Title: Machine learning analysis identifies *Drosophila Grunge/Atrophin* as an important learning and memory gene required for memory retention and social learning.

Authors: Balint Z Kacsoh[†], Casey S. Greene[§] and Giovanni Bosco[†*]

Affiliations:

[†] Department of Molecular and Systems Biology, Geisel School of Medicine at Dartmouth, Hanover, NH, 03755, USA.

[§] Department of Systems Pharmacology and Translational Therapeutics, Perelman School of Medicine, University of Pennsylvania, PA, 19104, USA.

*Corresponding author: (GB) giovanni.bosco@dartmouth.edu

RUNNING TITLE: Machine learning predicts novel genes in Drosophila learning and memory.

KEY WORDS

- *D. melanogaster*
- Learning and memory
- Long-term memory
- Behavior
- Histone modification
- Machine-learning
- Gene prediction

ABSTRACT

High throughput experiments are becoming increasingly common, and scientists must balance hypothesis driven experiments with genome wide data acquisition. We sought to predict novel genes involved in Drosophila learning and long-term memory from existing public high-throughput data. We performed an analysis using PILGRM, which analyzes public gene expression compendia using machine learning. We evaluated the top prediction alongside genes involved in learning and memory in IMP, an interface for functional relationship networks. We identified *Grunge/Atrophin* (*Gug/Atro*), a transcriptional repressor, histone deacetylase, as our top candidate. We find, through multiple, distinct assays, that *Gug* has an active role as a modulator of memory retention in the fly and its function is required in the adult mushroom body. Depletion of *Gug* specifically in neurons of the adult mushroom body, after cell division and neuronal development is complete, suggests that *Gug* function is important for memory retention by regulating neuronal activity, and not simply by altering neurodevelopment. Our study provides a previously uncharacterized role for *Gug* as a possible regulator of neuronal plasticity at the interface of memory retention and memory extinction.

INTRODUCTION

A unique and fundamental characteristic of higher order organisms is that they posses the ability to recall and remember past experiences. The ability to remember past experiences proves to be advantageous when an organism is presented with challenges bearing resemblance to previous experiences. The biology of memory and memory formation has been probed with a vast array of experimental techniques that range from behavioral testing to physiologically

correlates to experimental perturbation through the use of chemical, anatomical, and pharmacological stresses (Greenspan 1995, 747-750) . Genetic manipulations provide important insight into behavioral memory output, allowing one to understand the genetic and molecular factors that govern memory formation, retention, and retrieval, while providing clues to evolutionarily conserved mechanisms that could allow for future vertical integration (Dubnau and Tully 1998, 407-444; Tully *et al.* 1994, 35-47) .

Drosophila melanogaster has proven to be an important genetic model system for understanding the genetic basis of memory (Greenspan 1995, 747-750; Davis 2005, 275-302; Margulies *et al.* 2005, R700-R713; McGuire *et al.* 2005, 328-347) . Drosophila memory has been analyzed through the use of associative and non-associative assays. Work using associative learning paradigms has provided genes and genetic functions that are now looked upon as "classical" learning. These memory genes were found through the use of associative-fear-conditioning assays (McGuire *et al.* 2005, 328-347; Tully 1987, 330-335; Tully *et al.* 1994, 35-47; Dubnau and Tully 1998, 407-444) . These gene candidates were further elucidated and expanded in other assays, including the use of non-associative olfactory memory (Das *et al.* 2011, E646-54; McCann *et al.* 2011, E655-62; Ramaswami 2014, 1216-1229) . Although these studies are critical in elucidating genetic and physiological mechanisms of learning and memory, these assays use non-ecologically relevant forms of stimuli, and thus, the genes identified might only represent a small fraction of learning and memory genes in the insect brain.

Other forms of memory have also been tested using these classical genes that involve more ecologically relevant forms of stimuli to the fly life cycle. For example, Drosophila larvae are regularly infected by endoparasitoid wasps. In natural *D. melanogaster* populations, upwards of 90% of fly larvae are found to be infected by wasps, suggesting they exert extremely strong

selection pressures on Drosophila populations (Driessen *et al.* 1989, 409-427; Fleury *et al.* 2004, 181-194; LaSalle 1993, 197-215) . Drosophila larvae can mount a cellular immune response (Carton and Nappi 1997, 218-227) , or adult *D. melanogaster* females can alter egg-laying behavior after an encounter with endoparasitoid wasps. A change in oviposition behavior entails at least two very different and quantifiable behavioral responses. First, if high ethanol containing food is made available to adult Drosophila, then female flies in the presence of wasps will actively prefer to lay eggs on ethanol-laden food (Kacsoh *et al.* 2015a, 1143-1157; Kacsoh *et al.* 2013, 947-950) . This behavior persists even after the wasp threat is removed, demonstrating a memory formation and retention (Kacsoh *et al.* 2015a, 1143-1157; Kacsoh *et al.* 2013, 947-950) . Second, Drosophila females will depress their oviposition rate in the presence of wasps and maintain this depression following wasp removal. Wasp exposed flies will also communicate the wasp threat to naïve flies that have never seen the wasp threat, which in turn will also depress their oviposition rate (Kacsoh *et al.* 2015b, 10.7554/eLife.07423; Lefevre *et al.* 2012, 230-233; Lynch *et al.* 2016) . Both of these behaviors persist in wild-type flies for multiple days after wasp encounter, allowing one to probe questions regarding memory formation and retention. These memories are regulated, in part, by known learning and memory genes (Kacsoh *et al.* 2015b, 10.7554/eLife.07423; Kacsoh *et al.* 2015a, 1143-1157; Kacsoh *et al.* 2013, 947-950) .

Previously identified learning and memory genes may only represent a small fraction of the range of learning and memory genes and gene functions in the insect brain required for memory acquisition, retention, and recall. Thus, it is valuable to define and mechanistically identify novel genes and gene products in addition to those known to be important for behaviors mainly observed in laboratory conditions that may not fully represent natural behaviors. More importantly, an understanding of how non-associative memories and behaviors persist over time

first requires the development of new methodologies that are independent of classical mutagenesis-based approaches.

In the sequencing age, vast compendia of existing data present the opportunity to address important biological questions (Faith *et al.* 2007, e8; Yan *et al.* 2010, e12139; Chikina *et al.* 2009, e1000417) . Given this abundance, the question becomes, how does one utilize so much data to ask specific questions? Genome-wide expression data capture a vast array of perturbations in a diverse range of organisms, biological processes, and tissues, are publically available and can be utilized to find new targets. Data-mining algorithms applied to large data sets can uncover novel, biologically relevant functions of genes that might be otherwise overlooked as candidates (Greene *et al.* 2014, 1896-1900; Greene and Troyanskaya 2012, 95-100) . PILGRM (the Platform for Interactive Learning by Genomics Results Mining) is a data-mining platform that allows its users to encode specific questions about a biological function of interest into user created gene sets. Probing for a specific biological function is done by the user curating a gold standard of genes that are relevant to a particular pathway or process, which PILGRM uses for a supervised machine learning analysis of global microarray expression collections (Greene and Troyanskaya 2011, W368-74) . The user can also provide a list of negative control genes, which allows the user to improve prediction specificity. PILGRM then trains a support vector machine (SVM (Joachims 2006, 217-226)) classifier with these gene sets to discover novel relevant genes (Greene and Troyanskaya 2011, W368-74) . The SVM analysis does this by assigning relatively high weights to conditions that differentiate gold standard (positive) genes from those in the negative standard (Greene and Troyanskaya 2011, W368-74) . Thus, this machine learning approach can allow one to identify novel gene targets/functions in order to further elucidate a biologically meaningful process, such as memory.

In this study, we use PILGRM to identify novel genetic candidates by using a gold standard comprised of classically identified learning and memory genes. We examined the top prediction from PILGRM, a histone-deacetylase, *Grunge* (*gug*, CG6964), in a *Drosophila melanogaster* functional relationship network and found that it was connected to many gold standard genes. We then use the natural wasp predator system to probe a role for this previously untested gene within *Drosophila melanogaster* learning and memory using three distinct learning and memory paradigms that test for long term memory formation, long term memory retention, and social learning. We test the hypothesis that *gug* may be a high level regulator of learning and memory. Histone-acetylation activity has been previously identified as a regulator of mouse hippocampal learning and memory (Mews *et al.* 2017) , where histone acetylation yields activation of early memory genes and perturbation of this system presents defects in memory formation. A functional role for histone deacetylase activity in memory is consistent previous reports showing that *Rpd3* (HDAC1) functions in the Drosophila mushroom body to facilitate long-term memory formation (Fitzsimons and Scott 2011, e29171) .

RESULTS

PILGRM ANALYSIS REVEALS NOVEL GENE TARGETS

In order to identify novel genes involved in learning and memory, we turned to the PILGRM data-mining tool (http://pilgrm.princeton.edu). We input a positive, gold standard list as well as a list of randomly selected negatives. Our positive standard list was comprised of previously identified learning and memory genes. Specifically, we input the genes: *for* (Donlea *et al.* 2012, 2613-2618) , *Akt1* (Guo and Zhong 2006, 4004-4014) , *Fmr1* (Kanellopoulos *et al.* 2012, 13111-13124; Kacsoh *et al.* 2015b, 10.7554/eLife.07423; Kacsoh *et al.* 2015a, 1143-1157)

, *S6k* (Vargas *et al.* 2010, 1006-1011) , *CrebB-17A* (Perazzona *et al.* 2004, 8823-8828) , *Adf1* (DeZazzo *et al.* 2000, 145-158; Kacsoh *et al.* 2013, 947-950; Kacsoh *et al.* 2015a, 1143-1157) , *amn* (Aldrich *et al.* 2010, 33-41; Kacsoh *et al.* 2015b, 10.7554/eLife.07423; Kacsoh *et al.* 2015a, 1143-1157) , *rut* (Scheunemann *et al.* 2013, 17422-17428; Kacsoh *et al.* 2015a, 1143-1157; Kacsoh *et al.* 2015b, 10.7554/eLife.07423) , *Nf1* (Li *et al.* 2013, 5821-5833) , *CaMKII* (Mehren and Griffith 2006, 686-689) , *orb2* (Krüttner *et al.* 2012, 383-395; Kacsoh *et al.* 2015b, 10.7554/eLife.07423; Kacsoh *et al.* 2015a, 1143-1157) , and *shi* (Iyengar *et al.* 2011, 883-900) (Supplementary file 1).

The analysis was performed using PILGRM's *Drosophila melanogaster* compendium, which consists of gene expression data sets from the Gene Expression Omnibus (GEO). PILGRM provides a visual interpretation for precision of predictions at different recall thresholds (Figure 1 A). We also gain visual analysis of true positive rate at various false positive rate thresholds and find that the area under the curve, shown as the shaded region, is 0.8130 (Figure 1 B). PILGRM also provides a list of novel predictions (Figure 1 C, Supplementary file 2). In this case, the top novel prediction is the gene *Grunge* (*gug*), whose molecular function is listed as a histone deacetylase (Wang *et al.* 2006, 525-530; Zhang *et al.* 2013, 575-583; Zhang *et al.* 2002, 45-56; Wang *et al.* 2008, 555-562; Yeung *et al.* 2017, 10.7554/eLife.23084) . This gene has no previously described role in learning and memory in Drosophila, however the Drosophila gene expression atlas (Fly:Atlas) reports *Gug* expression as most highly enriched in the fly brain (Chintapalli *et al.* 2007, 715-720) . Recent work has shown that aberrations in mouse hippocampal histone acetylation yields mice with impaired long-term spatial memory formation, a cognitive process that relies on histone acetylation to activate early memory genes (Mews *et al.* 2017; Kandel *et al.* 2014, 163-186; Korzus *et al.* 2004, 961-972; Gräff and Tsai 2013, 97-111;

Wood *et al.* 2005, 111-119) . In Drosophila, the HDAC1 homolog, *Rpd3*, also is implicated in learning and memory functions (Fitzsimons and Scott 2011, e29171) . Given these observations, in conjunction with observations showing epigenetic deregulation as a mechanism for neuropsychiatric diseases (Kandel *et al.* 2014, 163-186; Zovkic *et al.* 2013, 61-74; Gräff and Tsai 2013, 97-111; Walker *et al.* 2015, 112-121) , we wished to elucidate the possibility of the histone deacetylase *Gug* playing a role in learning and long-term memory.

GUG SHOWS HIGH CONNECTIVITY TO KNOWN LEARNING AND MEMORY GENES

In order to further examine *Gug* before moving to molecular experiments, we used the integrative multi-species prediction (IMP) webserver to determine the extent to which *Gug* was connected with our known learning and memory gene set (Wong *et al.* 2012, W484-90; Wong *et al.* 2015, W128-33) . IMP integrates gene-pathway annotations from the selected organism in addition to mapping functional analogs. This system has been shown to provide accurate gene-process predictions (Chikina and Troyanskaya 2011, e1001074) .

