

Electronic Thesis and Dissertation Repository

11-9-2017 11:00 AM

Gene Expression Changes in the Mushroom Body of *Drosophila melanogaster* During a Time Course of Long-Term Memory Formation and Maintenance

Spencer G. Jones, *The University of Western Ontario*

Supervisor: Kramer, Jamie M, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Biology

© Spencer G. Jones 2017

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Genetics Commons](#), [Genomics Commons](#), and the [Molecular Genetics Commons](#)

Recommended Citation

Jones, Spencer G., "Gene Expression Changes in the Mushroom Body of *Drosophila melanogaster* During a Time Course of Long-Term Memory Formation and Maintenance" (2017). *Electronic Thesis and Dissertation Repository*. 5045.
<https://ir.lib.uwo.ca/etd/5045>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

Abstract

Long-term memory (LTM) requires gene transcription. However, there is still much to learn about which genes are transcriptionally regulated during LTM and the biological roles they play. Here, gene expression changes were characterized in *Drosophila melanogaster* over a time course of LTM formation and maintenance in neurons of the mushroom body (MB), a structure required for normal learning and memory. I identified 120 genes differentially expressed ($q < 0.2$, fold change > 1.3) 24h after LTM induction. Among these were 13 potential downstream targets for RNA localization by the known memory genes *pumilo*, *staufer* and *oskar*, several genes encoding chromatin regulators and seven genes with cAMP response elements (CRE) that may be regulated by cAMP response element binding (CREB)-mediated transcription. Taken together, the results of this study provide a rich data-set of transcriptionally-regulated LTM candidate genes for further study.

Keywords:

Long Term Memory Formation, Long Term Memory Maintenance, RNA-sequencing, Transcriptome Analysis, CREB, Mushroom Body, *Drosophila melanogaster*, INTACT

Co-Authorship Statement

All experiments in this thesis were performed by myself. The bioinformatics pipeline for the analysis of RNA-sequencing data was designed and optimized in co-ordination with Kevin Nixon. Kartik Pradeepan assisted with fly collection for genetic experiments. All experiments were designed, funded and supervised by Dr. Jamie Kramer.

Acknowledgments

There are many people that I am extremely grateful to have worked with during the duration of this project.

First and foremost, I would like to thank my supervisor, Dr. Jamie Kramer, for his expertise, guidance and support throughout the entirety of my M.Sc. experience. Without your encouragement and dedication, this work would definitely not be possible. I genuinely believe that with you at the helm, the Kramer Lab is certain to consistently generate high quality work.

I would like to thank Kevin Nixon for his invaluable contributions, patience and support. I have thoroughly enjoyed working together over the past two years, even when our protocols stop working for reasons unknown, and I look forward to working together again in future projects.

I would like to extend thanks to all members of the Kramer Lab, past and present, for creating an incredible working environment. I would especially like to thank Melissa Chubak, Max Stone, Kevin Nixon and Taylor Lyons. Without your support and humour this experience would not been nearly as enjoyable.

I would like to thank my volunteer, Kartik Pradeepan, for his versatility and commitment to helping me complete this project. I am grateful for your contributions and am very lucky to have had you work with me.

I would like to thank my advisory committee members, Dr. Brent Sinclair and Dr. Nathalie Berube, for their feedback throughout my degree. A special thank you to Dr. Sinclair for his comments on my thesis- they vastly improved the quality of this work.

I would also like to thank David Carter and The London Regional Genomics Centre for providing the valuable sequencing services required to bring this thesis to fruition.

Finally, I would like to give a special thank you to my parents Nancy and Ray Jones, as well as my grandparents, Doreen and Bruce Gordon, for their endless support. Without the scientific curiosity you instilled in me, I certainly would not have pursued this field of study.

