INVESTIGATION

New *Drosophila* Long-Term Memory Genes Revealed by Assessing Computational Function Prediction Methods

Balint Z. Kacsoh,*1 Stephen Barton,*1 Yuxiang Jiang,† Naihui Zhou,‡ Sean D. Mooney,§ Iddo Friedberg,† Predrag Radivojac,** Casey S. Greene,†† and Giovanni Bosco*,2

*Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, NH, 03755, USA
†Department of Computer Science, Indiana University, Bloomington, IN
‡Department of Veterinary Microbiology and Preventive Medicine, Iowa State University, Ames, Iowa 50011
§Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, PA, 19104

ORCID IDs: 0000-0001-9171-0611 (B.Z.K.); 0000-0001-7955-7677 (S.B.); 0000-0003-4657-925X (Y.J.); 0000-0003-2654-0833 (S.D.M.); 0000-0002-1789-8000 (I.F.); 0000-0002-6769-0793 (P.R.); 0000-0001-8713-9213 (C.S.G.); 0000-0002-8889-9895 (G.B.)

ABSTRACT

A major bottleneck to our understanding of the genetic and molecular foundation of life lies in the ability to assign function to a gene and, subsequently, a protein. Traditional molecular and genetic experiments can provide the most reliable forms of identification, but are generally low-throughput, making such discovery and assignment a daunting task. The bottleneck has led to an increasing role for computational approaches. The Critical Assessment of Functional Annotation (CAFA) effort seeks to measure the performance of computational methods. In CAFA3, we performed selected screens, including an effort focused on long-term memory. We used homology and previous CAFA predictions to identify 29 key *Drosophila* genes, which we tested via a long-term memory screen. We identify 11 novel genes that are involved in long-term memory formation and show a high level of connectivity with previously identified learning and memory genes. Our study provides first higher-order behavioral assay and organism screen used for CAFA assessments and revealed previously uncharacterized roles of multiple genes as possible regulators of neuronal plasticity at the boundary of information acquisition and memory formation.

KEYWORDS

*D. melanogaster*  
Parasitoid wasp  
Learning and memory  
Long-term memory  
Behavior  
Bioinformatics  
Gene function prediction  
Critical assessment

In the sequencing age, public data present the opportunity to the experimental biologist to address important biological questions that might otherwise remain hidden (Faith *et al.* 2007, e8; Yan *et al.* 2010, e12139; Chikina *et al.* 2009, e1000417; Chikina and Troyanskaya 2011, e1001074). However, the data abundance and availability can be a formidable project to tackle. Thus, we ask the question, how and to what extent can one utilize so much data to ask a very particular series of questions?

Data-mining algorithms have been designed and subsequently utilized to dissect large data sets. This utilization can present the discovery of novel and biologically relevant pathways and gene functions that might be overlooked due to the inability to pass detection threshold of other methodologies (Greene *et al.* 2009, 1289-1296; Greene *et al.* 2014, 1896-1900; Greene and Troyanskaya 2012, 95-100; Greene and Troyanskaya 2011, W368-74). By utilizing a bioinformatics-based approach, an accurate computer-generated list of gene targets can be generated to shed light on important biological process. This allows the biologist to rapidly conceive of, and subsequently test,
new hypotheses on new targets that would be otherwise overlooked. Evaluating such methods frequently relies on curated databases of gene functions derived from the literature. We sought to complement these evaluations with assays of the behavioral phenomenon: long-term memory formation.

The ability to perceive, process, respond to, and remember past, present, and potentially upcoming environmental challenges is essential for the survival of any organism. This ability is observed across the animal kingdom and remains a ripe, and understudied field of study. The field of memory and the accompanying neurobiology has arisen from the utilization of multiple approaches, ranging from behavior testing, to experimental perturbation that includes pharmacological, biochemical, and anatomical stressors in model and non-model systems (Dubnau et al. 2001, 476; Dubnau and Tully 1998, 407-444; Dubnau et al. 2001, 476-480; Margulies et al. 2005, R700-R713; Tully et al. 1994, 35-47; Tully 1987, 330-335). Genetic manipulations have provided key insight into behavioral plasticity that facilitates memory formation and retention. The use of genetics allows the researcher to understand the hereditary components that regulate neuronal plasticity. Utilization of genetic systems allows one to find clues illuminating the extent of evolutionary conservation among the different molecular mechanisms that govern learning and memory (Greenspan 1995, 747-750).

*Drosophila melanogaster* has been employed as a model system to understand the genetic and molecular basis of memory. Using associative (McGuire et al. 2005, 328-347; Tully et al. 1994, 35-47; Tully 1987, 330-335; Tully and Quinn 1985, 263-277) and non-associative paradigms (Das et al. 2011, E646-54; McCann et al. 2011, E655-62; Ramaswami et al. 2013, 727-736; Ramaswami 2014, 1216-1229), insight into the molecular mechanism governing the formation of long-term memories has been provided. However, the genes and proteins discovered in this manner might only be a small representation of the underlying learning and memory factors (i.e., genes, pathways, stimuli) that manifest in the insect brain. Therefore, it remains valuable to develop and utilize ways in which we may define potentially new roles for known genes and proteins. These novel elements that may be involved in the complex process of memory can be identified through an unbiased approach and subsequent mechanistic analysis.

There exists a strong need to have a means to understand the process of genes and gene function, including within the study of memory. The approach to understanding this process can be and has been expedited through the use of computational algorithms. These algorithms aim to automatically annotate and categorize genes to a given function. Genetic annotation via these algorithms can be based on multiple factors including the nucleic acid sequence and similarity to other known genetic elements (Wilson et al. 2000, 233-249; Cozzetto and Jones 2017, 55-67), evolutionary relationships (Pellegrini et al. 1999, 4285-4288; Engelhardt et al. 2005, e45), the gene expression patterns (Greene and Troyanskaya 2011, W368-74; Greene et al. 2014, 1896-1900; Greene et al. 2016, 557-567; Greene and Troyanskaya 2012, 95-100), the gene interaction partner(s) (Marcotte et al. 1999, 751-753; Vazquez et al. 2003, 697), and other features (Lan et al. 2013, S8; Sokolov et al. 2013, S10; Cozzetto et al. 2016, 31865). Given so many possible modes of prediction, there exists a need for a commonly accepted and utilized modality. The critical assessment of functional annotation (CAFA) was conducted to address this need (Jiang et al. 2016, 184; Radivojac et al. 2013, 221). CAFA has historically relied on curated databases of gene functions, but there remains a need to complement these literature-derived resources with biological assays (Greene et al. 2016, 557-567). This remains especially true in biological processes, such as learning and memory, where relatively few proteins have experimental annotations.