We queried the network with *Gug* and our gold standard set. We set the network filter to only show interactions whose minimum relationship confidence was 0.85, a stringent threshold. We found that *Gug* was connected, both directly and indirectly, to our known learning and memory genes (Figure 2). Unsurprisingly, this network was enriched for the biological processes of long-term memory (26.3%, p-value: 1.29e-8), short-term memory (15.8%, p-value: 1.85e-6), learning (21.1%, p-value: 1.21e-5), and associative learning (18.4%, p-value: 4.51e-5). Given the PILGRM and IMP results, we sought to test the role of *Gug* in a variety of assays probing for its role in memory formation, retention, and social learning.

GUG HAS A ROLE IN LONG TERM MEMORY RETENTION

We first wished to identify a role for *gug* in long-term memory formation and retention. We tested this question using two assays, both involving the predatory wasp *Leptopilina heterotoma* (strain Lh14). Both behavior assays utilize adult *D. melanogaster*, who can alter its egg-laying behavior when encountering endoparasitoid wasps. We first used an ethanol seeking food preference assay where, following a wasp exposure, female flies will actively prefer to lay eggs on ethanol-laden food (Kacsoh *et al.* 2013, 947-950; Kacsoh *et al.* 2015a, 1143-1157) . Briefly, we exposed 3-5 day old flies in 3 batches of 100 female flies and 20 male flies with 50 female wasps. Control flies underwent the same process, but lacked wasps. Following a 24-hour exposure, flies and wasps were separated. Exposed and unexposed flies are transferred in batches of 5 female and 1 male flies to a fly corral, a Petri dish with holes, where the center of the two holes were 6 cm apart, and the edge of the two holes were 7.2 cm apart. We place Falcon tube caps to the holes. The caps contain 0.375 grams of flaky instant blue hydrated fly food with either distilled-water or distilled water with ethanol (6% by volume). Caps are changed every 24-hours for 3-days and egg counts were performed on caps. All counts are blinded (see methods) (Kacsoh *et al.* 2015a, 1143-1157) .

We find that wild-type *D. melanogaster* continue to oviposit on ethanol laden food across each of the 3, 24-hour time points tested following wasp exposure (Figure 3 A). Interestingly, we find that *Gug* heterozygous mutants (Spradling *et al.* 1999, 135-177) (*Gug*/+) show a defect in persistence of oviposition on ethanol-laden food (Figure 3 B). These flies show a defect in memory retention, starting on day 2. Day 1 of the assay does not show a memory defect, suggesting that memory formation is not affected. However, these experiments cannot not

exclude the possibility that the *Gug* gene product is required in non-neural tissues, and that the behavior is not specific to neuronal function.

The mushroom body (MB) of the adult brain is thought to be required for behaviors that are dependent on learning and memory (Aso *et al.* 2009, 156-172; Claridge-Chang *et al.* 2009, 405-415; Schwaerzel *et al.* 2003, 10495-10502; Masse *et al.* 2009, R700-R713) . The behaviors we set out to test *Gug*'s role in have been shown to be MB dependent (Kacsoh *et al.* 2015b, 10.7554/eLife.07423; Kacsoh *et al.* 2015a, 1143-1157) . A critical question that arose after testing the *Gug* heterozygotes was whether *Gug* was required in the MB for memory formation and/or retention. We used the GAL4/UAS system to drive expression of an RNA-hairpin targeting *Gug* mRNA (Gug^{RNAi}) in conjunction with the MB drivers OK107 and MB247. We find that driver lines and the line containing the *Gug* hairpin have wild-type memory retention (Supplementary Figure 1 A-C). However, flies expressing the Gug^{RNAi} in the MB phenocopy the *Gug* heterozygote memory retention defect (Figure 3 C-D), showing no defect in memory formation.

While the MB drivers in conjunction with the Gug^{RNAi} showed a phenotype, the constitutive expression of the hairpin presents the possibility of *Gug*'s function is essential for early neuronal development. This could mean that the knockdown phenotypes may simply reflect developmental defects that precluded proper adult MB functions. In order to test this possibility, we turned to the GAL4-based Gene-Switch system where the GAL4 transcription factor is fused to the human progesterone ligand-binding domain (Burcin *et al.* 1999, 355-360) . We used flies expressing the Gene-Switch transgene specifically in the MB where only an administration of the pharmacological Gene-Switch ligand, RU486, could activate the GAL4 transcription factor (Mao *et al.* 2004, 198-203) . We find that feeding of RU486 to the outcrossed

MB GeneSwitch driver has wild-type memory retention, showing that feeding of the ligand does not perturb memory formation or retention (Supplementary Figure 1 D). We also observe wild-type memory retention when the MB GeneSwitch is in combination with the *Gug*^{RNAi} when fed vehicle control (Figure 3 E). When fed RU486, these flies show the same memory defect as when *Gug*^{RNAi} is driven by a constitutively active MB driver and the *Gug* heterozygotes (Figure 3 E). Thus, we believe that the memory retention defect observed is a result of *Gug* being a necessary gene product in the MB to facilitate memory retention in the adult Drosophila brain. Given these data, we use the MB GeneSwitch line to drive expression of the *Gug* hairpin for all subsequent experiments, in order to delineate the role of *Gug* in memory retention from other important functions it may also perform during development.

We decided to examine the role of *Gug* in memory retention using a second assay. This assay utilizes another *D. melanogaster* behavior following wasp exposure. In this behavior, Drosophila female flies depress their oviposition rate following wasp exposure. This depression is observed for multiple days after the predator threat is removed (Kacsoh *et al.* 2015b, 10.7554/eLife.07423; Lefevre *et al.* 2012, 230-233) . We used the Fly Condos to measure oviposition rate (Kacsoh *et al.* 2015b, 10.7554/eLife.07423) . Briefly, *D. melanogaster* were exposed for 24 hours to wasps in cylindrical 7.5cm long by 1.5cm diameter tubes of the Fly Condos. Each tube contains five female flies and one male fly, either with three female wasps (exposed) or with no wasps at all (unexposed) (see methods). After 24-hours, wasps are removed, and all flies are placed into new, clean Fly Condos. We repeat this transfer for each of 3 days following wasp exposure. At every 24-hour time point, food-plates are removed, replaced with new food plates, and embryos are counted in a blinded manner.

Consistent with previous observations, wild-type flies depress oviposition in the presence of wasps and for 3 days following wasp exposure (Figure 4 A, Supplementary Figure 2 B) (Lefevre *et al.* 2012, 230-233; Kacsoh *et al.* 2015b, 10.7554/eLife.07423) . We find that *Gug* heterozygotes (*Gug*/+) have wild-type memory on day 1 following wasp exposure, but show accelerated memory decay (Figure 4 B). In order to ascertain the specificity of the phenotype with respect to memory, we again utilized the MB GeneSwitch line. We find that the Gug^{RNAi} construct outcrossed to *Canton S* has wild-type memory retention (Supplementary Figure 2 A) (Kacsoh *et al.* 2015b, 10.7554/eLife.07423) . When the Gug^{RNAi} is expressed in conjunction with the MB GeneSwitch fed vehicle only, we observe wild-type memory retention (Figure 4 C). When this line is fed RU486, leading to *Gug* knock-down, we observe a memory retention defect comparable to the heterozygote, where day one is unaffected, but day two begins to show accelerated memory decay (Figure 4 D). These data again show no defect in memory formation. Collectively, the data suggest that *Gug* has a functional role in the MB, modulating memory retention.

GUG HAS A ROLE IN TEACHING BEHAVIOR

We observe a memory retention, but not formation, defect in *Gug* heterozygous mutants and *Gug* knockdown in the MB. We sought to ask whether this memory retention defect translates to other behaviors that are present following wasp exposure. Following wasp exposure, flies depress oviposition and communicate the threat of wasps to naïve, unexposed flies. We term the experienced flies as “teachers” and the naïve flies “students.” We hypothesized that the memory retention defects observed (Figures 3-4) may translate to defects in teaching behavior. To ask this, we utilized the Fly Condos (Kacsoh *et al.* 2015b, 10.7554/eLife.07423) . Briefly, we

exposed wild-type flies to wasps for 24-hours, followed by wasp removal. We then placed the exposed and unexposed teacher flies into new Fly Condos with three naïve female flies expressing Histone-RFP (His-RFP) for an additional 24-hours (see methods). Following a 24-hour co-incubation, we removed His-RFP students, placed exposed and unexposed teachers into new Fly Condos containing a new batch of 3 female naïve His-RFP flies. We repeated this process 3 times following wasp exposure as a means of testing the maintenance and ability of teaching behavior. New batches of students were used each day. At every 24-hour time point, food-plates were removed, replaced with new food plates, and embryos were counted in a blinded manner. The His-RFP line was ideal for discriminating mixed populations of non-RFP and RFP embryos, allowing us to probe specificity of student and teacher behavior.

Consistent with previous data, we find that wild-type flies can instruct multiple batches of students across the 3 days tested (Figure 5 A) (Kacsoh *et al.* 2015b, 10.7554/eLife.07423) . We find that *Gug* heterozygotes (*Gug*/+) have wild-type teaching ability on day 1, which quickly decays on days 2 and 3 (Figure 5 B). To again probe for memory specificity, we turned to the MB GeneSwitch line. We find that outcrossed Gug^{RNAi} flies have wild-type teaching behavior across each time-point tested (Supplementary Figure 3). When the Gug^{RNAi} is expressed in conjunction with the MB GeneSwitch line fed vehicle only, we observe wild-type teaching behavior across all 3 days. When this line is fed RU486, allowing for expression of Gug^{RNAi}, we find wild-type teaching ability on day 1, but find decaying teaching ability across days 2 and 3, similar to the heterozygote. Collectively, the data suggest a role for *Gug* in memory maintenance of teaching behavior.

GUG HAS A ROLE IN STUDENT BEHAVIOR

Finally, we wished to test the role of *Gug* in social learning behavior. Since both *Gug* heterozygotes and RNAi depleted Drosophila were able to perceive and respond to wasp presence and form memory one prediction might be that *Gug* deficient flies should also be able to learn from wasp-exposed teacher flies. It is alternatively possible that learning in the context of social interactions with other flies could be different from learning that takes place directly from wasp-exposure. In this latter case, the molecular, cellular and/or neuronal requirements for social learning might be different than those for non-social learning. In order to test the role of *Gug* in social learning, we modified the Fly Condo oviposition depression assay. Briefly, we exposed His-RFP flies to wasps for 24-hours, followed by wasp removal. We then placed the exposed and unexposed teacher flies into new Fly Condos with three naïve female flies that were either wild-type, or containing a *Gug* RNAi-hairpin construct, for an additional 24-hours. Following a 24-hour co-incubation, food-plates were removed, and embryos were counted in a blinded manner.

Consistent with previous results, we find that wild-type flies are able to learn from His-RFP teacher flies (Figure 6 A). Interestingly, we find that *Gug* heterozygous (*Gug*/+) students have an impaired social learning ability when compared to wild-type students (Figure 6 B). We then wished to test the role of *Gug* in the MB in a social learning context. We find that the outcrossed Gug^{RNAi} show wild-type learning ability (Supplementary Figure 4). When expressed in combination with the MB GeneSwitch line and fed vehicle only, we observe wild-type learning (Figure 6 C). When this line is fed RU486, allowing for expression of Gug^{RNAi}, we observe an impaired social learning ability (Figure 6 D). Collectively, these data strongly suggest that *Gug* has an important function in social learning and that this function specifically occurs in the MB information-processing center of the fly brain. The data also suggest that the

type of learning taking place from wasp exposure is fundamentally different from the type of learning taking place during social learning.

DISCUSSION

In this study, we have presented the use of a machine-learning platform, PILGRM, combined with functional network analysis via IMP, to identify and evaluate novel regulators of learning and memory in Drosophila. We identified *Gug*, a transcriptional repressor known to complex with the Rpd3/HDAC histone deacetylase (Fitzsimons and Scott 2011, e29171) . To our knowledge *Gug* itself has not been previously implicated to function in *Drosophila* memory. We tested *Gug*'s role in memory formation and long-term memory retention through two unique assays and find that, while memory formation is normal, memory retention is severely perturbed when the protein is missing in the fly mushroom body. We note that this function in memory retention for *Gug* is different from what has previously been reported for Rpd3, and therefore constitutes a novel function for *Gug* specifically in maintenance of long term memory. This observation raises the interesting possibility that the HDAC activity of Rpd3 recruited to chromatin by Gug could also be important in memory retention. Importantly, the memory defects in *Gug* deficient flies are not simply a result of developmental defects, given the data utilizing the MB GeneSwitch lines, where the depletion of *Gug* in adult MB neurons takes place after cell division and development has been completed. This suggests that *Gug* has an active role in post-mitotic neurons throughout memory retention and/or retrieval in the fly MB, most likely via its transcriptional repressor function. Additionally, we find that perturbations in *Gug* yield teaching behavior that is normal on day one, but as the memory decays, so does the teaching behavior. This suggests that both memory and teaching behavior are modulated by *Gug* activity. Of

greatest interest, we find that social learning is also perturbed in *Gug* deficient flies, where students are not able to interpret information from wild-type teacher flies. This is important because it suggests that molecular requirements for learning in a social context from experienced flies may require fundamentally different molecular and/or neuronal circuitry, and it is therefore distinct from learning in non-social contexts. Interestingly, humans with trinucleotide expansions within the *Atrophin 1* gene, the human *Gug* homolog, exhibit autism-like behaviors (Licht and Lynch 2002, 51-54) .