Table of Contents

Abstract.....	i
Co-Authorship Statement.....	ii
Acknowledgments.....	iii
Table of Contents.....	iv
List of Figures.....	vi
List of Tables.....	vii
List of Appendices.....	viii
List of Abbreviations.....	ix
Chapter 1.....	1
1 Introduction.....	1
1.1 The molecular pathways of memory.....	1
1.2 The cellular correlates of memory.....	4
1.3 <i>Drosophila melanogaster</i> as a model organism for learning and memory processes...5	5
1.4 The mushroom body.....	7
1.5 Courtship conditioning as a learning paradigm for <i>D. melanogaster</i>	9
1.6 Transcriptome analyses of LTM of <i>D. melanogaster</i>	11
1.7 Study objective.....	14
Chapter 2.....	15
2 Methods.....	15
2.1 Flystocks.....	15
2.2 LTM induction using courtship conditioning.....	15
2.3 Isolation of nuclei tagged in a specific cell-type (INTACT).....	17
2.4 RNA isolation and RNA-sequencing sample preparation.....	18
2.5 INTACT validation by qPCR.....	18

2.6 RNA-seq data analysis.....	20
2.7 GO and motif enrichment analysis.....	21
Chapter 3.....	22
3 Results.....	22
3.1 MB-UNC84 males show normal LTM.....	22
3.2 Validation of MB-enrichment with INTACT.....	24
3.3 Analysis and quality control of MB-UNC84 RNA-seq data.....	25
3.4 DE analysis reveals a list of candidate genes differentially expressed 24h post-training	32
3.5 GO analysis of DE genes reveals terms enriched for learning and memory processes.....	34
3.6 Identification of known and <i>de novo</i> motifs within the promoters of DE genes.....	37
Chapter 4.....	39
4 Discussion.....	39
4.1 Genes with greater transcript abundance 24h post-training.....	39
4.2 Genes with lower transcript abundance 24h post-training.....	43
4.3 Enriched known and <i>de novo</i> motifs within the promoter regions of DE genes.....	43
4.4 Limitations.....	45
4.5 Genetic tools for further study.....	46
4.6 Summary and conclusions.....	48
References.....	49
Appendices.....	59
Curriculum Vitae.....	66

List of Figures

Figure 1.1: The Canonical Molecular Pathway for Memory Formation.....	3
Figure 1.2: Structure of the mushroom body (MB) and transmission of an environmental input through intrinsic MB neurons.....	8
Figure 1.3: Isolation of Nuclei in a Specific Cell Type (INTACT).....	13
Figure 2.1: Schematic showing approach used for sample collection and LTM validation.....	16
Figure 3.1: Long term courtship memory is intact in MB-UNC84 flies.....	23
Figure 3.2: qPCR Validates MB-Specificity of INTACT.....	24
Figure 3.3: Alignment efficiency and association of reads with genomic features for INTACT-obtained RNA-sequencing results.....	29
Figure 3.4: Sequencing data for RNA isolated from INTACT-obtained nuclei shows MB-enrichment	30
Figure 3.5: Volcano plots displaying genes identified as differentially expressed between experimental conditions.....	33
Figure 3.6: GO results for upregulated and downregulated DE genes	35

List of Tables

Table 2.1: Primers used to validate MB-Specificity of INTACT.....	19
Table 3.1: Raw read distribution of RNA-sequencing data.....	27
Table 3.2: Raw count distribution of RNA-sequencing data.....	28
Table 3.3: Consistency of MB-enrichment between samples.....	31
Table 3.4: Genes contributing the greatest source of variance for the first two principal components.....	31
Table 3.5: DE gene results reveal a list of genes previously associated with learning and memory.....	36
Table 3.6: HOMER identifies CRE motifs within promoter regions of DE genes.....	37
Table 3.7: HOMER identifies <i>de novo</i> motifs within promoter regions of DE genes.....	38
Table 4.1: Genes with known physical interactions to pumilo, oskar and staufer within the upregulated DE gene list.....	42

List of Appendices

Appendix A: Supplementary Tables	59-65
---	-------

Abbreviations

AC	Adenylyl cyclase
ATP	Adenosine triphosphate
cAMP	Cyclic adenosine monophosphate
CRE	cAMP response element
CREB	cAMP response element-binding protein
CS	Conditioned stimuli
CVA	11-cis-Vaccenyl acetate
DE	Differential expression
DNA	Deoxyribonucleic acid
GFP	Green fluorescent protein
GO	Gene ontology
gPCR	G protein-coupled receptors
HOMER	Hypergeometric optimization of motif enrichment
INTACT	Isolation of nuclei tagged in specific cell-type
KC	Kenyon cells
LTM	Long-term memory
LTP	Long-term potentiation
MB	Mushroom body
PDE	Phosphodiesterase
PKA	Protein kinase A
RNA	Ribonucleic acid
STM	Short-term memory
TSS	Transcriptional start site
qPCR	Real-time polymerase chain reaction
UAS	Upstream activating sequence
US	Unconditioned stimuli