The inherent plasticity present in an organism allowing it to recall information can allow the individuals and/or groups to modify many facets of their behavior, including reproduction. Modulation of reproductive behavior, and anticipation of conditions that effect offspring survival, which may affect fitness, is necessary for the survival of an organism. For example, *Drosophila melanogaster* females will retain their eggs (oviposition depression) when environmental conditions are determined to be nutritionally unsuitable (Spradling et al. 1999, 135-177). Drosophila will also depress their oviposition rates during and following an encounter with an endoparasitoid wasp by inducing effector caspases in a stage specific manner in the ovary (Lefevre et al. 2012, 230-233; Lynch et al. 2016; Kacsoh et al. 2017; Kacsoh et al. 2018, e1007430; Kacsoh et al. 2015b, 10.7554/eLife.07423). These endoparasitoids regularly infect Drosophila larvae, upwards of 90% in nature (Driessen et al. 1989, 409-427; Fleury et al. 2004, 181-194; LaSalle 1993, 197-215). Egg depression following wasp-exposure is maintained and is dependent on long-term memory formation (Kacsoh et al. 2018, e1007430; Kacsoh et al. 2017; Kacsoh et al. 2015b, 10.7554/eLife.07423; Bozler et al. 2017, e1007054). It is important to note that adult Drosophila are under no threat from these wasps. This makes all behavioral changes in response to the wasps anticipatory as a means to benefit their offspring. These changes are mediated, in part, by the visual system (Kacsoh et al. 2018, e1007430; Kacsoh et al. 2015b, 10.7554/eLife.07423; Kacsoh et al. 2015a, 1143-1157; Kacsoh et al. 2013, 947-950). There has also been evidence to demonstrate the involvement of long-term memory genes and proteins that mediate the maintenance of this oviposition depression behavior. In particular, the role of these genes in the mushroom body (MB), the region of the fly brain implicated in learning and memory (Aso et al. 2009, 156-172; Aso et al. 2014, e04577; Claridge-Chang et al. 2009, 405-415; Masse et al. 2009, R700-R713; Schweerzel et al. 2003, 10495-10502), has been established to function in the maintenance of the egg depression behavior (Kacsoh et al. 2017; Kacsoh et al. 2015b, 10.7554/eLife.07423; Kacsoh et al. 2015a, 1143-1157; Kacsoh et al. 2018, e1007430; Kacsoh et al. 2015a, 1143-1157; Bozler et al. 2017, e1007054).

Given the current findings, we might only be scratching the surface of the alien world that is the foundation of neurobiology and the study of memory—not just in the Drosophila brain, but also in other organisms. These yet-to-be-identified factors might be necessary and/or sufficient for the processes of information acquisition, memory induction and maintenance, following the ability of memory recall. Given this extremely multifaceted process containing multiple layers of signaling, it remains valuable to identify and test novel genetic candidates. In order to find new genes and pathways, we must identify, create, implement, and validate new methodologies, which are independent of classical mutagenesis-based approaches.

In this study, we performed assays for CAFA3 designed to assess the performance of algorithms predicting novel genetic candidates that may be involved in long-term memory. We selected 29 genes to screen that we expected would be particularly useful for this evaluation. This study focuses on the biological findings resulting from the screen. Through the use of RNA interference (RNAi) of each gene by expression of an RNAi transgene in the MB, we identify 3 genes involved in perception of the stimulus (wasp exposure), and 12 genes involved in long-term memory formation. Using a functional relationship network analysis, we find that the 3 perception genes are enriched in developmental and response to stimuli processes, while our long-term memory candidates are highly connected to learning and memory genes, including ones previously identified to be involved in this particular behavioral paradigm, and that the accompanying network is enriched for the process of long-term memory.
MATERIALS AND METHODS

Insect Species/Strains

The *D. melanogaster* strain Canton-S (CS) was used as the wild-type strain, in addition to outcrosses. For crosses:

- *ft* RNAi, *kn1* RNAi, *sens* RNAi, *rut* RNAi, *shp* RNAi, *shf* RNAi, *sna* RNAi, *S-H17* RNAi, *GluR1* RNAi, *Tkr86C* RNAi, *wdr* RNAi, *Snap25* RNAi, *Dop161* RNAi, *Rnp48* RNAi, *Nrk* RNAi, *sdx* RNAi, *CG5815* RNAi, *Cgd87A* RNAi, *OctobetaR* RNAi, *CG17119* RNAi, *Capr* RNAi, *CG18812* RNAi, *nnb* RNAi, *Mob4* RNAi, *Oaz* RNAi, *Trpm* RNAi, *Dop1R2* RNAi, *p38* RNAi, *CG12744* RNAi, *CG5815* RNAi, and *Capr* RNAi were used.

These stocks were acquired from the Bloomington Drosophila Stock Center and the Vienna Drosophila Resource Center (Supplementary Table 1). Stock numbers respectively:

- 29566, 36664, 27287, 27035, 28924, 55867, 28679, 32471, 27521, 31884, 35654, 27306, 31765, 35457, 55184, 33412, 22203, 28716, 58179, 40823, 110272, 42777, 28888, 65236, 35715, 44503, 51423, 64846 and 67262 from Bloomington Drosophila Stock Center and 110272 and 22203 from the Vienna Drosophila Resource Center (Supplementary table 1).

Flies that were aged 4-6 days post-eclosion, grown on fresh Drosophila media, were utilized in all experiments. Flies were maintained at room temperature, approximately 20°C, with a 12:12 light:dark cycle at light intensity 16, with 30–55% humidity dependent on weather. We measured light intensity with a Sekonic L-308DC light meter using shutter speed 120, sensitivity at iso8000, and a 1/10 step measurement value (%). These are conditions used in previous studies (Kacsoh et al. 2018, e1007430; Kacsoh et al. 2017; Kacsoh et al. 2015b, 10,75545/ eLife.07423; Kacsoh et al. 2015a, 1143–1157).

We utilized the generalist Figitid larval endoparasitoid *Leptopilina heterotoma* (strain Lh14), which is known to infect a wide array of Drosophilids (Schlenke et al. 2007, e158; Kacsoh and Schlenke 2012, e34721; Kacsoh et al. 2014, 697-715) for all experiments (exposed condition). *L. heterotoma* strain Lh14 originated from a single female collected in Winters, California in 2002, and has been maintained in the lab ever since. To propagate wasp stocks, we used the D. melanogaster (strain Canton-S) in batches of 40 females and 15 males per vial (Genesse catalog number 32-116). Adult flies were disposed of, and not used to propagate the stock.

Flies aged 7 days post eclosion were used for exposed vials. Batches of 10 female, 5 male flies were placed into standard Drosophila vials (Genesse catalog number 32-116) of 20 females, 2 males. Three vials were prepared to host each line. Flies were removed from these treatment vials and eggs were counted from the vials for both days tested. All raw egg counts and corresponding p-values are provided in supplementary file 1.

Cross set-up

For RNAi experiments, male flies containing the RNAi-hairpins were crossed to virgin female flies containing the MB247 GAL4 and/or the MB8571 GAL4, which were kindly provided by Mani Ramaswami (University of Dublin, Ireland) and the Bloomington Drosophila Stock Center, stock number 48571, respectively. 10 female MB247 GAL4 virgin flies were mated to 5 males containing the selected hairpin.

Crosses were performed in standard Drosophila vials containing 5 mL standard Drosophila media (Genesse catalog number 32-116) supplemented with 10 granules of activated yeast. Crosses were kept at approximately 20°C, with a 12:12 light:dark cycle at light intensity 16, with 30–55% humidity dependent on weather. Crosses were moved to new vials every two days until egg production declined, at which point the adults were disposed of. Upon F1 eclosion, F1 flies were moved to a new vial and allowed to age to 4 days before use in experiments.

Wasp response and memory assay

F1 flies from above described crosses are CO2, anesthetized and sorted by genotype to ensure that both the GAL4 and hairpin are in the same background. Genotypic sorting was only required when UAS-RNAi lines contained a balancer chromosome which had to be sorted against. Flies are sorted into groups of 10 females and 5 males. Unexposed flies are placed into standard Drosophila vials (Genesse catalog number 32-116) containing 5 mL standard Drosophila media. Once unexposed vials are set up, wasps are anesthetized on the CO2 pad next to flies that will be used for exposed vials. Batches of 10 female, 5 male flies are placed into standard Drosophila vials (Genesse catalog number 32-116) with 20 female Lh14 wasps. Vials contain 5mL standard Drosophila media. Exposed flies were separated from wasps following anesthetization. These wasp-exposed flies were placed into new vials containing 5mL standard Drosophila media. Eggs were counted from the vials that flies were removed from, which corresponds to the acute response, 0-24-hour period. After an additional 24-hours in new vials, all flies were removed from all treatment vials and eggs were counted to measure the memory response (24-48-hour period).