Previous work using a murine hippocampal model demonstrates a role for histone acetylation and subsequent expression of neuronal genes involved in memory formation (Mews *et al.* 2017) . We propose a model where histone acetylation and decetylation are acting in concert to promote neuronal plasticity (Figure 7). Given our results, we hypothesize that deacetylase activity is important in the memory retrieval process, where certain gene activity is turned off, or repressed. This repression may help elongate the memory retention, as without deacetlyase activity, we observe accelerated memory decay. The targets of *Gug* in the MB may serve as memory extinction genes (Abel and Lattal 2001, 180-187; Rudenko *et al.* 2013, 1109-1122) , or genes that promote memory decay, and repression of these genes promotes memory maintenance.

Collectively, our data highlights the value of machine-learning approaches for biological function prediction via biologist-driven analyses. PILGRM enables data-driven experimentation in conjunction with knowledge-based discovery through the selection of a biologically relevant gold standard. Given that PILGRM utilizes a large collection of public gene expression data, it may identify genes that are not differentially regulated in an investigator's own experiment, but act in unison with curated genes of interest. It can also identify consistent patterns across a

compendium. This approach may help to identify regulators where small, potentially non-statistically significant changes in transcript abundances can have biologically meaningful changes. These changes can be subtle, yet could lead to neuron plasticity. Epigenetic mechanisms are continually found to be important regulators of neuronal plasticity and functional output. These mechanisms are implicated in maintaining neuronal homeostasis, and perturbations have been implicated in neuropsychiatric diseases (Walker *et al.* 2015, 112-121; Mews *et al.* 2017; Gräff and Tsai 2013, 97-111; Kandel *et al.* 2014, 163-186) . Our study provides a previously uncharacterized role for *Gug* as a possible regulator of neuronal plasticity at the intersection of memory retention and memory extinction.

Acknowledgements: We thank Greg Roman, FlyBase, and the Bloomington Drosophila Stock Center, for stocks. We thank Sassan Hodge for technical assistance. We thank Julianna Bozler for technical and graphical assistance. We acknowledge grants from Geisel School of Medicine at Dartmouth, a Gordon and Betty Moore Foundation Data-Driven Discovery Investigator grant GBMF4552 (CSG), the National Institute of Health Pioneer grant 1DP1MH110234 (GB), and the Defense Advanced Research Projects Agency grant HR0011-15-1-0002 (GB).

Bibliography

Abel, T. and K. M. Lattal, 2001 Molecular mechanisms of memory acquisition, consolidation and retrieval . Current opinion in neurobiology **11**: 180-187.

Aldrich, B. T., J. Kasuya, M. Faron, H. Ishimoto, and T. Kitamoto, 2010 The amnesiac gene is involved in the regulation of thermal nociception in Drosophila melanogaster . Journal of neurogenetics **24**: 33-41.

Aso, Y., K. Grübel, S. Busch, A. B. Friedrich, I. Siwanowicz*, et al*., 2009 The mushroom body of adult Drosophila characterized by GAL4 drivers . Journal of neurogenetics **23**: 156-172.

Burcin, M. M., G. Schiedner, S. Kochanek, S. Y. Tsai, and B. W. O'Malley, 1999 Adenovirus-mediated regulable target gene expression in vivo . Proceedings of the National Academy of Sciences of the United States of America **96**: 355-360.

Carton, Y. and A. Nappi, 1997 Drosophila cellular immunity against parasitoids . Parasitology today **13**: 218-227.

Chikina, M. D. and O. G. Troyanskaya, 2011 Accurate quantification of functional analogy among close homologs . PLoS Comput Biol **7**: e1001074.

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.

Chintapalli, V. R., J. Wang, and J. A. Dow, 2007 Using FlyAtlas to identify better Drosophila melanogaster models of human disease . Nature genetics **39**: 715-720.

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.

Das, S., M. K. Sadanandappa, A. Dervan, A. Larkin, J. A. Lee, *et al*., 2011 Plasticity of local GABAergic interneurons drives olfactory habituation . Proceedings of the National Academy of Sciences of the United States of America **108**: E646-54.

Davis, R. L., 2005 Olfactory memory formation in Drosophila: from molecular to systems neuroscience . Annu.Rev.Neurosci. **28**: 275-302.

DeZazzo, J., D. Sandstrom, S. de Belle, K. Velinzon, P. Smith, *et al*., 2000 nalyot, a mutation of the Drosophila myb-related Adf1 transcription factor, disrupts synapse formation and olfactory memory . Neuron **27**: 145-158.

Donlea, J., A. Leahy, M. S. Thimgan, Y. Suzuki, B. N. Hughson, *et al*., 2012 Foraging alters resilience/vulnerability to sleep disruption and starvation in Drosophila . Proceedings of the National Academy of Sciences of the United States of America **109**: 2613-2618.

Driessen, G., L. Hemerik, and J. J. Van Alphen, 1989 Drosophila species, breeding in the stinkhorn (Phallus impudicus Pers.) and their larval parasitoids . Netherlands Journal of Zoology **40**: 409-427.

Dubnau, J. and T. Tully, 1998 Gene discovery in Drosophila: new insights for learning and memory . Annual Review of Neuroscience **21**: 407-444.

Faith, J. J., B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, *et al*., 2007 Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles . PLoS biol **5**: e8.

Fitzsimons, H. L. and M. J. Scott, 2011 Genetic modulation of Rpd3 expression impairs long-term courtship memory in Drosophila . PLoS One **6**: e29171.

Fleury, F., N. Ris, R. Allemand, P. Fouillet, Y. Carton, *et al*., 2004 Ecological and Genetic Interactions in Drosophila–parasitoids Communities: A Case Study with D. Melanogaster, D. Simulans and their Common Leptopilina Parasitoids in Southe-astern France . Genetica **120**: 181-194.

Gautier, L., L. Cope, B. M. Bolstad, and R. A. Irizarry, 2004 affy--analysis of Affymetrix GeneChip data at the probe level . Bioinformatics (Oxford, England) **20**: 307-315.

Gentleman, R. C., V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, *et al*., 2004 Bioconductor: open software development for computational biology and bioinformatics . Genome biology **5**: R80.

Gräff, J. and L. Tsai, 2013 Histone acetylation: molecular mnemonics on the chromatin . Nature Reviews Neuroscience **14**: 97-111.

Greene, C. S., J. Tan, M. Ung, J. H. Moore, and C. Cheng, 2014 Big data bioinformatics . Journal of cellular physiology **229**: 1896-1900.

Greene, C. S. and O. G. Troyanskaya, 2012 Accurate evaluation and analysis of functional genomics data and methods . Annals of the New York Academy of Sciences **1260**: 95-100.

Greene, C. S. and O. G. Troyanskaya, 2011 PILGRM: an interactive data-driven discovery platform for expert biologists . Nucleic acids research **39**: W368-74.

Greenspan, R. J., 1995 Flies, genes, learning, and memory . Neuron **15**: 747-750.

Guo, H. F. and Y. Zhong, 2006 Requirement of Akt to mediate long-term synaptic depression in Drosophila . The Journal of neuroscience : the official journal of the Society for Neuroscience **26**: 4004-4014.

Huttenhower, C., M. Schroeder, M. D. Chikina, and O. G. Troyanskaya, 2008 The Sleipnir library for computational functional genomics . Bioinformatics (Oxford, England) **24**: 1559-1561.

Irizarry, R. A., L. Gautier, and L. M. Cope, 2003 An R package for analyses of Affymetrix oligonucleotide arrays, 102-119 in *The analysis of gene expression data*. Springer.

Iyengar, B., C. J. Chou, K. Vandamme, M. Klose, X. Zhao, *et al*., 2011 Silencing synaptic communication between random interneurons during Drosophila larval locomotion . Genes, Brain and Behavior **10**: 883-900.

Kacsoh, B. Z., J. Bozler, S. Hodge, M. Ramaswami, and G. Bosco, 2015a A novel paradigm for nonassociative long-term memory in Drosophila: predator-induced changes in oviposition behavior . Genetics **199**: 1143-1157.

Kacsoh, B. Z., J. Bozler, M. Ramaswami, and G. Bosco, 2015b Social communication of predator-induced changes in Drosophila behavior and germ line physiology . eLife **4**: 10.7554/eLife.07423.

Kacsoh, B. Z., Z. R. Lynch, N. T. Mortimer, and T. A. Schlenke, 2013 Fruit flies medicate offspring after seeing parasites . Science (New York, N.Y.) **339**: 947-950.

Kandel, E. R., Y. Dudai, and M. R. Mayford, 2014 The molecular and systems biology of memory . Cell **157**: 163-186.

Kanellopoulos, A. K., O. Semelidou, A. G. Kotini, M. Anezaki, and E. M. Skoulakis, 2012 Learning and memory deficits consequent to reduction of the fragile X mental retardation protein result from metabotropic glutamate receptor-mediated inhibition of cAMP signaling in Drosophila . The Journal of neuroscience : the official journal of the Society for Neuroscience **32**: 13111-13124.

Korzus, E., M. G. Rosenfeld, and M. Mayford, 2004 CBP histone acetyltransferase activity is a critical component of memory consolidation . Neuron **42**: 961-972.

Krüttner, S., B. Stepien, J. N. Noordermeer, M. A. Mommaas, K. Mechtler, *et al*., 2012 Drosophila CPEB Orb2A mediates memory independent of Its RNA-binding domain . Neuron **76**: 383-395.

LaSalle, J., 1993 Parasitic Hymenoptera, biological control and biodiversity. Hymenoptera and Biodiversity.: 197-215.

Lefevre, T., J. C. de Roode, B. Z. Kacsoh, and T. A. Schlenke, 2012 Defence strategies against a parasitoid wasp in Drosophila: fight or flight?. Biology letters **8**: 230-233.

Li, W., M. Cressy, H. Qin, T. Fulga, D. Van Vactor, *et al*., 2013 MicroRNA-276a functions in ellipsoid body and mushroom body neurons for naive and conditioned olfactory avoidance in Drosophila . The Journal of neuroscience : the official journal of the Society for Neuroscience **33**: 5821-5833.

Licht, D. J. and D. R. Lynch, 2002 Juvenile dentatorubral-pallidoluysian atrophy: new clinical features . Pediatric neurology **26**: 51-54.

Lynch, Z. R., T. A. Schlenke, and J. Roode, 2016 Evolution of behavioural and cellular defences against parasitoid wasps in the Drosophila melanogaster subgroup . Journal of Evolutionary Biology.

Mao, Z., G. Roman, L. Zong, and R. L. Davis, 2004 Pharmacogenetic rescue in time and space of the rutabaga memory impairment by using Gene-Switch . Proceedings of the National Academy of Sciences of the United States of America **101**: 198-203.

Margulies, C., T. Tully, and J. Dubnau, 2005 Deconstructing memory in Drosophila . Current biology **15**: R700-R713.

Masse, N. Y., G. C. Turner, and G. S. Jefferis, 2009 Olfactory information processing in Drosophila . Current Biology **19**: R700-R713.

McCann, C., E. E. Holohan, S. Das, A. Dervan, A. Larkin, *et al*., 2011 The Ataxin-2 protein is required for microRNA function and synapse-specific long-term olfactory habituation . Proceedings of the National Academy of Sciences of the United States of America **108**: E655-62.

McGuire, S. E., M. Deshazer, and R. L. Davis, 2005 Thirty years of olfactory learning and memory research in Drosophila melanogaster . Progress in neurobiology **76**: 328-347.

Mehren, J. E. and L. C. Griffith, 2006 Cholinergic neurons mediate CaMKII-dependent enhancement of courtship suppression . Learning & memory (Cold Spring Harbor, N.Y.) **13**: 686-689.

Mews, P., G. Donahue, A. M. Drake, V. Luczak, T. Abel, *et al*., 2017 Acetyl-CoA synthetase regulates histone acetylation and hippocampal memory . Nature.

Perazzona, B., G. Isabel, T. Preat, and R. L. Davis, 2004 The role of cAMP response element-binding protein in Drosophila long-term memory . The Journal of neuroscience : the official journal of the Society for Neuroscience **24**: 8823-8828.

