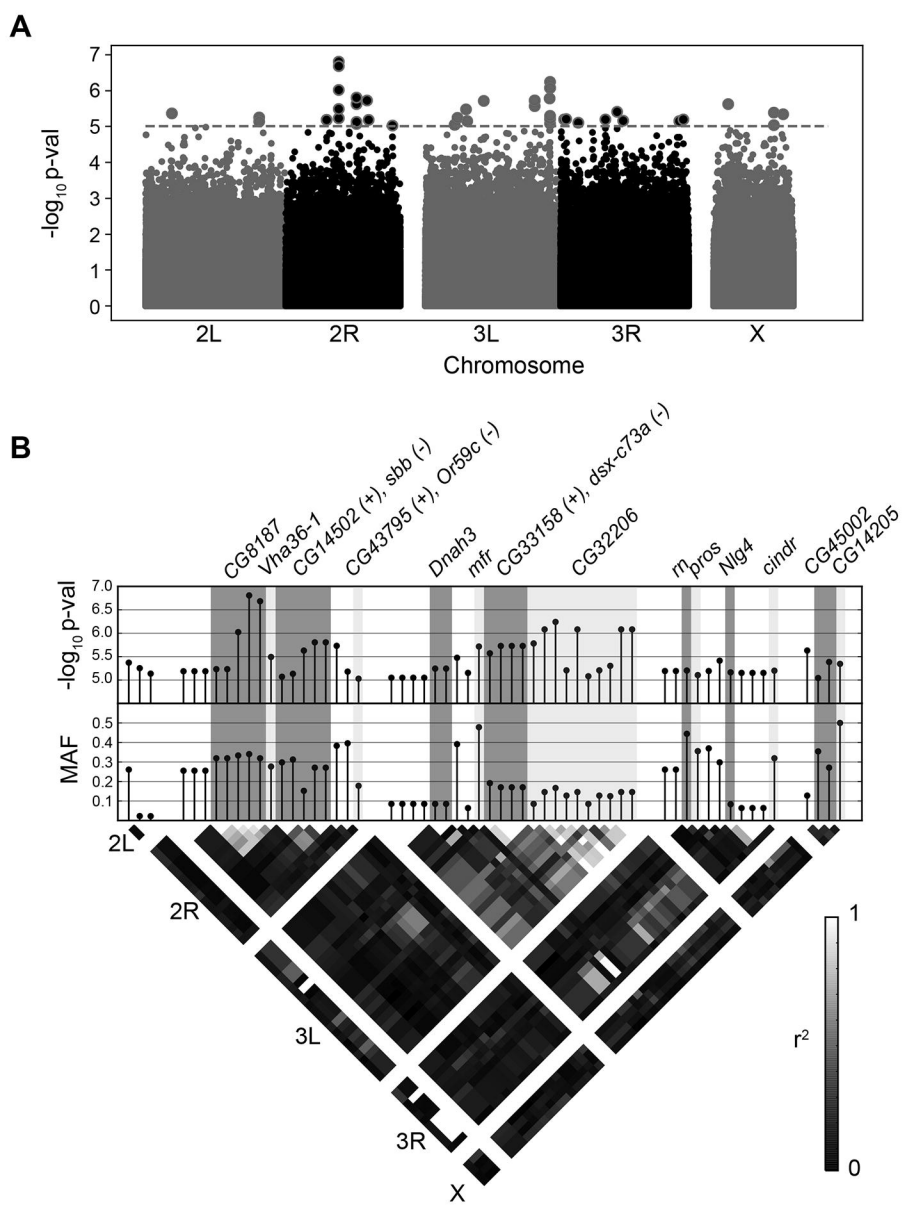
association between SNP genotype and *CG14205* mRNA expression was identified ($p < 0.05$; one-way ANOVA), whereby Low aggregating lines had higher levels of expression than High aggregating lines. Note that Low and High lines segregated by genotype, as shown in (C). **(C)** Transformed p-values for 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 red arrow head), falls just past the exon-intron boundary within intron 5 and is positioned to potentially effect mRNA splicing.