All experimental treatments were run at 25°C with a 12:12 light:dark cycle at light intensity 16, using twelve replicates at 40% humidity unless otherwise noted. To avoid bias, gene categorization following CAFA analysis was withheld until completion of all experiments. Additionally, all vials were coded and scoring was blind as the individual counting eggs was not aware of treatments or genotypes and random subsets were counted by two individuals to ensure accuracy. Eggs were counted immediately after fly removal for both days tested. All raw egg counts and corresponding p-values are provided in supplementary file 1.

Immunofluorescence

Brains expressing MB247 GAL4 or MB8571 and CD8-GFP were dissected in a manner previously described (Kacsoh et al. 2015a, 1143-1157). Flies were placed in batches into standard vials (Genese catalog number 32-116) of 20 females, 2 males. Three vials were produced to produce three replicates to account for batch effects of fluorescence. We observed no batch effects. Brains that were prepared for immunofluorescence were fixed in 4% methanol-free formaldehyde in PBS with 0.001% Triton-X for approximately five minutes. The samples were then washed in PBS with 0.1% Triton-X. Samples were then blocked in a 4% NGS solution, prepared in a 0.1% Triton X solution. Brains were stained with nc82 (Hybridoma Bank at the University of Iowa, Antibody registry ID: AB 2314866) at 1:10 ratio of antibody:0.1% Triton X. Staining was performed overnight at 4°C. Following the overnight stain, brains were washed with 0.1% Triton X solution three times. Secondary Cy3 antibody (Jackson Immunoresearch) was then
placed on samples in a 1:200 ratio in 0.001% Triton-X. Following secondary staining, samples were washed in PBS with 0.1% Triton-X three times. This was followed by a 10-minute nuclear stain with 4', 6-diamidino-2-phenylindole (DAPI). Samples were then washed again a 0.1% Triton X solution three times and mounted in Vectashield (Vector Laboratories, Catalog number: H-1200) and immediately imaged.

**Imaging**

A Nikon A1R SI Confocal microscope was used for imaging GFP in the MB (Figure 2 C, Supplementary Figure 2-3). Image averaging of 16x during image capture was used for all images.

**IMP analysis**

Following all knockdown experiments, we input our acute response gene hits and memory response gene hits separately into IMP (http://imp.princeton.edu) (Figure 5, 6) (Wong et al. 2015, W128-33; Wong et al. 2012, W484-90). We did this analysis after all experiments were completed to further remove any bias. We utilized a stringent minimum gene connection of 0.85 confidence, indicated with blue lines, as well as a less stringent cutoff of 0.4, indicated with pink lines.

**Statistical analysis**

We performed statistical tests on exposed v unexposed conditions in Microsoft Excel. Our conditions utilized Welch’s two-tailed t-tests for all comparisons. P-values reported are included in supplemental file 1.

**Data Availability Statement**

All raw data are available in supplemental file 1. All strains used shown in supplemental table 1 and are available upon request. Supplemental material available at Figshare: https://doi.org/10.25387/g3.7364774.

**RESULTS**

**CAFA analysis reveals novel Gene targets**

We employed critical assessment of functional annotation (CAFA) analysis to provide new genetic factors in learning and memory. The CAFA challenge is organized to evaluate computational methods for protein function prediction through a community experiment (Jiang et al. 2016, 184; Radijovac et al. 2013, 221). Every three years, the organizers of the challenge release a large number of proteins whose functions are unknown or partially known. The predictors then use data and methodology of their choice to predict and submit predictions for these targets by the submission deadline set to occur four months after the target release. The submission is subsequently closed and the organizers collect the predictions. After a waiting period of 6-12 months, the organizers collect protein functions published after the submission deadline and evaluate the predictor performance on newly accumulated annotations. The organizational structure and further details of the experiment have been previously discussed by Friedberg and Radijovac (Friedberg and Radijovac 2017, 133-146).

The availability of the predictions accumulated during the first two CAFA challenges provided an opportunity to not only identify top-performing computational models, but also to feed predictions back into the experimental pipeline. We used these accumulated predictions from previous CAFA challenges to set up one of the CAFA3 experiments. The goal of this experiment was twofold: (1) to identify novel genes involved in learning and memory, and (2) to evaluate protein function prediction methods in CAFA3 according to their ability to identify learning and memory genes. This means that the set of genes that was ultimately selected for experimental evaluation had to contain both novel targets and also targets not involved in learning and memory. Furthermore, the selected genes would desirably be informative with respect to the ability of the organizers to accurately rank computational methods as well as their potential to improve the prediction of this biological function. In this study, we describe the pipeline for identifying memory genes in *Drosophila*. The evaluation of computational methods will be described in a separate paper along with all other CAFA3 challenges.

Specifically, to identify novel genes involved in learning and memory and evaluate prediction tools, we turned to a computational approach using the principles of the machine learning technique known as active learning (Sverchkov and Craven 2017, e1005466). Twenty-nine candidate genes for screening were selected based on an ensemble of top performing predictors in the CAFA2 challenge (Jiang et al. 2016, 184).

That is, we collected prediction scores from top ten computational models that submitted predictions on 3195 Drosophila target genes (Figure 1 A). Their raw prediction scores on each gene for the biological process “long-term memory” (GO:0007616) were first converted into quartiles (Q) with five possible outcomes from 0 indicating a negative prediction up to 1 indicating a positive prediction with a step size 0.25. We then averaged the quartiles from these ten models for each gene so as to group them into three categories: (i) “likely positive” with average Q greater than or equal to 0.5 and covered by at least 7 models; (ii) “likely negative” with average Q of 0.00 and agreed by at least 7 models; and (iii) “variable” with average Q between 0.1 and 0.25 and standard deviation greater than 90% quantile (i.e., top 10% variable) and covered by at least 7 models (Figure 1 B). After excluding essential genes and those already associated with the GO:0007616 term in FlyBase, we randomly selected subsets of genes from the positive class (Mob4, rut, CG17119, GluRIA, Snap25), the negative class (Rnp4F, sens, sna, Capr, CG12744, CG5815, Oaz, Trpm, wdn) and the variable class (5-HT7, Cad87A, Dopr1R, Dopr2R, mmb, NrK, Octbeta1R, CG18812, fl1, knurl, p38c, sbr, sdk, shf, Tkr86C).