Ramaswami, M., 2014 Network plasticity in adaptive filtering and behavioral habituation . Neuron **82**: 1216-1229.

Rudenko, A., M. M. Dawlaty, J. Seo, A. W. Cheng, J. Meng, *et al*., 2013 Tet1 is critical for neuronal activity-regulated gene expression and memory extinction . Neuron **79**: 1109-1122.

Scheunemann, L., P. Skroblin, C. Hundsrucker, E. Klussmann, M. Efetova, *et al*., 2013 AKAPS act in a two-step mechanism of memory acquisition . The Journal of neuroscience : the official journal of the Society for Neuroscience **33**: 17422-17428.

Schlenke, T. A., J. Morales, S. Govind, and A. G. Clark, 2007 Contrasting infection strategies in generalist and specialist wasp parasitoids of Drosophila melanogaster . PLoS Pathog **3**: e158.

Schwaerzel, M., M. Monastirioti, H. Scholz, F. Friggi-Grelin, S. Birman, *et al*., 2003 Dopamine and octopamine differentiate between aversive and appetitive olfactory memories in Drosophila . The Journal of neuroscience : the official journal of the Society for Neuroscience **23**: 10495-10502.

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.

Tully, T., T. Preat, S. Boynton, and M. Del Vecchio, 1994 Genetic dissection of consolidated memory in Drosophila . Cell **79**: 35-47.

Tully, T., 1987 Drosophila learning and memory revisited . Trends in neurosciences **10**: 330-335.

Vargas, M. A., N. Luo, A. Yamaguchi, and P. Kapahi, 2010 A role for S6 kinase and serotonin in postmating dietary switch and balance of nutrients in D. melanogaster . Current Biology **20**: 1006-1011.

Walker, D. M., H. M. Cates, E. A. Heller, and E. J. Nestler, 2015 Regulation of chromatin states by drugs of abuse . Current opinion in neurobiology **30**: 112-121.

Wang, L., B. Charroux, S. Kerridge, and C. C. Tsai, 2008 Atrophin recruits HDAC1/2 and G9a to modify histone H3K9 and to determine cell fates . EMBO reports **9**: 555-562.

Wang, L., H. Rajan, J. L. Pitman, M. McKeown, and C. C. Tsai, 2006 Histone deacetylase-associating Atrophin proteins are nuclear receptor corepressors . Genes & development **20**: 525-530.

Wong, A. K., A. Krishnan, V. Yao, A. Tadych, and O. G. Troyanskaya, 2015 IMP 2.0: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks . Nucleic acids research **43**: W128-33.

Wong, A. K., C. Y. Park, C. S. Greene, L. A. Bongo, Y. Guan, *et al*., 2012 IMP: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks . Nucleic acids research **40**: W484-90.

Wood, M. A., M. P. Kaplan, A. Park, E. J. Blanchard, A. M. Oliveira, *et al*., 2005 Transgenic mice expressing a truncated form of CREB-binding protein (CBP) exhibit deficits in hippocampal synaptic plasticity and memory storage . Learning & memory (Cold Spring Harbor, N.Y.) **12**: 111-119.

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.

Yeung, K., A. Boija, E. Karlsson, P. H. Holmqvist, Y. Tsatskis, *et al*., 2017 Atrophin controls developmental signaling pathways via interactions with Trithorax-like . eLife **6**: 10.7554/eLife.23084.

Zhang, S., L. Xu, J. Lee, and T. Xu, 2002 Drosophila atrophin homolog functions as a transcriptional corepressor in multiple developmental processes . Cell **108**: 45-56.

Zhang, Z., J. Feng, C. Pan, X. Lv, W. Wu, *et al*., 2013 Atrophin-Rpd3 complex represses Hedgehog signaling by acting as a corepressor of CiR . The Journal of cell biology **203**: 575-583.

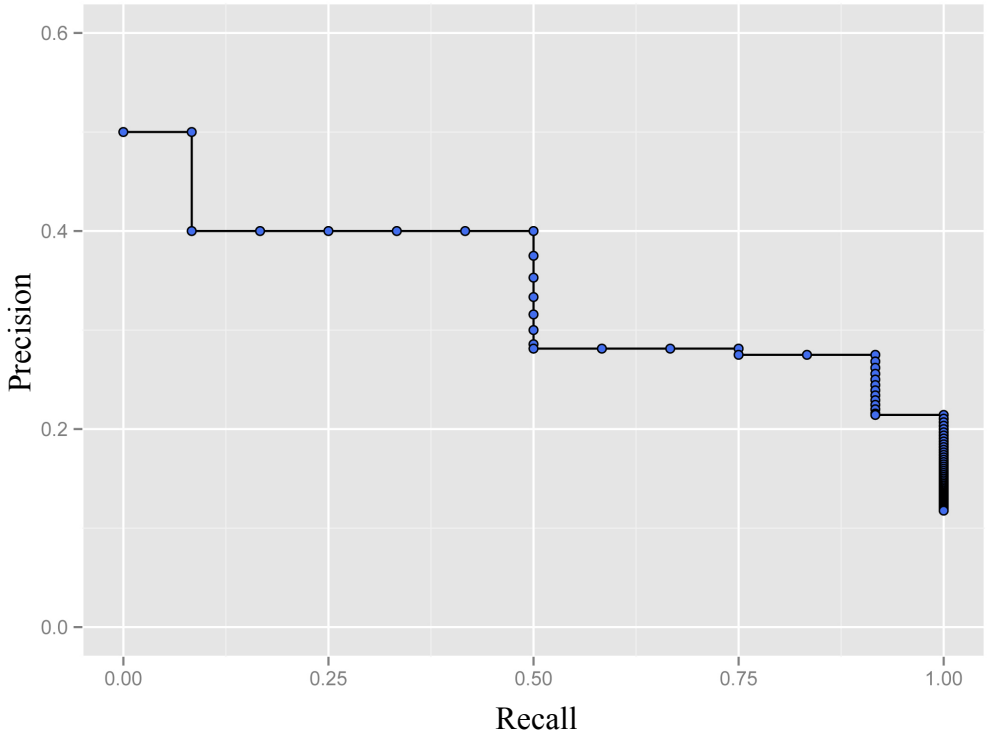
Zovkic, I. B., M. C. Guzman-Karlsson, and J. D. Sweatt, 2013 Epigenetic regulation of memory formation and maintenance . Learning & memory (Cold Spring Harbor, N.Y.) **20**: 61-74.

FIGURE LEGENDS

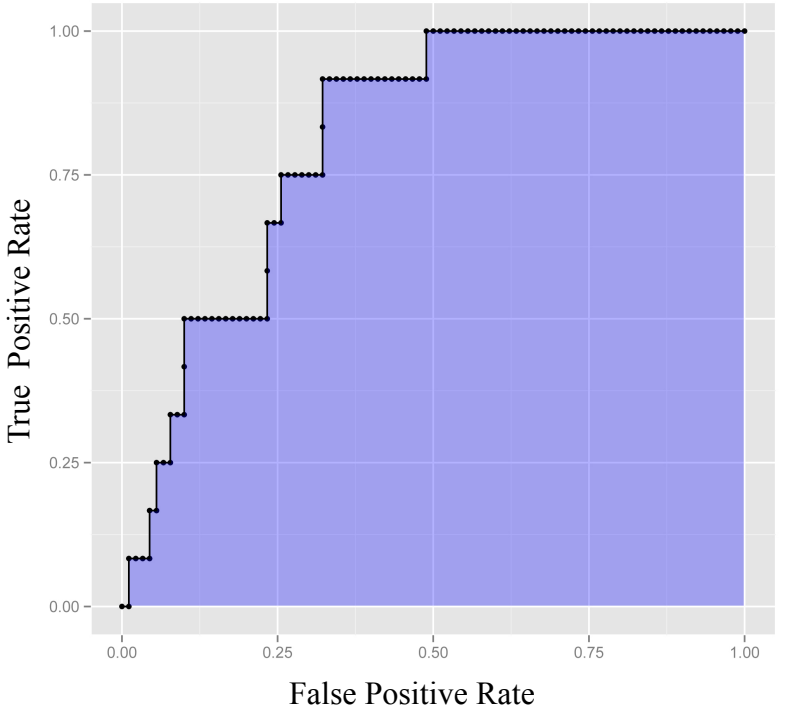
Figure 1. PIGLRM analysis reveals genes that may have a role in Drosophila memory, including the histone deacetylase, *Gug*.

Using known learning memory genes as the positive standard, we performed PILGRM analysis to identify a list of novel target genes that may be involved in memory. The analysis provides visual analysis of precision of predictions at different recall thresholds (A), visual analysis of true positive rate at various false positive rate thresholds (the area under the curve, shown as the shaded region, for this analysis is 0.8130) (B), and the top novel predictions based on the analysis (C).

A



B



C

Gene Name	Gene Score	Estimated Probability
Gug	0.599059	0.523282761
gem	0.583578	0.506416955
tws	0.580922	0.503520835
Taf4	0.571693	0.493456879
crol	0.571595	0.493350027
Sap47	0.570648	0.492317519
CG14478	0.569926	0.491530371
CG13624	0.562456	0.483389508
Sdc	0.560679	0.481454056
spen	0.559114	0.479749966
l(3)82Fd	0.557465	0.477954921
Atpalpha	0.555083	0.475362975
Pde8	0.55315	0.473260571
cora	0.551914	0.471916743
Csk	0.550481	0.47035924

Figure 2. IMP analysis indicates high network connectivity of *Gug* and known memory genes.

A visual representation of a gene network utilizing our positive standards from Figure 1 with gug shows a high degree of network connectivity. Lines indicate a minimum relationship confidence of 0.85.

A

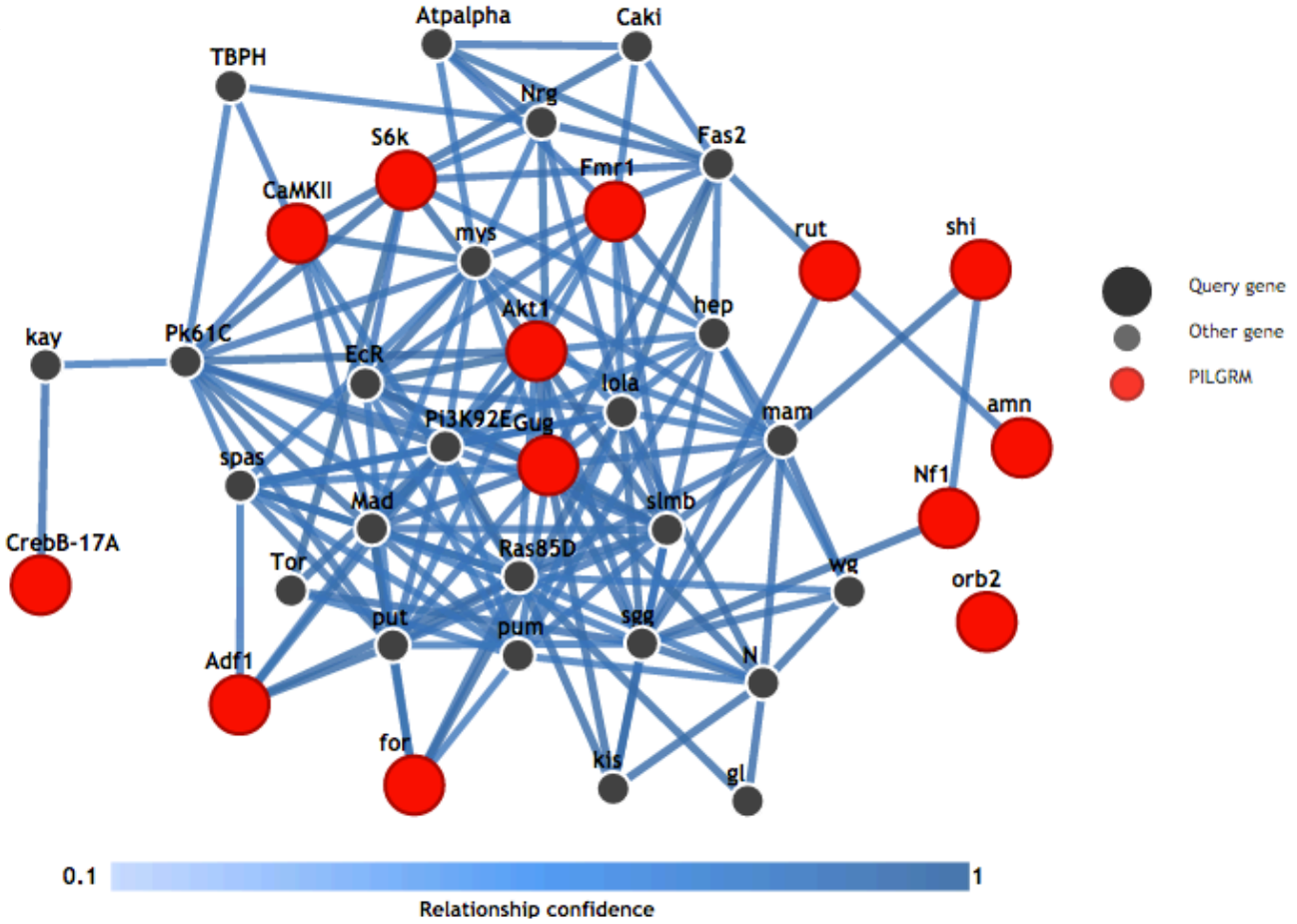
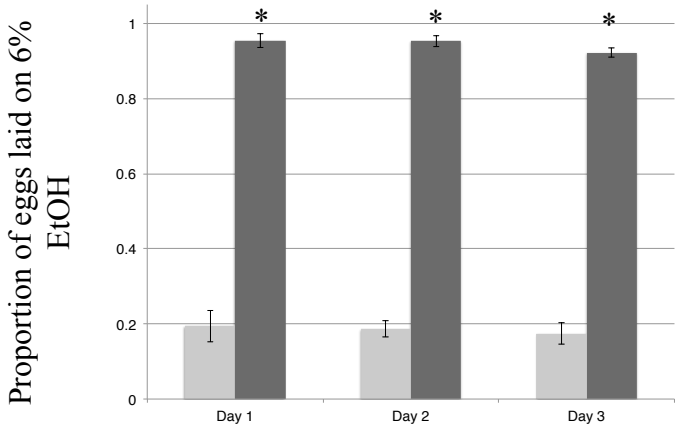