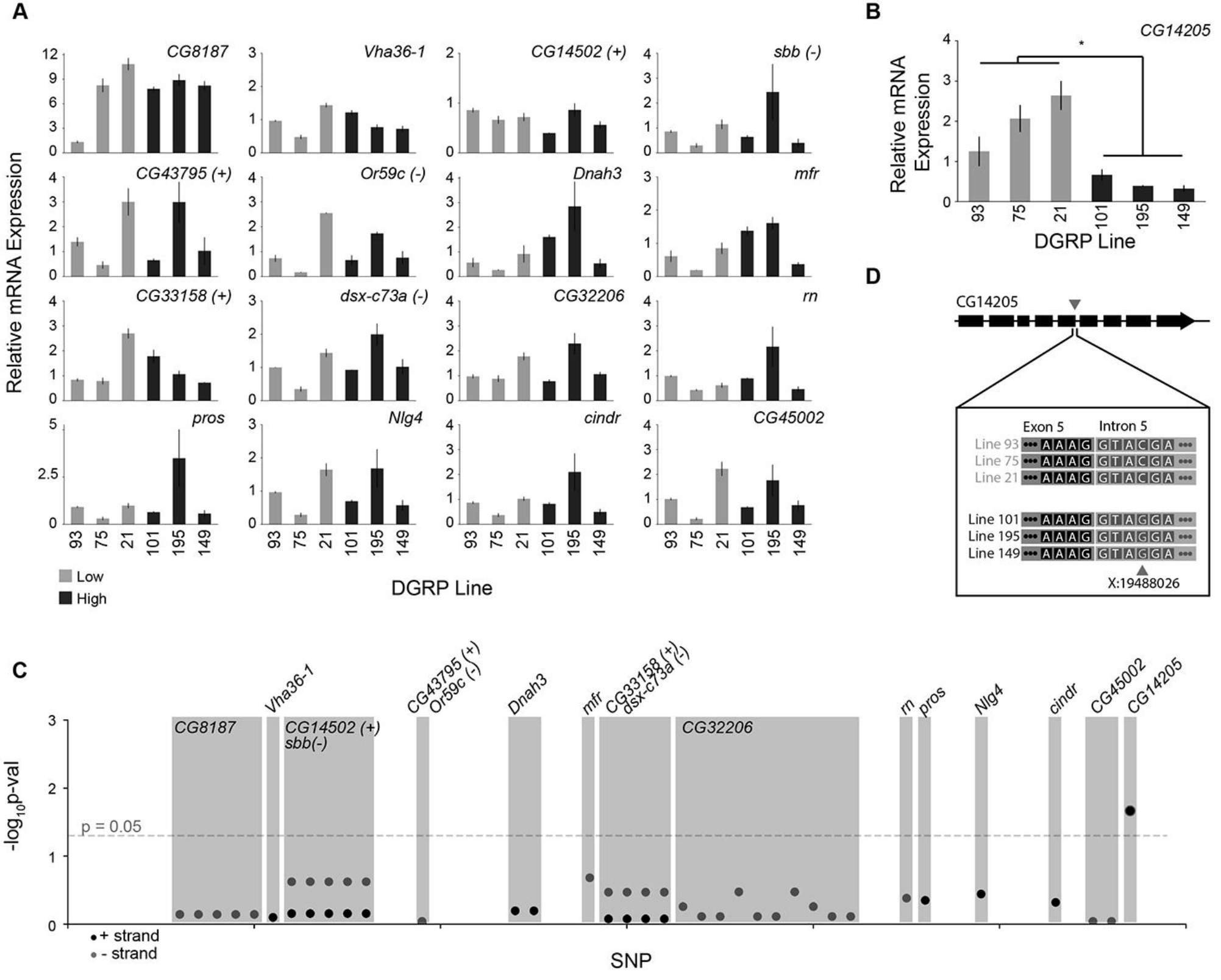
Figure 4: Neuronal knockdown of some candidate genes leads to altered aggregation behavior. (A) Pan-neuronal RNAi-mediated knockdown of SNP-associated genes (UAS-RNAi lines from the Vienna Drosophila Resource Center). Knockdown of *CG8187* (n=5-8, $p < 0.05$), *CG14502* (n=8–9, $p < 0.05$), *CG32206* (n=7–9, $p < 0.01$), or *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 Tukeys 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 Students 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