**flies depress oviposition during and following wasp exposure**

In order to elucidate the roles of these genes in memory, we turned to an assay using the Drosophila-endoparasitoid-wasp system. Our assay utilizes the behavior of *D. melanogaster* following exposure to parasitoid wasps. One of the behaviors during and following wasp encounter by female flies is a depression their egg-laying (oviposition) rate. This oviposition depression occurs during wasp exposure and following wasp removal. The oviposition depression following wasp exposure is maintained for an extended period following wasp removal. Of unique interest is that this behavior is dependent on learning and memory genes (Kacsoh et al. 2018, e1007430; Kacsoh et al. 2015b, 10.7554/eLife.07423; Kacsoh et al. 2017; Lefèvre et al. 2012, 230-233; Lynch et al. 2016). We performed an egg depression assay similar to previously published protocols (see methods). Briefly, flies are CO2 anesthetized and selected by genotype, meaning that both the GAL4 is in conjunction with the hairpin, or the wild type chromosome copy as a result of the outcross, are contained in the same animal. Flies are then sorted into groups of 10 females and 5 males. Unexposed flies are placed into standard Drosophila vials containing 5 mL standard Drosophila media. For exposed groups, batches of 10 female, 5 male flies are placed into standard Drosophila vials with 20 female wasps. Following a 24-hour exposure period (acute response), flies from both unexposed and exposed conditions are briefly, but independently anesthetized. Unexposed flies are placed into new vials. Exposed flies are separated from wasps following anesthetization. These wasp-exposed flies are
then placed into new vials. Eggs are counted from the vials that flies were removed from, which corresponds to the acute response, 0-24-hour period. After an additional 24-hours in new vials, all flies are removed from all treatment vials and eggs are counted to measure the memory response (24-48-hour period) (Figure 2 A). Control, unexposed treatments are performed simultaneously with exposed treatments to ensure appropriate comparisons. We performed all experiments at 25°C with a 12:12 light:dark cycle at light intensity 160 μW/cm², using twelve replicates at 40% humidity. To avoid bias, prediction scores for the genes being tested remained unknown until the completion of the study. A random number was assigned to each stock and only relevant genotypic information—i.e., the presence of a balancer in the UAS line to sort against—where we collect top-performing methods for Drosophila genes on GO:0007616. (B) indicates the process where we convert raw prediction scores to quantile (Q) and accumulate for each protein. We subsequently group these genes into three categories, where we picked 29 candidates for experimental verification.

As previously shown, we find that wild type Canton-S flies depress oviposition during the acute response period (in the presence of wasps) and maintain the depression during the memory response period (following wasp removal) (Figure 2 D) (Lefevre et al. 2012, 230-233; Lynch et al. 2016; Kacsoh et al. 2017). We test an additional control line, where we expressed RNAi-hairpins targeting the white gene in the MB, using both driver lines. This genetic combination yields wild-type behavior, which demonstrates that induction of an RNAi-hairpin construct alone does not induce an RNAi-hairpin mediated white knockdown in the MB, reconfirms that the MB247 driver line alone (outcrossed to Canton-S) has wild-type memory retention, in addition to demonstrating that the secondary MB driver line (outcrossed to Canton-S) has wild-type memory retention (Supplementary Figure 1 A-B) (Kacsoh et al. 2015b, 10.7554/eLife.07423; Kacsoh et al. 2017). We test an additional control line, where we expressed an RNAi-hairpin targeting the white gene in the MB, using both driver lines. This genetic combination yields wild-type behavior, which demonstrates that induction of an RNAi-hairpin construct alone does not induce a deficient acute response or memory formation (Figure 2 E, Supplementary Figure 1 C). Outcrossed driver lines and RNAi-hairpin mediated white knockdown were run simultaneously to ensure appropriate conclusions about background lines and the effect of RNAi.

Using this system as a means to probe stimulus response and memory, we utilized the CAFA analysis to provide a series of potential long-term memory genes (Table 1). We then used the GAL4/UAS system to drive expression of an RNAi-hairpin targeting each of the predicted gene mRNAs in conjunction with the MB driver. Our assay allows us to measure the role of each of these genes for the acute response and memory response. This RNAi based screen is designed to test the efficacy of computational prediction. Secondary screens can be performed where alternative RNAi-hairpins may be tested and/or RNAi validation may be performed.
Figure 2 Wild-type flies depress oviposition following wasp exposure. Experimental design for oviposition depression assay (A) indicates a 0-24-hour period while wasps are present (acute exposure) and a 24-48-hour period following wasp removal (memory). In order to screen for long term memory genes, we drive expression of various RNAi-hairpin constructs in the fly mushroom body (MB) (B). A merged confocal image of the MB247 driver line expressing a CD8-GFP in green, DAPI in teal, and nc82 in magenta is shown in (C). Using this assay, we present the percentage of eggs laid by exposed flies normalized to eggs laid by unexposed flies. Wild-type flies exposed to wasps lay fewer eggs than unexposed flies in the presence of and following wasp removal (D). Flies expressing a white RNAi in the MB using the MB247 driver also show wild-type behavior, indicating that the RNAi machinery does not perturb the acute response nor the memory response (E). Error bars represent standard error (n = 12 biological replicates) (∗P < 0.05).
Table 1 Bioinformatic analysis reveals a series of genes that may be involved in long-term memory formation

| Gene list                  | acute response | memory response |
|----------------------------|----------------|-----------------|
| CS (WT control)            | Y              | Y               |
| RNAi of White, MB247       | Y              | Y               |
| MB247/+                    | Y              | Y               |
| RNAi of White, MB48571     | Y              | Y               |
| MB48571/+                  | Y              | Y               |
| FBgn0001075 ft             | N              | N               |
| FBgn0001323 knrl           | N              | N               |
| FBgn0002573 sens           | Y              | Y               |
| FBgn0003301 rut            | Y              | N               |
| FBgn0003321 sbr            | Y              | Y               |
| FBgn0003390 shf            | Y              | Y               |
| FBgn0003448 sna            | Y              | N               |
| FBgn00045735 5-HT7         | Y              | N               |
| FBgn0004619 GluR1A         | Y              | Y               |
| FBgn0004841 TkR86C         | N              | N               |
| FBgn0005642 whn            | Y              | Y               |
| FBgn0011288 Snap25         | Y              | Y               |
| FBgn0014024 Rnp4F          | Y              | N               |
| FBgn0020391 Nrk            | N              | N               |
| FBgn0037963 Cad87A         | Y              | Y               |
| FBgn0011582 Dop1R1         | Y              | Y               |
| FBgn0021764 sdk            | Y              | Y               |
| FBgn0038980 Octbeta1R      | N              | N               |
| FBgn0039045 CG17119        | Y              | Y               |
| FBgn0042135 CG18812        | Y              | Y               |
| FBgn0259483 Mob4           | Y              | N               |
| FBgn0259168 mnb            | Y              | Y               |
| FBgn0261613 Oaz            | Y              | Y               |
| FBgn0265194 Trpm           | Y              | Y               |
| FBgn0266137 Dop1R2         | Y              | Y               |
| FBgn0267339 p38c           | Y              | Y               |
| FBgn0033459 CG12744        | Y              | N               |
| FBgn0027574 CG5815         | Y              | Y               |
| FBgn0042134 capr           | Y              | Y               |

Genes tested and results are shown. A binary Yes (Y) / No (N) output is shown for each gene with respect to the presence of an acute response and a memory response. Both an acute and memory phenotype are listed if the egg count is significantly different from unexposed. Genes that perturb the acute response might be involved in memory formation, but their function in memory cannot be teased apart in the manner tested.

Three CAFA-selected genes affect in the acute response

We tested a total of 29 unique RNAi hairpins against 29 unique mRNA species in the MB and found 3 genes that, when knocked down via the expression of an RNAi transgene, elicit no acute response. These genes are fat (ft), knirps-like (knrl), and tachykinin-like receptor at 86C (TkR86C) (Figure 3, Table 1). All three of these genes fall under the variable class in the CAFA analysis.