Atpalpha
Caki
TBPH
Nrg
S6k
Fas2
Fmr1
CaMKII
mys
rut
shi
Akt1
hep
kay
Pk61C
EcR
lola
mam
amn
Pi3K92E
Gug
spas
Nf1
Mad
slmb
CrebB-17A
Tor
Ras85D
wg
orb2
put
pum
sgg
Adf1
N
for
kis
gl
Query gene
Other gene
PILGRM
0.1
1
Relationship confidence

Figure 3. *Gug* is involved in memory retention using ethanol choice memory assay

Proportion of eggs laid on 6% ethanol oviposition cap following wasp exposure after three 24-hour time points is shown. Wild-type flies continue to oviposit on 6% ethanol after wasp exposure (A). *Gug*/+ heterozygotes (B) and flies expressing a Gug^{RNAi} hairpin in the MB show impaired memory retention (C-D). MB switch flies fed vehicle control show wild-type memory (E), while flies expressing a Gug^{RNAi} via RU486 feeding show impaired memory retention (F). Error bars represent 95% confidence intervals (n = 10 biological replicates) (*$p < 0.05$).

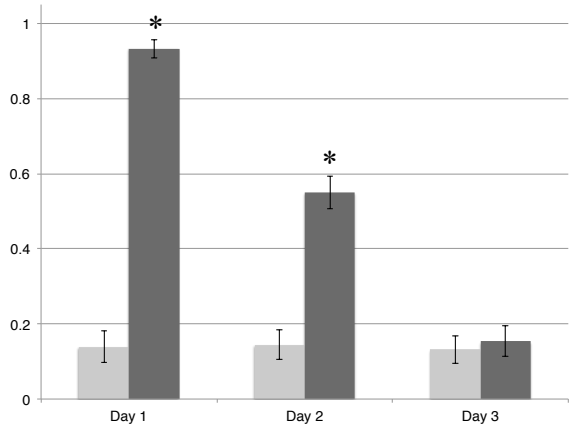
A

Wild-type flies have show memory retention



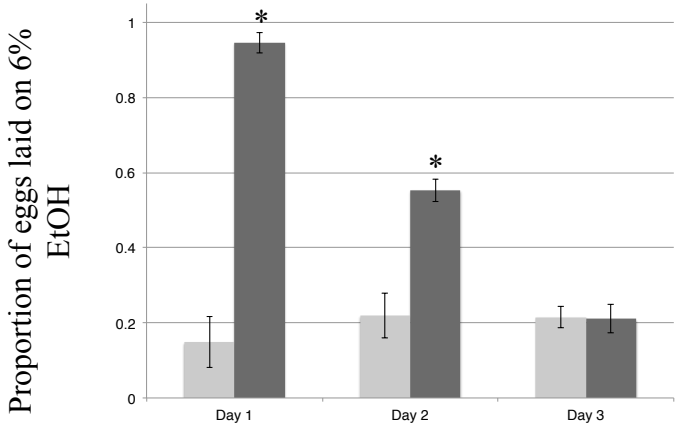
B

Gug/+ flies show accelerated memory decay



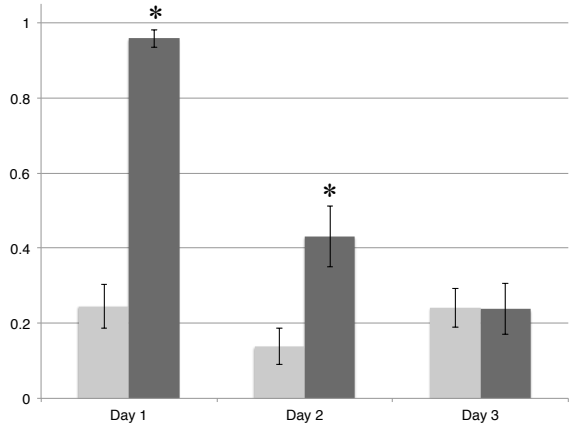
C

OK107;*Gug*^{RNAi} flies have impaired memory retention



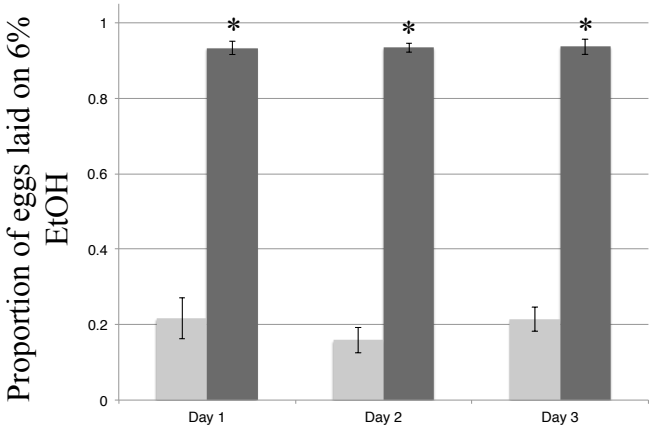
D

MB247;*Gug*^{RNAi} flies have impaired memory retention



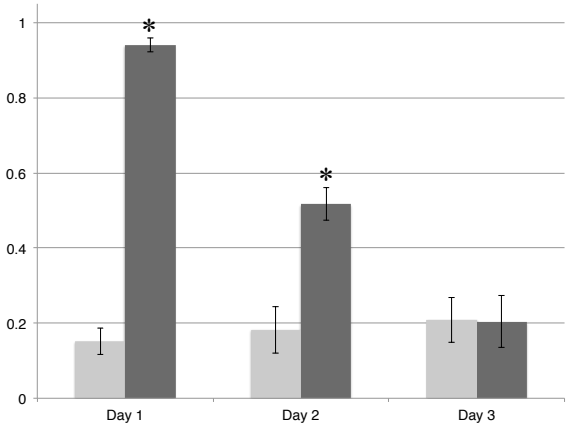
E

MBswitch;*Gug*^{RNAi} flies fed vehicle only show wild-type memory retention



F

MBswitch;*Gug*^{RNAi} flies fed RU486 only show wild-type memory retention



Unexposed Exposed

Figure 4. *Gug* is involved in memory retention using egg retention assay

Percentage of eggs laid by exposed flies normalized to eggs laid by unexposed flies is shown. Wild-type flies exposed to wasps lay fewer eggs than unexposed flies for multiple days (A). *Gug*/+ flies have impaired memory retention (B). MB switch flies fed vehicle control show wild-type memory (C), while flies expressing a Gug^{RNAi} via RU486 feeding show impaired memory retention (D). Error bars represent standard error (n = 24 biological replicates) (*$p < 0.05$).

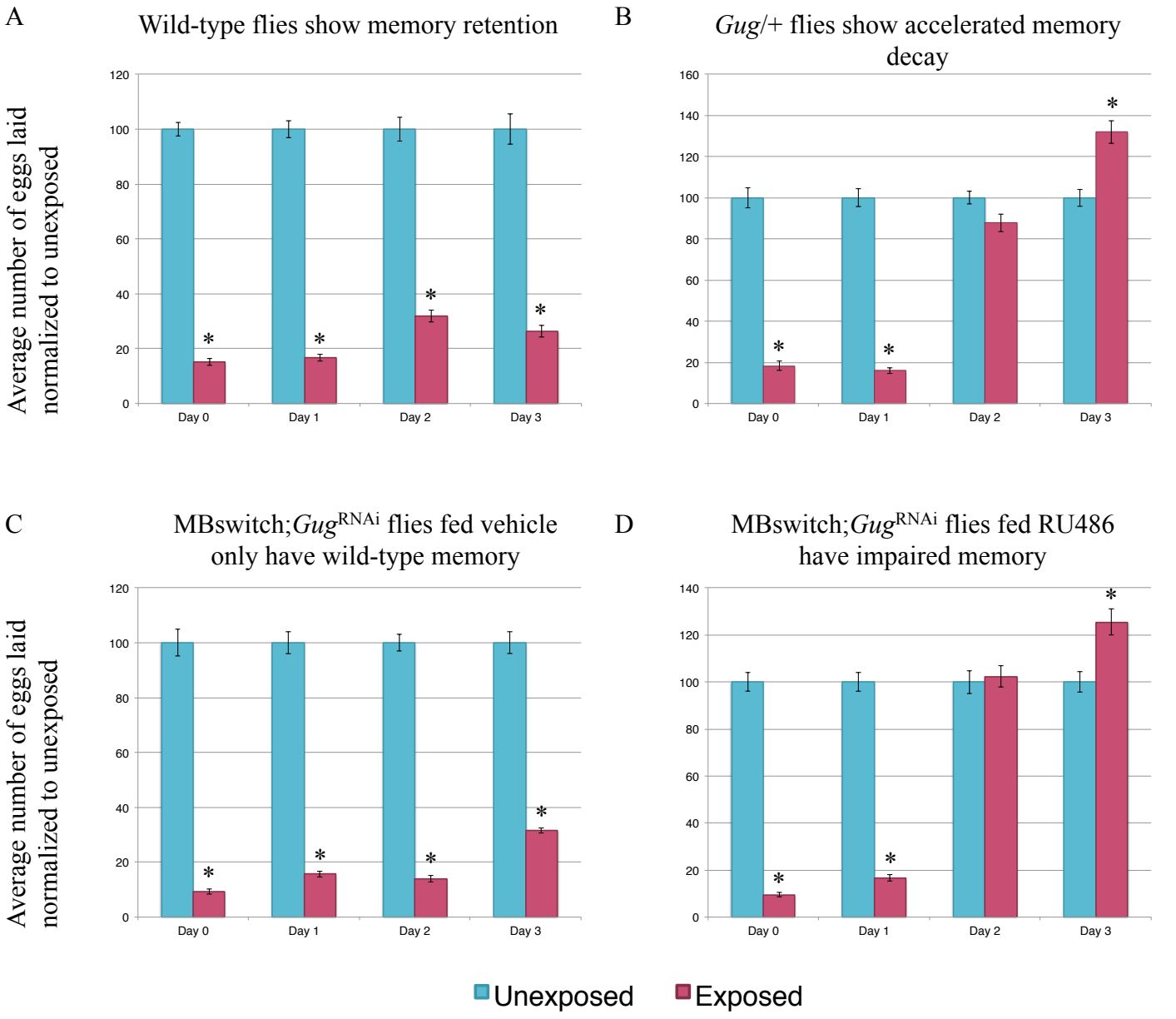
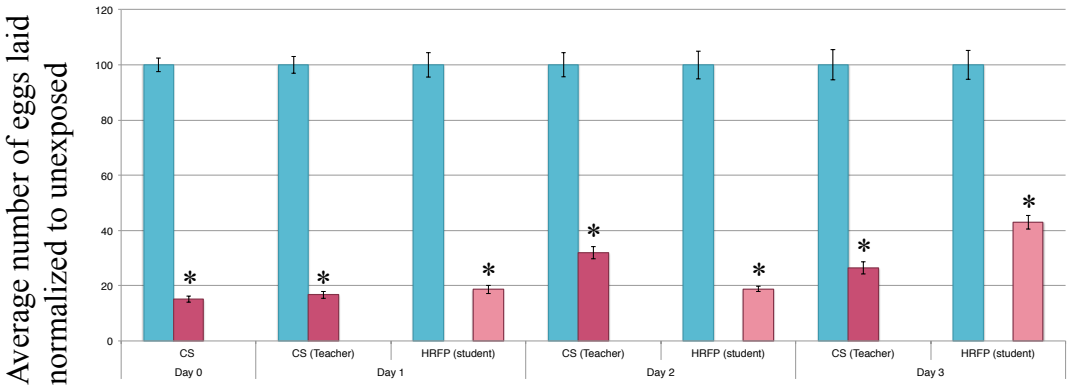

A
Wild-type flies show memory retention
B
Gug/+ flies show accelerated memory decay
C
MBswitch;GugRNAi flies fed vehicle only have wild-type memory
D
MBswitch;GugRNAi flies fed RU486 have impaired memory
Average number of eggs laid normalized to unexposed
Day 0
Day 1
Day 2
Day 3
Unexposed
Exposed

Figure 5. *Gug* is involved in teaching ability

Percentage of eggs laid by exposed flies normalized to eggs laid by unexposed flies is shown. Wild-type flies exposed to wasps can teach multiple student cohorts (RFP-Histone) across three days (A). *Gug*/+ flies have impaired teaching ability (B). MB switch flies fed vehicle control show wild-type teaching ability (C), while flies expressing a Gug^{RNAi} via RU486 feeding show impaired teaching ability (D). Error bars represent standard error (n = 24 biological replicates) (*$p < 0.05$).