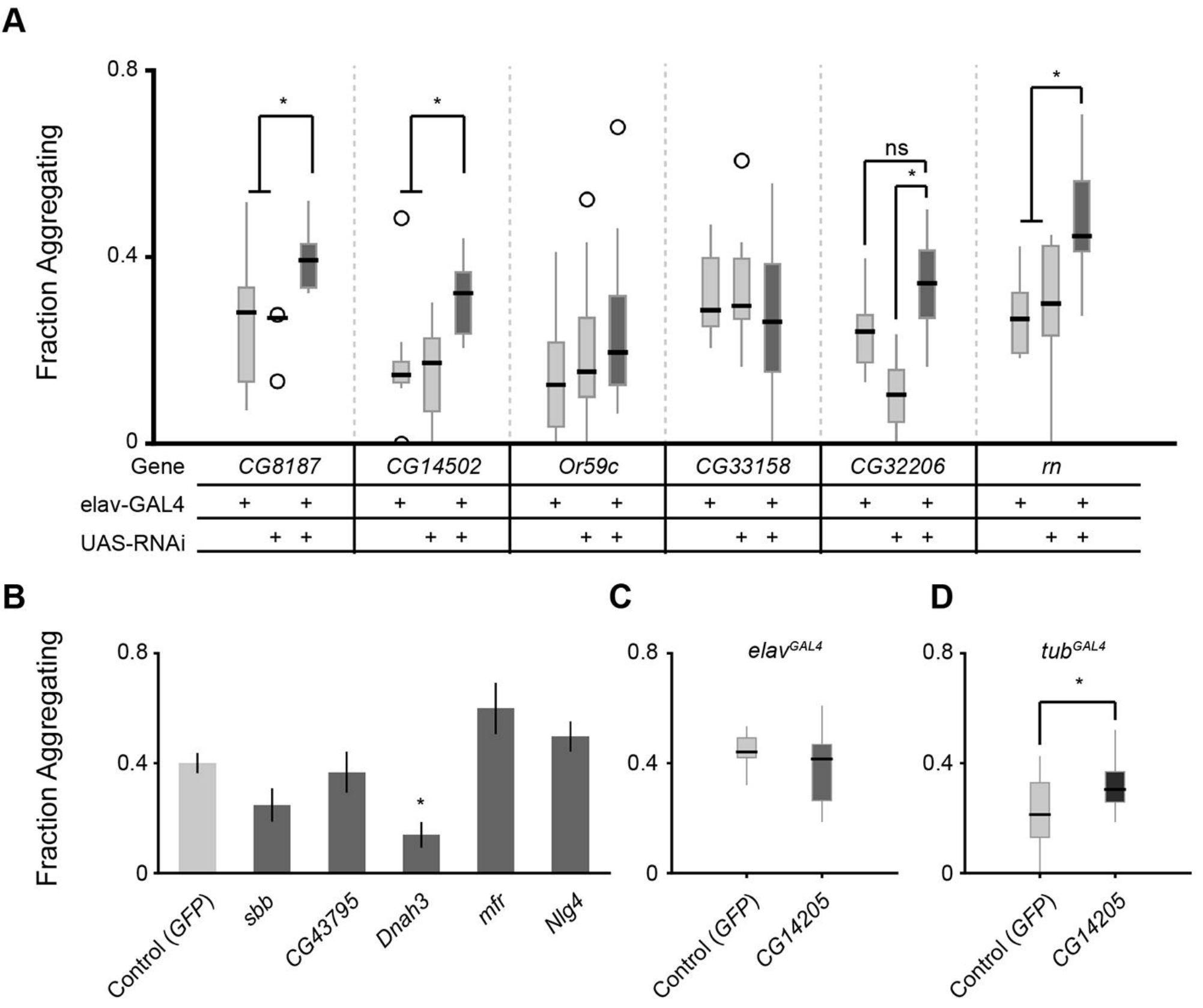
552 Bonferroni correction. **(C)** TRiP-RNAi-mediated knockdown of *CG14205* in neural
553 tissues, using the *elav*-GAL4 driver, did not lead to altered aggregation (n=8 per group,
554 $p > 0.05$; onetailed Students T-test). **(D)** TRiP-RNAi-mediated knockdown of *CG14205*
555 in all tissues, using the *tubulin*-GAL4 driver, led to a significant increase in the fraction
556 of larvae aggregating compared to control (n=11–12, $p = 0.025$; one-tailed Students T-
557 test).