Fat is a tumor suppressor gene that encodes a large cadherine transmembrane protein. Additionally, it functions as a receptor in the Hippo signaling pathway and the Dachous-Fat planar cell polarity pathway (Matakatu and Blair 2006, 2315-2324; Matakatu and Blair 2004, 3785-3794). It has GO annotation enrichments in cell cycle, development, and response to stimuli. Interestingly, ft has been implicated in neurodegenerative disorders that are polyglutamine (ploQ) diseases, where a functional Hippo/Fat pathway is protective (Napoletano et al. 2011, 945-958). The acute response may have been perturbed due to early onset neurodegeneration as a result of the ft knockdown (Figure 3 A). Interestingly, exposed flies show a significant increase in oviposition rate when in the presence of wasps, which is lost during the memory response. Given this observation, we tested ftRNAi in conjunction with the secondary MB driver, in addition to outcrossing the UAS-ftRNAi line (ftRNAi/+). We find this combination recapitulates the finding utilizing the MB247 driver where the acute response is ablated and exposed flies lay significantly more eggs in the presence of wasps (Supplementary Figure 4 A). The increased oviposition rate of wasp-exposed flies presents an interesting biological role for ft, possibly in response to stimulus regulation. Additionally, ft could be a negative regulator of nutrient sensing, where knockdown promotes ovi-position. Future experiments utilizing overexpression of ft can further elucidate its role in this paradigm. The outcrossed driver line behaves as wild-type, suggesting perturbation of the acute response is due to knockdown of ft and not genetic background effects (Supplementary Figure 4 C).

Knirps-like is a nuclear hormone receptor with a C4 zinc finger motif, lacking a ligand-binding domain. It is a target of hedgehog, wingless, and the Notch signaling pathways, with strong GO term enrichment in development (Chen et al. 1998, 4959-4968; Chen 1999; Curators 2004) (Figure 3 B). Interestingly, we find again a knockdown where exposed flies show a significant increase in oviposition rate when in the presence of wasps, which remains present during the memory response. Given this observation, we tested knrlRNAi in conjunction with the secondary MB driver, in addition to outcrossing the UAS-knrlRNAi line (knrlRNAi/+). We find this combination recapitulates the finding where the acute response is ablated. However, we do not observe the increased oviposition rate in the presence of wasps, suggesting that the significantly increased oviposition rate may be a result of genetic background (Supplementary Figure 4 B). The outcrossed driver line behaves as wild-type, suggesting perturbation of the acute response is due to knockdown of knrl and not genetic background effects (Supplementary Figure 4 D).

TKR86C is a G-protein coupled receptor and is rhodopsin-like. Its molecular function is described as neuropeptide receptor activity and is involved in many biological processes such as neuropeptide signaling and inter-male aggression (Jiang et al. 2013, E3526-34; Johnson et al. 2003, 52172-52178; Asahina et al. 2014, 221-235; Poels et al. 2009, 545-556). The data suggest that TKR86C may be involved in the acute response, perhaps by way of neuropeptide signaling (Figure 3 C, Table 1).

Twelve CAFA-selected genes affect in the memory response

Of the total 29 RNAi-hairpins against unique mRNA species in the MB tested, we identified a total of 12 unique mRNA species that are required in the MB for long-term memory formation: senseless (sens), rutabaga (rut), snail (sna), 5-hydroxytryptamine (serotonin) receptor 7 (5-HT7), RNA-binding protein 4F (Rnp4F), Neuro-specific receptor kinase (Nrk), Cadherin 87A (Cad87A), Dopamine 1-like receptor 1 (Dop1R1), Octopamine B1 receptor (Octb1R), MOB kinase activator 4 (Mob4), Minibrain (mbn), and Dopamine-like receptor 2 (Dop2R2) (Figure 4, Table 1).

With respect to the CAFA classification, these genes fall under multiple categories. From the positive class, we find Mob4 and rut. From the negative class, Rnp4F, sens, and sna were identified to be involved in memory formation. Finally, from the variable class, we identify 5-HT7, Cad87a, Dop1R1, Dop2R2, mnb, Nrk, and Octb1R. Previous work utilizing this assay has implicated the MB in long-term memory formation (Kacsoh et al. 2015b, 10.7554/eLife.07423), and given our results in this study, we have now identified additional gene products that also modulate the formation of this long-term memory.

We identify 12 genes involved in long-term memory formation, 11 of which are novel members of long-term memory formation using this
paradigm. We define a gene to be a candidate for long-term memory formation if there is a statistically significant egg depression during the acute response, but the depression is no longer statistically significant following wasp removal. 

Sens is a transcription factor that stimulates expression of pro-neural genes by regulating the differentiation of the peripheral nervous system and the R8 photoreceptor (Figure 4 A). It is GO term enriched in development and in responses to stimuli (Frankfort and Mardon 2004, 563-570; Salzberg et al. 1994, 269-288; Morey et al. 2008, 795; Mishra et al. 2013, e1004027). Rut encodes a membrane bound calcium2+/calmodulin activated adenylyl cyclase that is responsible for the synthesis of cAMP (Levin et al. 1992, 479-489). It is GO term enriched in behavior, responses to stimuli, and nervous system processes. Rut has been implicated in long-term memory formation before in associative learning (McGuire et al. 2003, 1765-1768; Zars et al. 2000, 18-31; Dubnau et al. 2001, 476-480; Alonso et al. 2003, 653-657; Margulies et al. 2005, R700-R713; Qin and Dubnau 2010, 203-212), as well as in non-associative learning using the fly-wasp assay (Kacsoh et al. 2015a, 1143-1157; Kacsoh et al. 2015b, 10.7554/eLife.07423), thus further validating our experimental protocol and results (Figure 4 B). Sna is a transcription factor that contributes to embryonic mesoderm development and nervous system development (Ashraf et al. 1999, 6426-6438; Mathew et al. 2011, 4978-4993). It is GO term enriched in development (Figure 4 C). 5-HT7 is a serotonin G-protein-coupled-receptor that binds and transmits the signal from the neurotransmitter 5-HT (serotonin) (Gaudet et al. 2010; Gasque et al. 2013, srep02120) and has been shown to be involved in olfactory learning and memory (Becnel et al. 2011, c02800; Luo et al. 2012, 471-484). It is GO term enriched in behavior and responses to stimuli (Figure 4 D). Rnp4F is predicted to be involved in mRNA binding, but little has been done to elucidate its function (Curators 2004; Lasko 2000, F51-6) (Figure 4 E). Nrk is a receptor tyrosine kinase that is expressed during neuronal development and regulates axon path finding (Kucherenko et al. 2011, 228-242; Oishi et al. 1997, 11916-11923; Marrone et al. 2011, 1). It is GO term enriched in development and response to stimulus (Figure 4 F). Cad87A is currently known to be cadherin-like, but its function and biological processes are unknown (Curators 2004) (Figure 4 G). Dop1R1 is known to be involved in dopamine neurotransmitter activity and visual learning and memory, not unlike the assay we utilized (Gotzes et al. 1994, 131-141; Seugnet et al. 2008, 1110-1117; Brody and Cravchik 2000, F83-8). Its GO term enriched in nervous system processes, behavior,
Figure 4 Expression of several genes in the mushroom body is necessary for long-term memory formation. Percentage of eggs laid by exposed flies normalized to eggs laid by unexposed flies is shown. Flies exposed to wasps with various RNAi-hairpins expressed in the MB using the MB247 driver show no memory response for sensRNAi (A), rutRNAi (B), snRNAi (C), 5-HT7RNAi (D), Rnp4frRNAi (E), NkrRNAi (F), Cad87aRNAi (G), Dop1R1RNAi (H), Oct1R1RNAi (I), Mob4RNAi (J), mnbRNAi (K), and Dop1R2RNAi (L). Error bars represent standard error (n = 12 biological replicates) (*P < 0.05).
responses to stimuli, and signaling. Interestingly, mutations in the human ortholog have been correlated with attention deficit disorder (Lowe et al. 2004, 348-356) (Figure 4 H). OctBR1 is known to be a G-protein-coupled-receptor and is rhodopsin-like. It has been shown to have a role in the negative regulation of synaptic growth at the neuro-muscular-junction (Balfanz et al. 2005, 440-451; Brody and Cravchik 2000, F83-8; Koon and Budnik 2012, 6312-6322). It is GO term enriched in responses to stimuli and signaling (Figure 4 I). Mob4 is involved in the negative regulation of synaptic growth at neuromuscular junctions, neuron projection morphogenesis, axodendritic transport, and negative regulation of the hippo pathway (He et al. 2005, 4139-4152; Schulte et al. 2010, 5189-5203; Zheng et al. 2017, 3612-3623). It is GO term enriched in cell cycle, development, responses to stimuli, and signaling (Figure 4 J). Mnb is a serine/threonine protein kinase and interacts with multiple signaling pathways. It is involved in multiple facets of behavior, and mutants have deficits in behaviors including visual learning and memory. The mutant has a morphologically smaller brain when compared to wild-type (Degoutin et al. 2013, 1176; Helfrich 1986, 321-343; Shahik et al. 2016, 3195-3205; Tejedor et al. 1995, 287-301; Sokolowski 2001, 879). It is GO term enriched in development, nervous system process, behavior, responses to stimuli, and signaling. Human orthologs exist of this gene (DYRK1A) and manifests as mental retardation (Moller et al. 2008, 1165-1170) (Figure 4 K). Dop1R2 belongs to the g-protein-coupled-receptor family and is implicated in adrenergic receptor activity, dopamine neurotransmitter activity, positive regulation of oxidative stress-induced neuron death, and cell signaling. These phenotypes are generally localized to the fan-shaped body of the brain, which is the visual learning and memory center of the fly brain that projects to the MB (Gaudet et al. 2010; Brody and Cravchik 2000, F83-8; Han et al. 1996, 1127-1135; Radford et al. 2002, 38810-38817; Cohn et al. 2015, 1742-1755; Pimentel et al. 2016, 333). This gene is GO term enriched in behavior, responses to stimuli, and signaling (Figure 4 L).