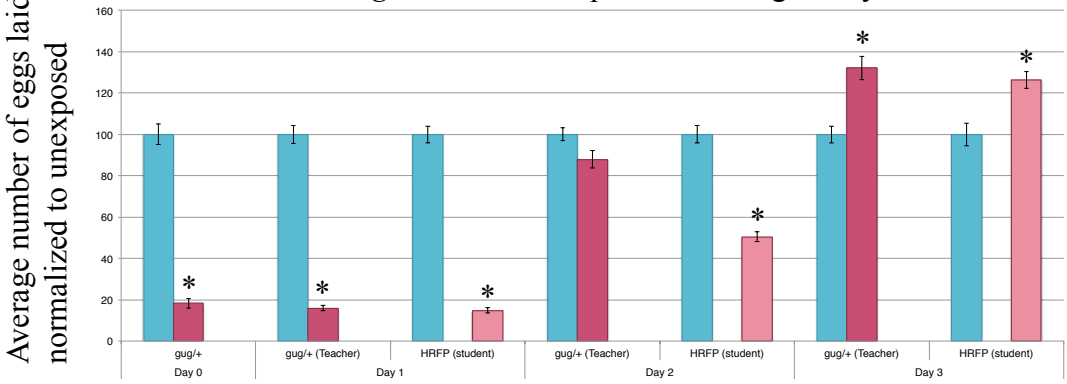
A

Wild-type flies are able to teach multiple cohorts of student flies



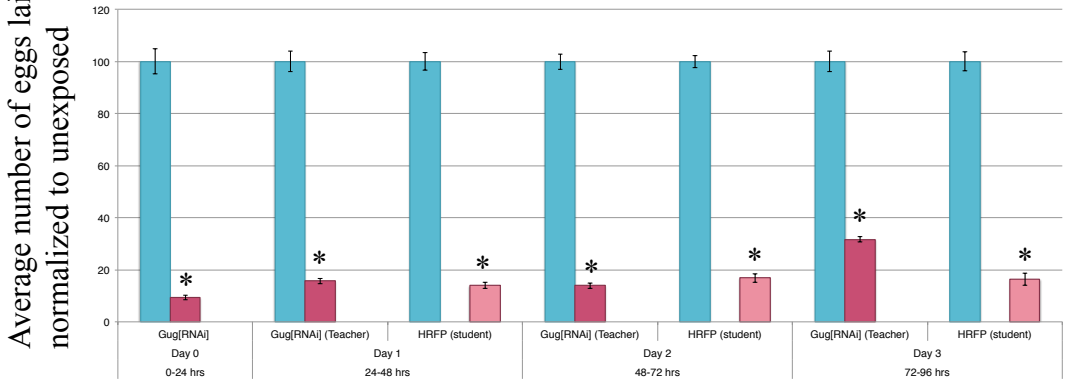
B

Gug/+ flies have impaired teaching ability



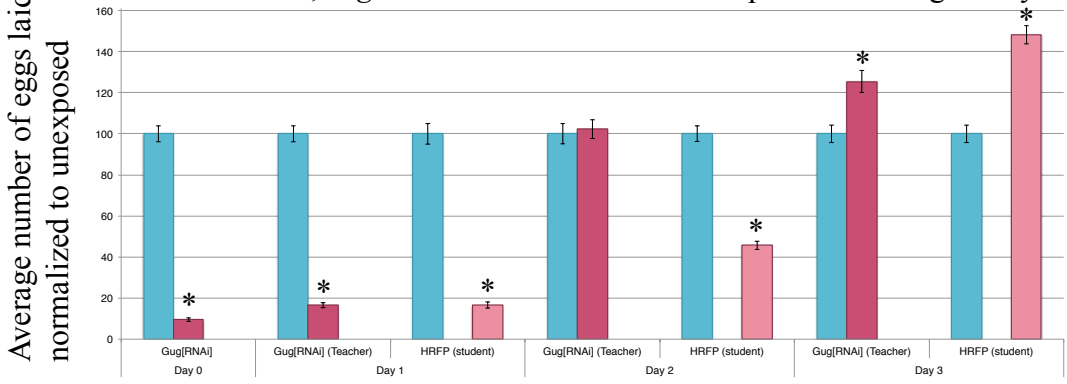
C

MBswitch;*Gug*^{RNAi} flies fed vehicle only have wild-type teaching ability



D

MBswitch;*Gug*^{RNAi} flies fed RU486 have impaired teaching ability



Unexposed

Exposed

Naïve

Figure 6. Gug has a role in social learning.

Percentage of eggs laid by exposed flies normalized to eggs laid by unexposed flies is shown. Wild-type flies (His-RFP) exposed to wasps can teach wild-type (CS) students (A). *Gug*/+ flies have impaired learning ability (B). MB switch flies fed vehicle control show wild-type learning ability (C), while flies expressing a Gug^{RNAi} via RU486 feeding show impaired learning ability (D). Error bars represent standard error (n = 24 biological replicates) (*$p < 0.05$).

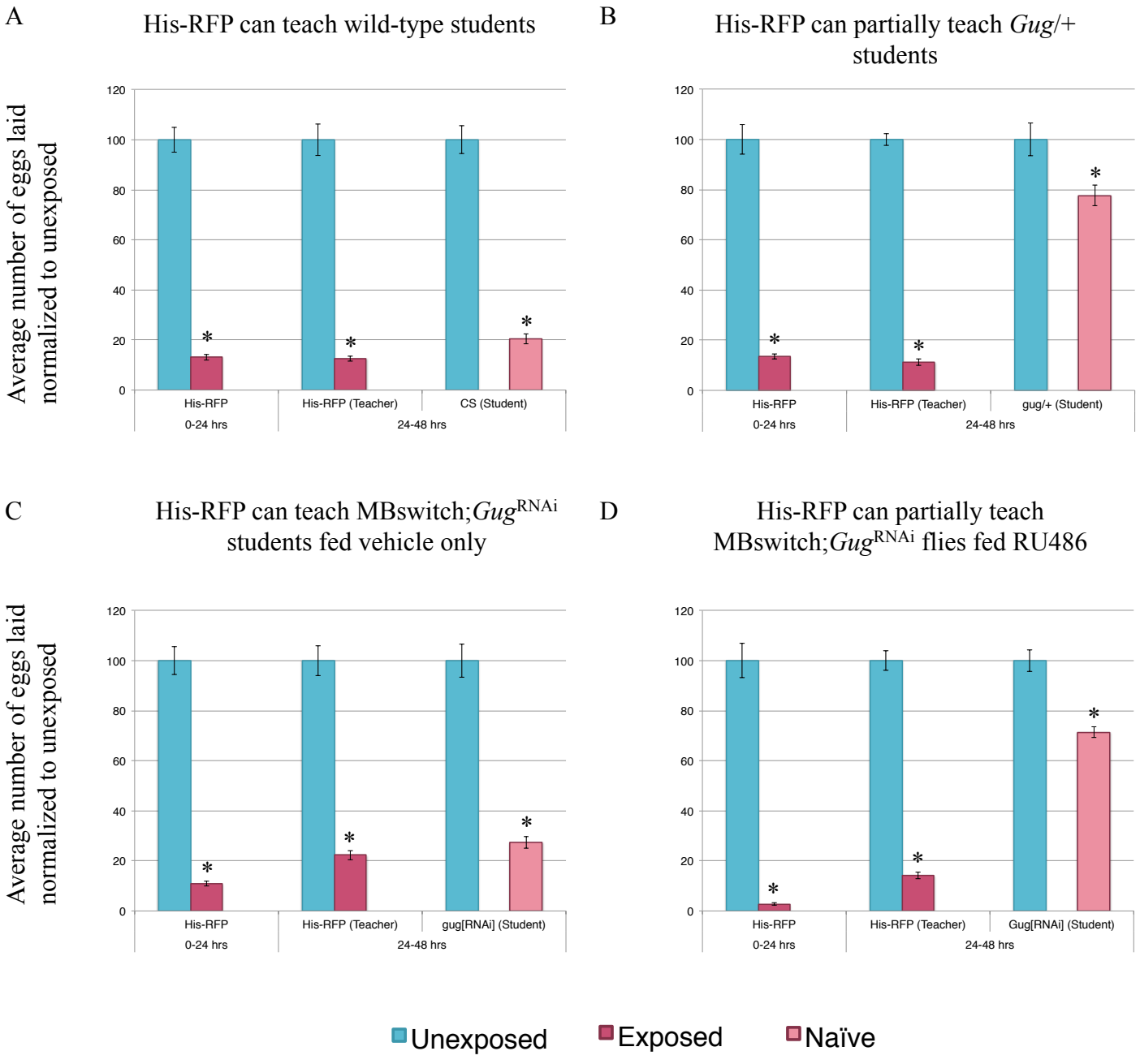
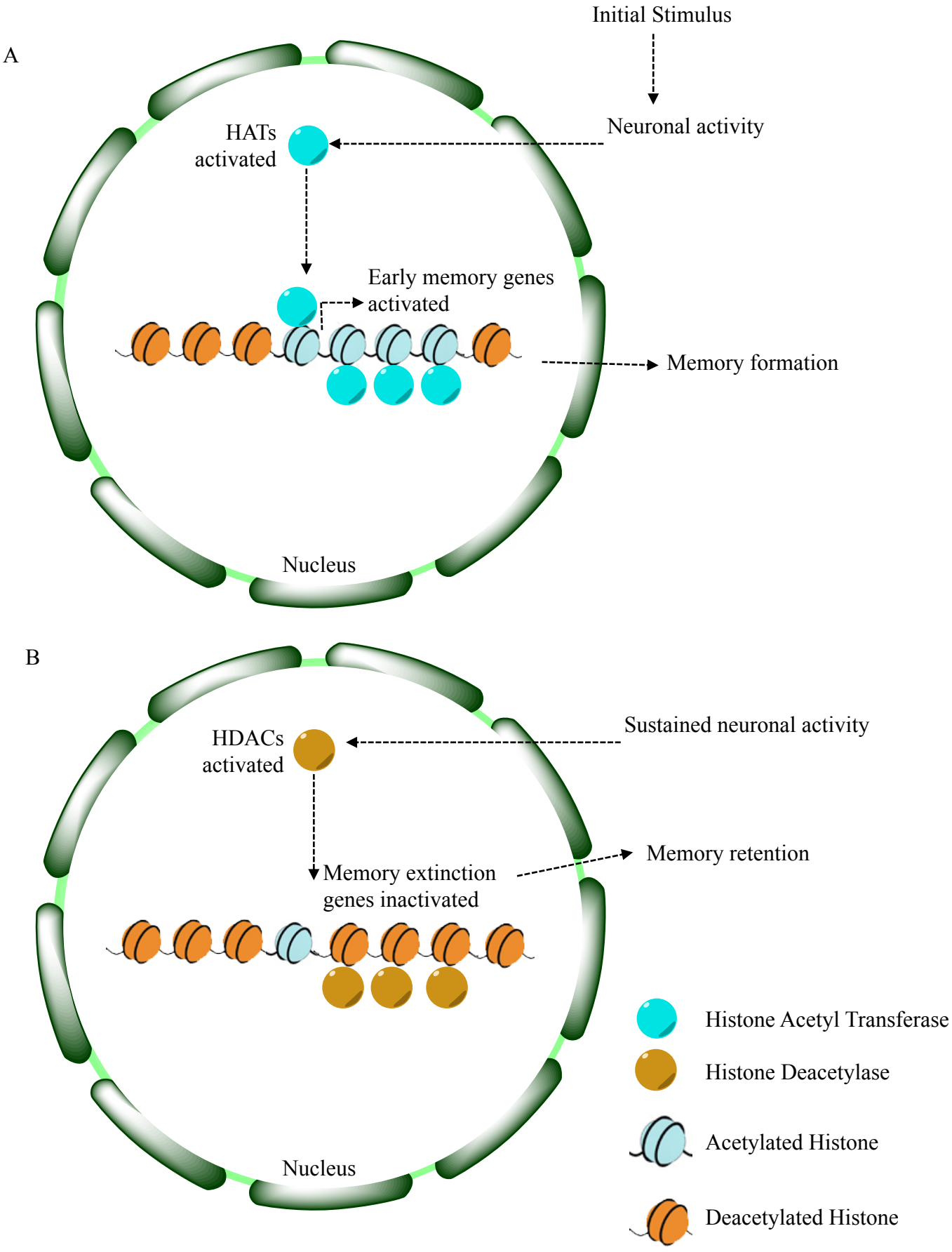

A
His-RFP can teach wild-type students
B
His-RFP can partially teach *Gug*/+ students
C
His-RFP can teach MBswitch;*Gug*^{RNAi} students fed vehicle only
D
His-RFP can partially teach MBswitch;*Gug*^{RNAi} flies fed RU486
Average number of eggs laid normalized to unexposed
120
100
80
60
40
20
0
His-RFP
His-RFP (Teacher)
CS (Student)
gug/+ (Student)
gug[RNAi] (Student)
Gug[RNAi] (Student)
0-24 hrs
24-48 hrs
Unexposed
Exposed
Naïve

Figure 7. Proposed model for histone deacetylation in memory retention.