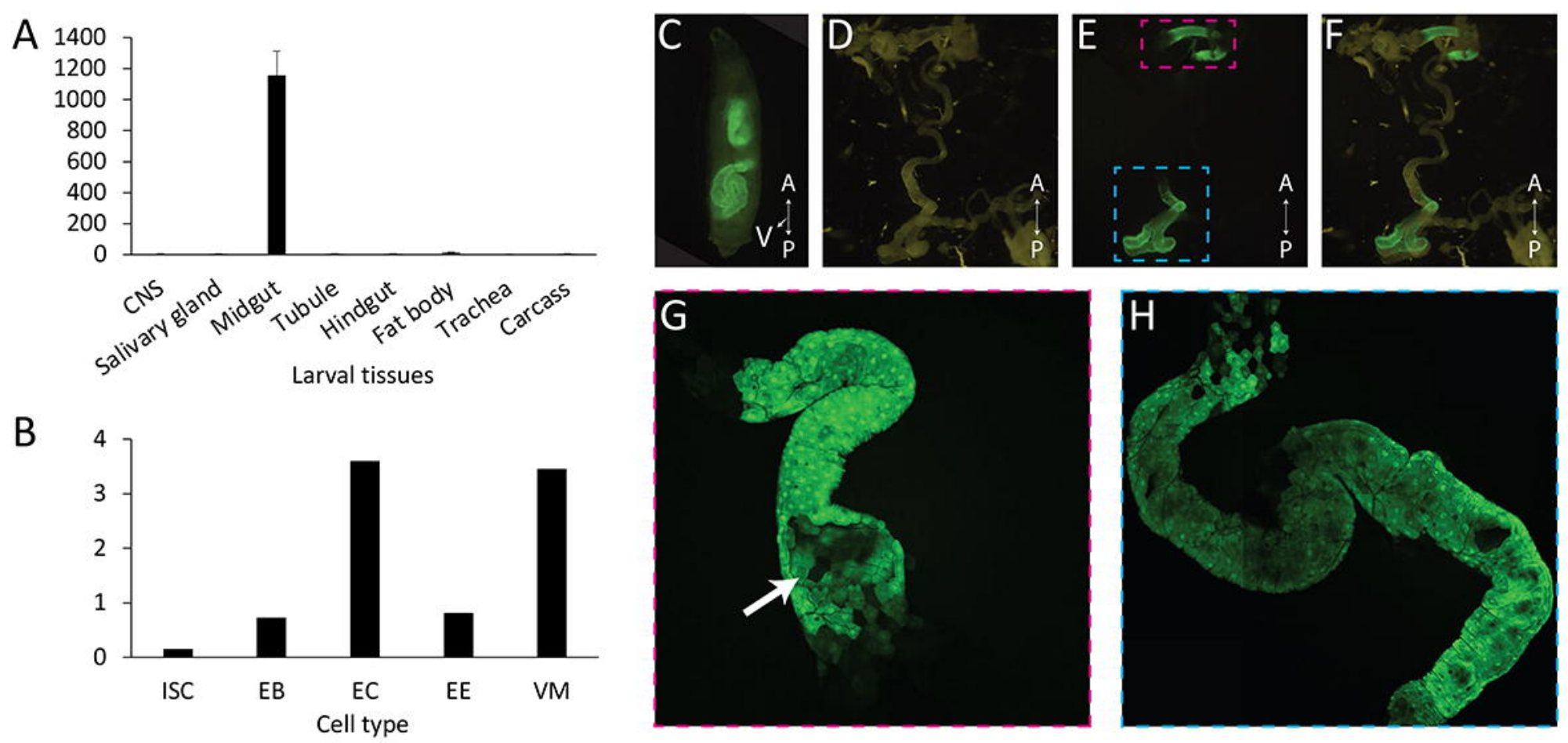
558 **Figure 5: Interaction network of SNP-containing genes. (A)** Expression levels of
559 *CG14205* across larval tissues. Data were extracted from the FlyAtlas database. **(B)**
560 Expression levels of *CG14205* across midgut cell types. ISC, Intestinal stem cells; EB,
561 Enteroblasts; EC, Enterocytes; EE, Enteroendocrine cells; VM, Visceral muscle. **(C)**
562 *CG14205* expression is restricted to an anterior and posterior regions of the larval
563 midgut. Image of an intact larva expressing GFP under the control of the *CG14205*
564 GAL4 line. **(D-F)** Image of 3rd instar dissected gut: (D) Visible light image, (E) GFP
565 image, (D) Overlay. **(G)** High resolution confocal image of *CG14205*-expressing
566 anterior region (pink dashed box in E). Arrow shows GFP expression in stereotypical
567 enterocytes. **(H)** High resolution confocal image of *CG14205*-expressing posterior
568 region (blue dashed box in E).


A
10 mm
Line 75
B
10 mm
Line 101
C
Fraction Aggregating
0.8
0.6
0.4
0.2
0
High
Low
93
75
21
101
195
149
DGRP Line


A
-log10 p-val
0
1
2
3
4
5
6
7
2L
2R
3L
3R
X
Chromosome
B
CG8187
Vha36-1
CG14502 (+), sbb (-)
CG43795 (+), Or59c (-)
Dnah3