We selected two additional lines for additional testing with our secondary MB driver—Dop1R1RNAi and mnbRNAi. These two lines were selected due to identified human orthologs involved in autism and mental retardation. We again observe wild-type acute responses in knockdown of Dop1R1 and mnb using the secondary MB driver (Supplementary Figure 5 A-B). We again find perturbation of the memory response, where neither knockdown demonstrates memory of the wasp response (Supplementary Figure 5 A-B). To rule out the effects of genetic backgrounds, we independently tested outcrossed UAS lines (Dop1R1RNAi/+ and mnbRNAi/+), where we find wild type acute and memory responses (Supplementary Figure 5 C-D). These observations suggest that, indeed, identification of these two genes by CAFA is correct in prediction a role in learning and memory. Collectively, these results demonstrate that our assay can be utilized to model human diseases involved in cognition. Future work may utilize knockdown of these genes in combination with drug or genetic treatments that might rescue observed deficits.

An additional 14 lines were tested and demonstrate wild-type, statistically significant, acute and memory responses during and following wasp exposure (Supplementary Figure 6, Table 1). These genes fall into multiple CAFA categories: positive class (CG17119, GlurIA, Snap25), negative class (Cpr, CG12744, CG5815, Osz, TrpM, wd), and the variable class (CG18818, p38c, shb, sdi, shj). These results further demonstrate that knockdown of any gene does not result in a perturbation—knockdown where we observe a perturbation of wild-type behavior does implicate said gene/gene product in the acute or memory process (Figures 3-4).

**Gene network analysis reveals a high degree of connectivity**

In order to further examine both the acute response genes and the memory genes identified from our screen, we employed the integrative multi-species prediction (IMP) webserver to determine the extent to which our hits are connected with either other response to stimulus genes or to known learning and memory genes (Wong et al. 2015, W128-33; Wong et al. 2012, W484-90). IMP is a user-friendly program that provides the integration of gene-pathway annotations from a selected organism (in our case, D. melanogaster), and subsequently maps functional analogs by providing a user-friendly visual output. IMP has been utilized to predict gene-process predictions that were subsequently experimentally validated (Chikina et al. 2009, e1000417; Chikina and Troyanskaya 2011, e1001074; Kacsoh et al. 2017).

We are utilizing IMP in a twofold manner—for both hypothesis validation and hypothesis generation. We perform hypothesis validation using IMP by identification of genes in the presented networks that have been previously implicated in the acute or memory phenotype. We are generating hypotheses using IMP through the generated networks by presentation of genes within the network that have not previously been identified in learning and memory paradigms.

We first queried the network with our acute response implicated genes. We set the network filter to only show interactions whose minimum relationship confidence was 0.85, which is a stringent threshold (Figure 5 A). We find that ft and knrl are connected in a large gene network, while Tkr86C is only connected to one partner, ft(1)G0020. The genes we identified have independently associated GO terms that are predicted with a 0.85 confidence score (Figure 5 B). Interestingly, ft is associated with R3/R4 cell differentiation, raising the possibility of perturbed vision in these animals accounting for the lack of the acute response. We find that the network is enriched for the biological processes of cell fate specification (34.6%, p-value: 1.93e-8), cell fate commitment involved in pattern formation (19.2%, p-value: 9.21e-5), skin development (19.2%, p-value: 6.33e-4), and neuroblast differentiation (15.4%, p-value: 6.29e-4).

We next queried the network with our memory response implicated genes. We set the network filter to only show interactions whose minimum relationship confidence was 0.85 (shown in blue), followed by a filter of 0.4 (shown in pink), in order to highlight strong and weak, but relevant, interactions (Figure 6 A). We find that almost every gene tested is highly interconnected. Given this level of interconnectivity with previously identified memory genes, the network suggests that we have identified potential bona fide learning and memory genes. Interestingly, hits are connected, both directly and indirectly, to known learning and memory genes. This finding is particularly important to note given that the predicted interactors of the hits identified have been implicated in learning and memory assays utilizing the fly-wasp system. These genes are FMR1, dunce, amnesiac, alcohol dehydrogenase factor 1, and grunge, each of which has been shown to have a wild type acute response, but perturbed memory response using the fly-wasp memory paradigm (Kacsoh et al. 2015b, 10.7554/eLife.07423; Kacsoh et al. 2015a, 1143-1157; Kacsoh et al. 2017; Kacsoh et al. 2013, 947-950). Therefore, the IMP analysis is further validation of the hits identified from the CAFA analysis. The GO terms associated with the genes in the network are very heavily connected to long-term memory, synaptic functions, and behavior (Figure 6 B). The network is enriched for the biological processes of associative learning (19.4%, p-value: 3.14e-5), learning (19.4%, p-value: 5.69e-5), positive regulation of growth (19.4%, p-value: 6.21e-5), and olfactory learning (16.7%, p-value: 1.27e-4). Only one of the hits, Nrk, was found to not be connected to the network.
Figure 5 IMP analysis of genes that yielded an acute response defect reveals connectivity and developmental functions. A visual representation of a gene network utilizing our positive acute response hits (ft, knrl, TkR86C) reveals a high level of connectivity between identified genes (A). Lines indicate a minimum relationship confidence of 0.85. GO term enrichment for acute response candidates reveal development specific processes (B).
Figure 6 IMP analysis of genes that yielded a memory defect reveals a high level of connectivity in addition to neuro-specific functions. A visual representation of a gene network utilizing our positive long-term memory hits (sens, rut, sna, 5-HT7, Rnp4F, Nrk, Cad87A, Dop161, Octbeta1R, Mob4, mnb, Dop1R2) reveals a high level of connectivity between some of the selected genes (A). Blue lines indicate a minimum relationship confidence of 0.8, while pink lines indicate a relationship confidence of 0.4. GO term enrichment scores for long-term memory candidates reveal neuro specific processes (B).
However, a connection was observed with multiple genes following a lowering of the relationship confidence score, indicating that \textit{Nrk} may have some interaction with the network.