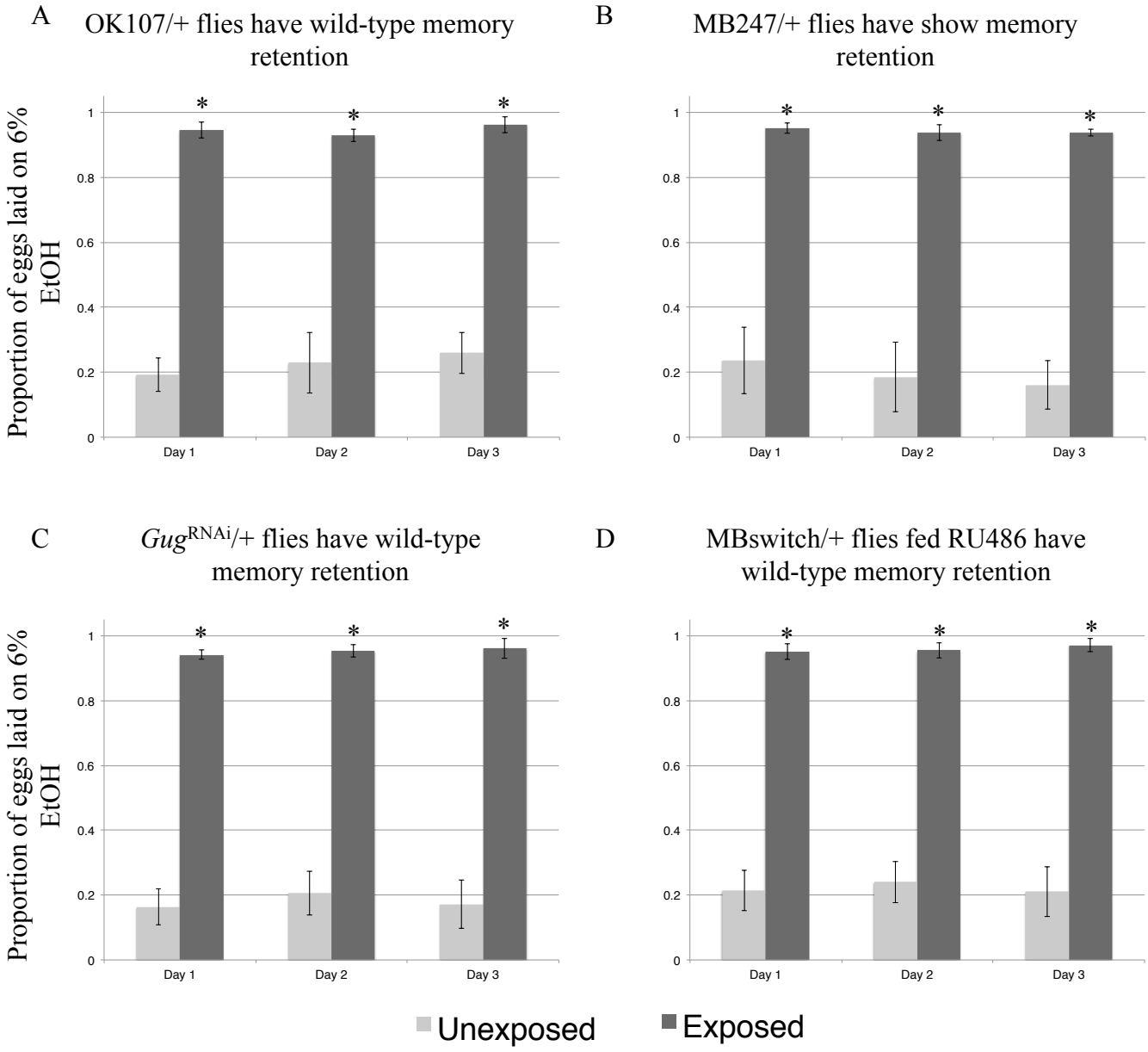
Model for learning and memory shows a role for *Gug*. At the start of memory formation, histone acetylases are upregulated following neuronal stimulation, promoting upregulation of transcription of memory related genes (A). Following memory consolidation, these gene sites on chromatin may become deacetylated, in addition to memory extinction genes, to promote memory retention (B).


A
Initial Stimulus
Neuronal activity
HATs
activated
Early memory genes
activated
Memory formation
Nucleus
B
HDACs
activated
Sustained neuronal activity
Memory retention
Memory extinction
genes inactivated
Nucleus
Histone Acetyl Transferase
Histone Deacetylase
Acetylated Histone
Deacetylated Histone

SUPPLEMENTARY FIGURE LEGENDS

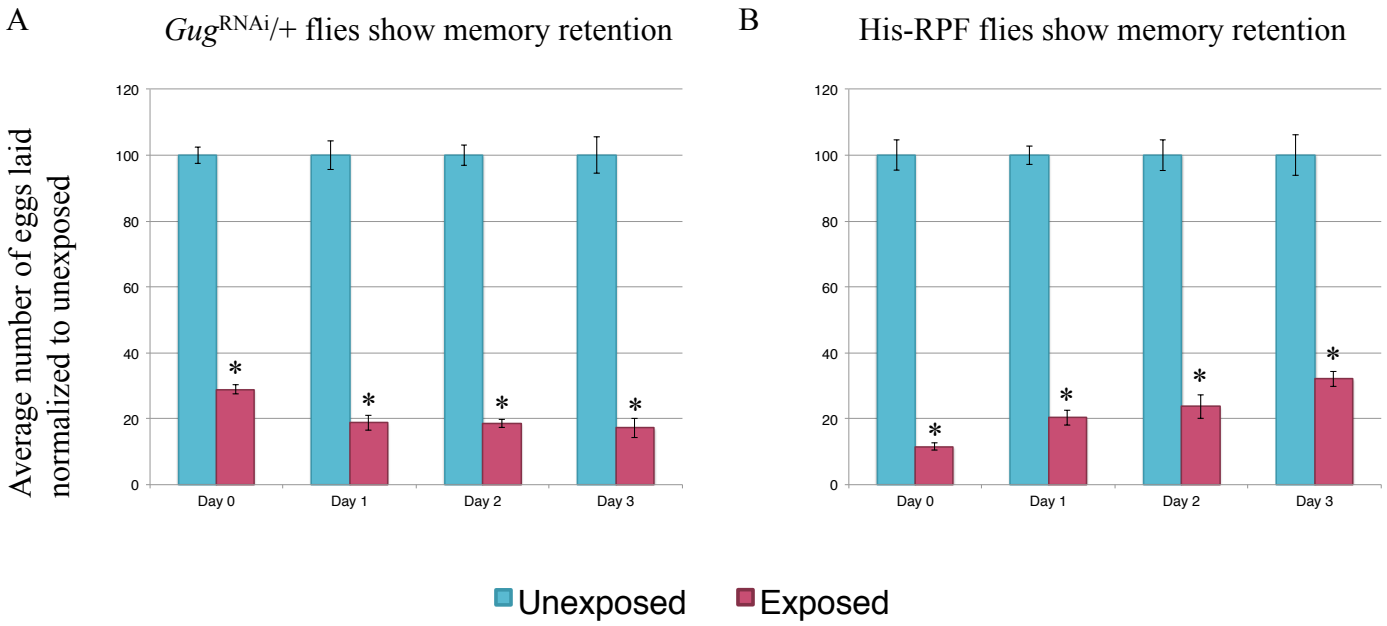
Supplementary Figure 1. Further evidence demonstrating the role of *Gug* in memory retention.

Proportion of eggs laid on 6% ethanol oviposition cap following wasp exposure after three 24-hour time points is shown. Controls outcrossed to Canton S continue to oviposit on 6% ethanol after wasp exposure. Outcrossed genotypes shown are OK107/+ (A), MB247/+ (B), Gug^{RNAi} (C), and MBswitch/+ fed RU486. Error bars represent 95% confidence intervals (n = 10 biological replicates) (*$p < 0.05$).


A
OK107/+ flies have wild-type memory retention
B
MB247/+ flies have show memory retention
C
Gug^RNAi/+ flies have wild-type memory retention
D
MBswitch/+ flies fed RU486 have wild-type memory retention
Proportion of eggs laid on 6% EtOH
0
0.2
0.4
0.6
0.8
1
Day 1
Day 2
Day 3
*
Unexposed
Exposed

Supplementary Figure 2. Further evidence demonstrating the role of *Gug* in egg-retention memory.

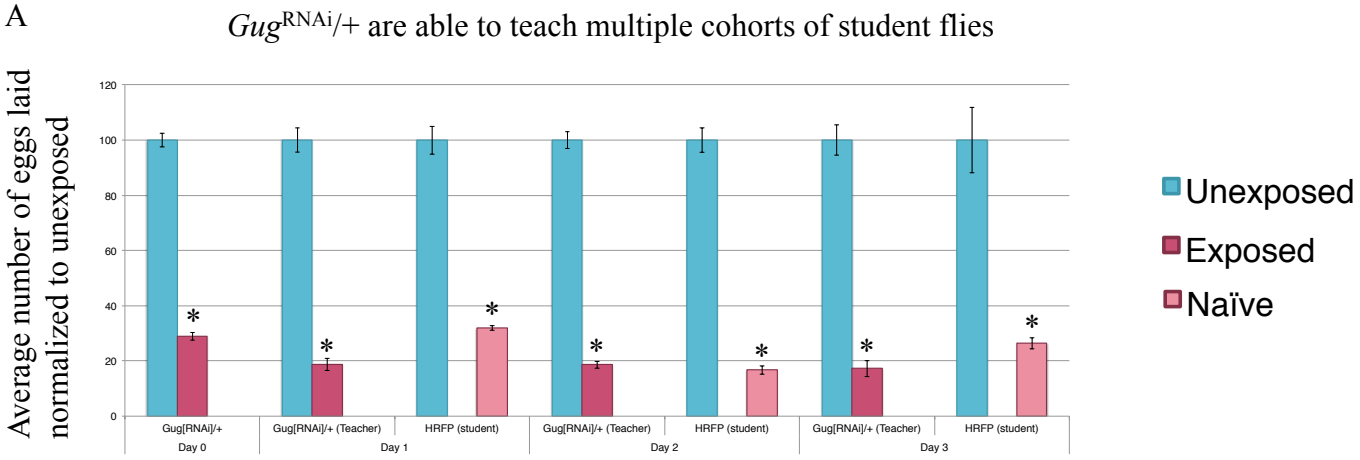
Percentage of eggs laid by exposed flies normalized to eggs laid by unexposed flies is shown. Wild-type flies exposed to wasps lay fewer eggs than unexposed flies for multiple days (A). Gug^{RNAi}/+ (A) and His-RFP flies have wild-type memory retention (B). Error bars represent standard error (n = 24 biological replicates) (*$p < 0.05$).


A
*Gug*RNAi/+ flies show memory retention
B
His-RPF flies show memory retention
Average number of eggs laid normalized to unexposed
120
100
80
60
40
20
0
Day 0
Day 1
Day 2
Day 3
Unexposed
Exposed

Supplementary Figure 3. Further evidence demonstrating the role of *Gug* in teaching behavior.