mfr
CG33158 (+), dsx-c73a (-)
CG32206
rn
pros
Nlg4
cindr
CG45002
CG14205
-log10 p-val
5.0
5.5
6.0
6.5
7.0
MAF
0.1
0.2
0.3
0.4
0.5
2L
2R
3L
3R
X
r2
0
1


A
CG8187
Vha36-1
CG14502 (+)
sbb (-)
CG43795 (+)
Or59c (-)
Dnah3
mfr
CG33158 (+)
dsx-c73a (-)
CG32206
rn
pros
Nlg4
cindr
CG45002
Relative mRNA Expression
93 75 21 101 195 149
DGRP Line
Low
High
B
CG14205
Relative mRNA Expression
*
93 75 21 101 195 149
DGRP Line
D
CG14205
Exon 5
Intron 5
Line 93 AAAG GTACGA
Line 75 AAAG GTACGA
Line 21 AAAG GTACGA
Line 101 AAAG GTAGGA
Line 195 AAAG GTAGGA
Line 149 AAAG GTAGGA
X:19488026
C
-log10p-val
p = 0.05
CG8187
Vha36-1
CG14502 (+)
sbb(-)
CG43795 (+)
Or59c (-)
Dnah3
mfr
CG33158 (+)
dsx-c73a (-)
CG32206
rn
pros
Nlg4
cindr
CG45002
CG14205
+ strand
- strand
SNP


A
Fraction Aggregating
0.8
0.4
0
*
*
ns
*
*
Gene
CG8187
CG14502
Or59c
CG33158
CG32206
rn
elav-GAL4
UAS-RNAi
B
Fraction Aggregating
0.8
0.4
0
*
Control (GFP)
sbb
CG43795
Dnah3
mfr
Nlg4
C
elav^GAL4
0.8
0.4
0
Control (GFP)
CG14205
D
tub^GAL4
0.8
0.4
0
*
Control (GFP)
CG14205


A
1400
1200
1000
800
600
400
200
0
CNS
Salivary gland
Midgut
Tubule
Hindgut
Fat body
Trachea
Carcass
Larval tissues
B
4
3
2
1
0
ISC
EB
EC
EE
VM
Cell type
C
A
V
P
D
A
P
E
A
P
F
A
P
G
H