**DISCUSSION**

In this study, we identify novel long-term memory genes as part of the CAFA3 experiment. We analyzed hits for the long-term memory process in Drosophila using a functional network analysis tool, IMP, where we validate our hits based on connectivity to previously identified genes and where we generate new genes for testing. We identify 3 genes involved in the acute response to wasps, and 12 genes implicated in long-term memory. To our knowledge, our acute response hits and 9 of our genes implicated in long-term memory have not been previously implicated to function in Drosophila memory. Of the remaining 3 genes, 2 have not been shown to be involved in this form of Drosophila memory.

Past work has demonstrated the need for an intact visual system in order to elicit a fly behavioral response in the presence of the wasp (Kacsoh et al. 2017; Kacsoh et al. 2015a, 1143-1157; Kacsoh et al. 2015b, 10.7554/eLife.07423; Kacsoh et al. 2013, 947-950; Kacsoh et al. 2018, e1007430). In this study, we provide the opportunity of direct interaction between the flies and wasps, permitting multi-sensory information acquisition. Using this setup, we identify 3 genes that are necessary in the MB to elicit the oviposition behavior. Thus, it becomes possible that these genes may affect certain sensory components of the nervous system, in addition to altering neuronal excitability of the MB. For example, the excitability threshold in the MB could be raised or lowered with genetic knockdown, changing the responses observed in our behavioral paradigm. This suggests that the genes might not be involved directly in the acute response or memory, but rather, the gene is altering neuronal properties. A wild-type fly might achieve maximal excitation and form a stable memory from the wasp interaction, but flies deficient in the 3 acute response genes identified are trained to a lesser degree resulting in an unstable ability to process the information. Following wasp removal, this instability might manifest as a lack of a memory formed. Future experiments that utilize isolated sensory stimuli, especially visual cues given the involvement of photoreceptor cell determination, would be able to examine these hypotheses.

We highlight that the stimulus (acute) response is not affected in flies lacking memory, as the day 0 (acute response) data are indistinguishable from wild-type levels. The presence of the acute response in these flies suggests that wasp identifying sensory systems are intact. Thus, RNAi of these genes in the MB gives us F1 animals that can respond to wasps as wild-type, but forget the exposure following wasp removal. Therefore, the genes that we define as a long-term memory hit are not required for the detection the input of “wasp”, but instead are required to make the initial memory. These results suggest that the wiring and development of the fly brain during knock-down of these select genes remains intact and produces a functioning brain that can respond to the initial stimulus as the wild-type controls (day 0). We observe phenotypic differences on day 1 (memory period), where we find the rapid extinction of memory, or the lack of memory formation. These two results are not possible to delineate with the current experimental setup, but remains as a future question. Wild type flies demonstrate a maintained oviposition response, demonstrating memory retention following the wasp exposure (Kacsoh et al. 2017; Kacsoh et al. 2015b, 10.7554/eLife.07423).

It remains a possibility that the genes presented as hits are not only involved in either the acute response or memory formation, but could be participating in development. The MB driver (MB247 or MB88571) in conjunction with each of the RNAi transgenes that presented either an acute response defect or a memory retention phenotype could be the result of the constitutive expression of the RNAi-hairpin throughout development and might have perturbed the development, whether morphological or excitatory element of the neuron, yielding the alterations in behavior. This possibility suggests that knockdown phenotypes have the potential to reveal undetected changes that preclude proper adult MB functions. This possibility does not preclude the potential hits we uncovered, nor the efficacy of the prediction tool, but instead provides an important avenue for future experiments.

**Figure 7** Proposed model for genes identified in the acute and memory response. Model for acute response and learning/memory.

First, a female fly sees a wasp. Second, acute response genes are activated. Third, effector caspases are activated. Fourth, oviposition depression is observed. Fifth, we see activation of memory genes. Sixth, we see a sustained reduction in oviposition. We hypothesize that some genes are involved in either development of the MB and/or processing in the MB during the acute response, while other genes are activated later and are involved in long-term memory formation. Following memory consolidation, these genes may be further expressed while the initial acute response genes might be dispensable.
Additionally, it remains a possibility that there might be variation in the efficacy of the RNAi lines utilized in this study. Further testing will need to be performed where validation of the RNAi lines will be undertaken or additional hairpins will be tested to ensure observed results are indeed biologically significant. It is also important to note that UAS-RNAi lines may independently have non-specific effects and that could contribute to perturbed behavior. We outcrossed 4 UAS-RNAi lines (Supplementary Figures 4-5 C-D) where we observed wild type behavior. However, the possibility remains that a UAS-RNAi line may independently contribute to observed deficiencies.

Finally, we graded our behavioral responses, the acute and memory responses, on a binary scale (yes/no) determined by significance testing when compared to control. This type of comparison does not account for subtle changes that could be contributed by these genes, such as an attenuated initial or learned response, or a heightened initial or learned response. To elucidate these different alternatives, additional testing will be required, including utilization of multiple RNAi transgenes targeting a single gene in conjunction with additional MB driver lines to examine the gene doses and behavioral effects.

Given our data, we propose a model where the initial wasp stimulus is partially mediated by genetic factors that could mediate development and/or excitability of the wasp response circuit. These factors, including other yet to be discovered members, may mediate oviposition depression, which results in the activation of effector caspases (Kacsoh et al. 2018, e1007430; Kacsoh et al. 2015b, 10.7554/elife.07423). The maintenance of this physiological change, and formation of a long-term memory, may then be mediated by a second group of genes that perpetuate the signal so that the flies continue to lay fewer eggs even after removal of the wasp threat (Figure 7).

Collectively, our data highlights the importance of incorporating bioinformatics-based approaches as a means to identify new genetic factors in phenotypes of interest, in addition to evaluating prediction tools. Some of the genes that we identified have human disease orthologs (Dop1R1 and mnb). We are provided with the opportunity to study human diseases in a genetic model system and we are able to show that knockdown of these orthologs has a relevant phenotype with respect to memory. Gene function annotations, which form the primary means of analysis in CAFA experiments, tend to be biased toward genes with large effect (Hibbs et al. 2009, e1000322; Hess et al. 2009, e1000407). Pairing these evaluations with targeted experimental assays might elucidate novel regulators and pathways that would have gone otherwise undetected due to detection threshold. These novel elements could have biologically important implications even though they only undergo smaller alterations as a function of a stimulus (Zovkic et al. 2013, 61–74). Such changes can be elusive to identify otherwise, yet, could be involved in important biological processes, such as neuronal plasticity and associated excitation. Our study provides a novel role for multiple genes that have functional connections to known and novel pathways, as potential regulators of either a stimulus response, or the maintenance of a response. These findings and pathways exist at the intersection of memory acquisition and memory retention.