Percentage of eggs laid by exposed flies normalized to eggs laid by unexposed flies is shown. Gug^{RNAi}/+ (outcrossed to Canton S) flies exposed to wasps can teach multiple student cohorts (RFP-Histone) across three days (A). Error bars represent standard error (n = 24 biological replicates) (*$p < 0.05$).

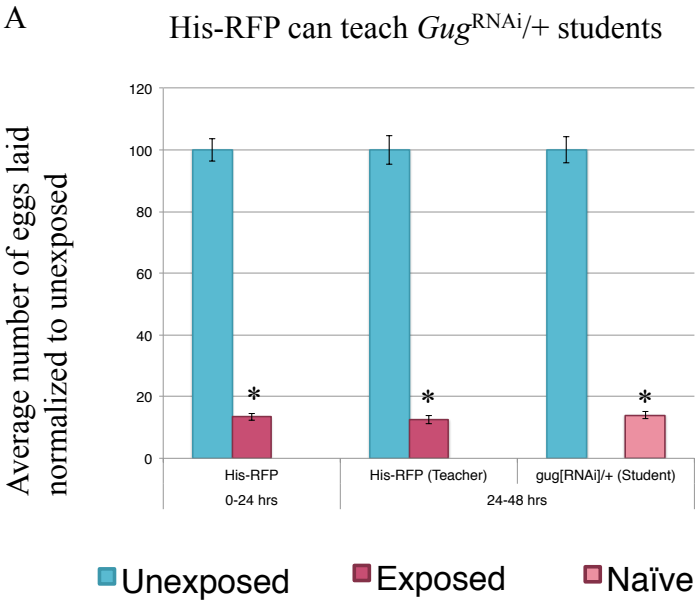
A


*Gug*RNAi/+ are able to teach multiple cohorts of student flies
Average number of eggs laid normalized to unexposed
120
100
80
60
40
20
0
Gug[RNAi]/+
Gug[RNAi]/+ (Teacher)
HRFP (student)
Gug[RNAi]/+ (Teacher)
HRFP (student)
Gug[RNAi]/+ (Teacher)
HRFP (student)
Day 0
Day 1
Day 2
Day 3
Unexposed
Exposed
Naïve

Supplementary Figure 4. Further evidence demonstrating the role of *Gug* in social learning.

Percentage of eggs laid by exposed flies normalized to eggs laid by unexposed flies is shown. Wild-type flies (His-RFP) exposed to wasps can teach Gug^{RNAi}/+ students (A). Error bars represent standard error (n = 24 biological replicates) (*$p < 0.05$).

A


His-RFP can teach *Gug*^{RNAi}/+ students
Average number of eggs laid normalized to unexposed
120
100
80
60
40
20
0
*
*
*
His-RFP
His-RFP (Teacher)
gug[RNAi]/+ (Student)
0-24 hrs
24-48 hrs
Unexposed
Exposed
Naïve

SUPPLEMENTARY FILES

Supplementary File 1. Gene list of positive and negative standards used in PILGRM analysis.

Genes in positive standard are indicated with a (1). Genes in negative standard indicated by a (-1).

Supplementary File 2. Novel genes as predicted by PILGRM.

Output gene list from PILGRM analysis.

MATERIALS AND METHODS

Insect Species/Strains

The *D. melanogaster* strains Canton-S (CS), and a Histone-RFP transgenic line, was used as the wild-type strain for oviposition preference after wasp exposure. The mushroom body Gene-Switch line (MB GeneSwitch) and the MB247 were kindly provided by Greg Roman (Baylor College of Medicine). Flies were maintained at room temperature with approximately 30% humidity. UAS-Gug^{RNAi}, Gug/TM3 (Spradling *et al.* 1999, 135-177) , and OK-107 were acquired from the Bloomington Drosophila Stock Center (stock numbers 32961, 11622, and 854, respectively). Flies aged 3-5 days were used for all experiments.

All species and strains used were maintained in fly bottles (Genesse catalog number 32-130) containing 50 mL of standard Drosophila media. Bottles were supplemented with 3 Kimwipes rolled together and placed into the center of the food.

The Figitid larval endoparasitoid *Leptopilina heterotoma* (strain Lh14 (Schlenke *et al.* 2007, e158)) was used in all experiments. In order to culture wasps, batches of 14 female and 5 male adult flies (*Drosophila melanogaster*, strain Canton S) were allowed to lay eggs in standard Drosophila vials containing standard Drosophila medium supplemented with activated yeast for 5 days. After 5 days, flies were removed and replaced with adult wasps (15 female, 6 male), which then attacked the developing fly larvae. Wasp vials were supplemented with approximately 500 μL of a 50% honey/water solution applied to the inside of the vial plugs. Wasps aged 3-7 days post-eclosion were used for all experiments.

PILGRM/IMP analysis

For the PILGRM analysis, we utilized the PILGRM version one data-mining algorithm (http://pilgrm.princeton.edu) (Greene and Troyanskaya 2011, W368-74) . The data set used in the analysis was derived from the Fruit Fly Compendium collected in August 2011. This data was collected from the Gene Expression Omnibus. The compendium used contained a total of 3,139 arrays from 186 different experiments. This data covers 12,864 Entrez gene identifiers. Within experiments, arrays were processed using the affy (Gautier *et al.* 2004, 307-315) bioconductor (Gentleman *et al.* 2004, R80) package. The gene-expression values were then normalized and summarized using the medianpolish method (Irizarry *et al.* 2003, 102-119) . The resulting experimental collections were then combined for learning using C++ Sleipnir library (Huttenhower *et al.* 2008, 1559-1561) .

Our gold standard was curated to include known learning and long-term memory genes. Our negative standards were selected by PILGRM in a randomized manner. Following the identification of gug, we input the gold standard and gug into IMP (http://imp.princeton.edu) (Wong *et al.* 2012, W484-90; Wong *et al.* 2015, W128-33) with a stringent minimum gene connection of 0.85 confidence.

Fly Oviposition Ethanol Choice

For fly memory assays, we used modified Petri dishes, termed Fly Corrals, as previously used and described (Kacsoh *et al.* 2015a, 1143-1157) . Holes are drilled into the Petri dish, where the center of the two holes is 6 cm apart. The diameter of each hole is ~1.2 cm. A nitex nylon mesh is melted onto the top of the dish in order to allow for ventilation. The mesh used has 120µm openings (Genesee Scientific catalog number 57-102). Dishes were cleaned using the Fisher Brand Sparkleen powder (catalog number 04-320-4) using a 10% Sparkleen solution.

Dishes are allowed to soak in the cleaning solution for at least 2 hours and subsequently rinsed in distilled water. Plates are then allowed to air dry for 24 hours. This cleaning protocol is followed both before and after an experiment was performed (Kacsoh *et al.* 2015a, 1143-1157) .

For preparation of the oviposition media in the Fly Corrals, we measured approximately 0.375 grams of instant blue Drosophila medium (Fisher Scientific Catalog number S22315C) into the caps of 15 mL Falcon Tubes (S22315C, Biological Resource Center, No.:22315C). For control (0% ethanol) food, we pipette 2250 µL of distilled water onto the instant food. For 6% ethanol food, we pipette 1966.5 µL of distilled water onto the instant food. Subsequent to this, we pipette 141.75 µL of 95% ethanol (190 Proof (95%), USP/NF/FCC/EP/BP/JP) onto the food. After mixing the liquid and the food, we immediately place one cap containing ethanol and one cap containing no ethanol onto the cage with lab tape (VWR) (Kacsoh *et al.* 2015a, 1143-1157) .

For memory assays utilizing the fly corral, 50 female flies and 10 male flies were co-incubated with 50 female Lh14 wasps for 24 hours in 2.25 cm diameter vials or sham exposed (control). Flies and wasps are then separated and flies are placed into fly corrals after the 24-hour exposure period with 5 females and 1 male fly placed per dish. Following a 24-hour period, caps are removed and replaced with freshly prepared caps (prepared in the exact manner the original caps are made). This process is repeated for 3 days following wasp exposure. Once caps are removed, the number of eggs on both the ethanol cap and the control cap are counted. Ten replicates are performed. All egg counts are blinded such that the counter is unaware of experimental condition and genotype (Kacsoh *et al.* 2015a, 1143-1157) .

Fly Oviposition Depression

We measured fly oviposition rates were using The Fly Condo™ (Genesee Scientific Cat # 59-110), as previously described (Kacsoh *et al.* 2015b, 10.7554/eLife.07423) . The Fly Condo contains 24 independent chambers, where each chamber is 7.5cm long by 1.5cm diameter. Each condo has a 24-well food plate into which we dispensed 2mL of standard Drosophila media. In order to assay egg retention of flies in the presence of wasps (acute exposure), we place 5 female flies and 1 male fly into one chamber of the Fly Condo. For exposed units, 3 female Lh14 wasps are also placed in the units. The oviposition plate from control and experimental condos are counted 24 hours later. To assay memory of wasp exposure, after the 24-hour exposure, we remove all wasps and transfer all flies to a new Fly Condo. The oviposition food plate is replaced with freshly poured food (2mL). We repeat this process for three days following the wasp exposure and egg counts are performed every 24-hours. All egg counts are blinded such that the counter is unaware of experimental condition.

Social communication and learning was tested as previously described (Kacsoh *et al.* 2015b, 10.7554/eLife.07423) . In order to assay fly-fly communication and the social learning period, 5 female flies and 1 male fly are placed into one chamber of The Fly Condo in the control, while 3 female Lh14 wasps are placed with the flies in the experimental setting for 24 hours. After the 24-hour wasp or sham (control) exposure, wasps are removed and replaced with 3 female "student" flies. These are naïve flies, never having seen a wasp. Flies are placed into new, clean fly condos for the second 24-hour period. For experiments with His-RFP teachers, the experiment terminates after one batch of students. For experiments using His-RFP as students, we replace the student flies with new batches of 3 female naïve His-RFP flies. This is repeated for 3 24-hour periods following the wasp exposure. This experiment measures teaching ability of flies. The oviposition plates contain 2mL of standard Drosophila media, replaced every 24-hours.

Fly embryo counts from each plate are made at each 24-hour time points. All egg counts are blinded such that the counter is unaware of experimental condition and genotype.

All treatments are run at 25ºC, at 40% humidity, with a 12:12 light:dark cycle in twenty-four replicates. Fly condos and oviposition plates are soaked thoroughly with 10% Sparkleen (Fisher catalog number 04-320-4) solution and rinsed with distilled water after every use and allowed to air-dry (Kacsoh *et al.* 2015b, 10.7554/eLife.07423) .

RU486 feeding

We perform RU486 feedings in both assays used, as previously described (Kacsoh *et al.* 2015a, 1143-1157; Kacsoh *et al.* 2015b, 10.7554/eLife.07423) . The method in which the RU486 is administered is similar in each behavior assay, but warrants two detailed methods. RU486 (Mifepristone) is used from Sigma (Lot number SLBG0210V).

For the ethanol seeking memory assay, we apply RU486 or vehicle only directly into the media. The caps for the fly corrals are made in the exact same manner as described above, but instead of using distilled water, an RU486 solution is used. The RU486 solution is prepared by dissolving 3.575 mg of RU486 in 800µL methanol (Fisher Scientific, Lot number 141313). This solution is added to 15.2 mL of distilled water. The total solution (16 mL) is thoroughly mixed and pipetted onto control food caps. For caps containing ethanol, the total solution of RU486 is changed to 14.992 mL and then pipetted onto the instant food. After pipetting the RU486 solution, we pipette 1.008 mL of 95% ethanol directly onto the food. Fresh RU486 solutions are prepared daily for experiments (Kacsoh *et al.* 2015a, 1143-1157) .

For the oviposition depression, teaching and social learning assays involving inducible knock-down, we prepare the fly condos by measuring 0.375 grams of flaky instant blue

Drosophila medium (Fisher Scientific Catalog number S22315C) into each well of The Fly Condo™ plates. For all food treatments, a total liquid volume of 2250 µL is directly pipetted onto the instant food. For experiments with RU486, an RU486 solution is used that is prepared by dissolving 3.575 mg of RU486 in 800µL methanol (Fisher Scientific, Lot number 141313). This solution is added to 15.2 mL of distilled water. The total solution (16 mL) is thoroughly mixed. From this mixed solution, we pipette 2250 µL onto the instant food. For plates containing no RU486 (vehicle, methanol only) 800µL methanol is mixed with 15.2 mL of distilled water. The total solution (16 mL) is thoroughly mixed. From this mixed solution, we pipette 2250 µL onto the instant food. RU486 solutions are prepped daily for experiments (Kacsoh *et al.* 2015b, 10.7554/eLife.07423) .

Statistical analysis

Statistical tests for behavioral assays are performed in Microsoft Excel. We use Welch's two-tailed t-tests for all egg count data. P-values are calculated for comparisons between paired treatment-group and unexposed (Kacsoh *et al.* 2015b, 10.7554/eLife.07423; Kacsoh *et al.* 2015a, 1143-1157) .