ACKNOWLEDGMENTS

We thank Mani Ramaswami, FlyBase, the Vienna Drosophila Resource Center, and the Bloomington Drosophila Stock Center, for stocks. We thank the Developmental Studies Hybridoma Bank at the University of Iowa for reagents. We acknowledge grants from Geisel School of Medicine at Dartmouth (GB and CSG), a Gordon and Betty Moore Foundation Data-Driven Discovery Investigator grant GBMF4552 (CSG), the National Science Foundation grant DBI-1458477, ABI 1458390 (CSG and GB), ABI 1458477 (PR), ABI 1458443 (SDM), and ABI 1458359 (IF), the National Institutes of Health Pioneer grant IDP1MH110234 (GB), and the Defense Advanced Research Projects Agency grant HR0011-15-1-0002 (GB).

LITERATURE CITED

Alonso, J. M., A. N. Stepanova, T. J. Leise, C. J. Kim, H. Chen et al., 2003 Genome-wide inserional mutagenesis of Arabidopsis thaliana. Science 301: 653–657. https://doi.org/10.1126/science.1086391

Asahina, K., K. Watanabe, B. J. Duistermars, E. Hoopfer, C. R. González et al., 2014 Tachykinin-expressing neurons control male-specific aggressive arousal in Drosophila. Cell 156: 221–235. https://doi.org/10.1016/j.cell.2013.11.045

Ashraf, S. I., X. Hu, J. Roote, and Y. T. Ip, 1999 The mesoderm determinant snail collaborates with related zinc-finger proteins to control Drosophila neurogenesis. EMBO J. 18: 6426–6438. https://doi.org/10.1093/emboj/18.22.6426

Aso, Y., K. Grübel, S. Busch, A. B. Friedrich, I. Siwanowicz et al., 2009 The mushroom body of adult Drosophila characterized by GAL4 drivers. J. Neurogenet. 23: 156–172. https://doi.org/10.1080/01677060802471718

Aso, Y., D. Hattori, Y. Yu, R. M. Johnston, N. A. Iyer et al., 2014 The neuronal architecture of the mushroom body provides a logic for associative learning. eLife 3: e04577. https://doi.org/10.7554/eLife.04577

Balfanz, S., T. Strünker, S. Frings, and A. Baumann, 2005 A family of octapamine receptors that specifically induce cyclic AMP production or Ca2 release in Drosophila melanogaster. J. Neurochem. 93: 440–451. https://doi.org/10.1111/j.1471-4159.2005.03034.x

Becnel, J., O. Johnson, J. Luo, D. R. Nääsel, and C. D. Nichols, 2011 The serotonin 5-HT7Dro receptor is expressed in the brain of Drosophila, and is essential for normal courtship and mating. PLoS One 6: e20800. https://doi.org/10.1371/journal.pone.0020800

Bozler, J., B. Z. Kacsoh, H. Chen, W. E. Theurkauf, Z. Weng et al., 2017 A systems level approach to temporal expression dynamics in Drosophila reveals clusters of long term memory genes. PLoS Genet. 13: e1007054. https://doi.org/10.1371/journal.pgen.1007054

Brody, T., and A. Cravchik, 2000 Drosophila melanogaster G protein-coupled receptors. J. Cell Biol. 150: F83–F88. https://doi.org/10.1083/jcb.150.2.F83

Chen, C., 1999. Control of cell migration and branch morphogenesis during Drosophila tracheal development. Chen, C. K., R. P. Kuhnlein, K. G. Eulenberg, S. Vincent, M. Affolter et al., 1998 The transcription factors KNIRPS and KNIRPS RELATED control cell migration and branch morphogenesis during Drosophila tracheal development. Development 125: 4959–4968.

Chikina, M. D., and O. G. Troyanskaya, 2011 Accurate quantification of functional analogy among close homologs. PLOS Comput. Biol. 7: e1001074. https://doi.org/10.1371/journal.pcbi.1001074

Chikina, M. D., C. Huttenhower, C. T. Murphy, and O. G. Troyanskaya, 2009 Global prediction of tissue-specific gene expression and context-dependent gene networks in Caenorhabditis elegans. PLOS Comput. Biol. 5: e1000417. https://doi.org/10.1371/journal.pcbi.1000417

Claridge-Chang, A., R. D. Roorda, E. Vrontou, L. Sjulson, H. Li et al., 2009 Writing memories with light-addressable reinforcement circuitry. Cell 139: 405–415. https://doi.org/10.1016/j.cell.2009.08.034

Cohn, R., I. Morante, and V. Ruta, 2015 Coordinated and compartmentalized neuromodulation shapes sensory processing in Drosophila. Cell 163: 1742–1755. https://doi.org/10.1016/j.cell.2015.11.019

Cozzetto, D., and D. T. Jones, 2017 Computational methods for annotation transfers from sequence, 55–67 in The Gene Ontology Handbook, Springer, New York.

Cozzetto, D., F. Minneci, H. Currant, and D. T. Jones, 2016 Ffpred 3: feature-based function prediction for all gene ontology domains. Sci. Rep. 6: 31865. https://doi.org/10.1038/srep31865

Curators, F., 2004 Swiss-Prot Project Members InterPro Project Members. Gene Ontology annotation in FlyBase through association of InterPro records with GO terms.
Sokolowski, M. B., 2001 Drosophila: genetics meets behaviour. Nat. Rev. Genet. 2: 879–890. https://doi.org/10.1038/35098592

Spradling, A. C., D. Stern, A. Beaton, E. J. Rhem, T. Laverty et al., 1999 The Berkeley Drosophila Genome Project gene disruption project: Single P-element insertions mutating 25% of vital Drosophila genes. Genetics 153: 135–177.

Sverchkov, Y., and M. Craven, 2017 A review of active learning approaches to experimental design for uncovering biological networks. PLOS Comput. Biol. 13: e1005466. https://doi.org/10.1371/journal.pcbi.1005466

Tejedor, F., X. Zhu, E. Kaltenbach, A. Ackermann, A. Baumann et al., 1995 minibrain: a new protein kinase family involved in postembryonic neurogenesis in Drosophila. Neuron 14: 287–301. https://doi.org/10.1016/0896-6273(95)90286-4

Tully, T., T. Preat, S. Boynton, and M. Del Vecchio, 1994 Genetic dissection of consolidated memory in Drosophila. Cell 79: 35–47. https://doi.org/10.1016/0092-8674(94)90398-0

Tully, T., 1987 Drosophila learning and memory revisited. Trends Neurosci. 10: 330–335. https://doi.org/10.1016/0166-2236(87)90090-7

Tully, T., and W. G. Quinn, 1985 Classical conditioning and retention in normal and mutant Drosophila melanogaster. J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol. 157: 263–277. https://doi.org/10.1007/BF01350033

Vázquez, A., A. Flammini, A. Maritan, and A. Vespignani, 2003 Global protein function prediction from protein-protein interaction networks. Nat. Biotechnol. 21: 697–700. https://doi.org/10.1038/nbt825

Wilson, C. A., J. Kreychman, and M. Gerstein, 2000 Assessing annotation transfer for genomics: quantifying the relations between protein sequence, structure and function through traditional and probabilistic scores. J. Mol. Biol. 297: 233–249. https://doi.org/10.1006/jmbi.2000.3550

Yan, H., K. Venkatesan, J. E. Beaver, N. Klitgord, M. A. Yildirim et al., 2010 A genome-wide gene function prediction resource for Drosophila melanogaster. PLoS One 5: e12139. https://doi.org/10.1371/journal.pone.0012139

Communicating editor: M. Arbeitman