Chapter 1: Introduction

Learning and memory can be defined as the creation, storage and recall of an altered behavioural response produced by an environmental input (Sweatt, 2010). Generally speaking, the processes behind learning and memory can be subdivided into three distinct phases (Tully, 2003; Hawkins et al., 2006). Acquisition, the process of learning, is the perception of a new experience. From this, a short-term memory (STM) is formed, which is malleable and transient. In the appropriate conditions, often due to repetition of the input, this experience may be consolidated and a long-term memory (LTM) formed. While organisms may have subtle differences in how they process learning and memory, often these phases are conserved. Study of these different phases can give insight into the cellular and molecular mechanisms of learning and memory.

1.1 The molecular pathways of memory

Simple forms of learning can be divided into two broad categories, associative and non-associative, which, while utilizing different procedures, induce learning and memory by similar biological processes (Lau et al., 2013). Non-associative forms of learning, like habituation, utilize repeated exposure to a single stimulus to produce a decrease in behavioural response (Groves & Thompson, 1970). Contrarily, associative learning requires input from two environmental signals to modify behaviour. Classical Pavlovian conditioning is a form of associative learning that pairs a biologically neutral stimulus, termed the conditioned stimulus (CS), with a stimulus which elicits an involuntary biological response, known as the unconditioned stimulus (US) (Domjan, 2005). Often the CS is a sensory input, usually a smell or visual cue, whereas the US involves reward or punishment, commonly through the provision of food or an aversive shock. Through the CS/US pairing and with sufficient training, the CS becomes associated with the innate response of the US, being able to produce the same biological response when presented alone. This CS/US pairing forms basis of many associative learning and memory paradigms.

At a molecular level, the CS/US pairing of associative learning converge to activate the 3', 5' cyclic adenosine monophosphate (cAMP) pathway (**Figure 1.1**). The cAMP signalling cascade has been consistently shown to be implicated in the different phases of learning and memory (Rall et al., 1956). The importance cAMP has in memory processes was initially realized through studies on the sea slug *Aplysia* (Brunelli, 1976). Through manipulation of a natural gill withdrawal reflex

in response to an electric shock, it was determined that cAMP was the key secondary molecule involved in the formation of the observed adapted behaviour. In parallel, single-gene mutants in *Drosophila melanogaster* further emphasized the importance of cAMP. In flies, mutants of the cAMP-generating *rutabaga* and the cAMP-inhibiting *dunce* have been shown to be required for both STM and LTM formation (Dudai et al., 1976; Livingstone et al., 1984; Blum et al., 2009). Initial activation of the cAMP pathway occurs when g-protein coupled receptors (gPCR), stimulated by US, activate a family of enzymes called adenylyl cyclases (AC) which function to catalyse the conversion of adenosine triphosphate (ATP) to cAMP (**Figure 1.1**). AC has not only been shown to be responsive to both gPCR activation but also to the influx of Ca^{2+} or its downstream effectors like calmodulin (Levin et al., 1992). The initial influx of calcium is attributed to the modulation of NMDA, cholinergic or GABAergic receptors and is a result of CS stimulation. This cross-talk between two distinct molecular pathways indicates that AC is the point of biological convergence between the CS/US pairing and acts as a molecular coincidence detector during associative learning, acting synergistically to increase cAMP levels (Tomchik & Davis, 2009).

Downstream of cAMP, the molecular pathways differ between types of memory (**Figure 1.2**). Protein kinase A (PKA) is a tetrameric enzyme consisting of two regulatory and two catalytic subunits. PKA is regulated by cAMP. In the absence of cAMP, PKA is incapable of kinase activity as the regulatory and catalytic subunits are bound together. However, when cAMP levels increase, these subunits do not bind and catalytic PKA is capable of phosphorylating downstream elements of cAMP pathway required for both STM and LTM formation (Drain et al., 1991). Relevant to STM formation is the inhibition of S-type K^+ channels, which increases cellular excitability (Kandel, 2001). For LTM formation, PKA phosphorylates the transcription factor cAMP-response element binding protein (CREB). To mediate transcription during memory formation and maintenance, CREB complexes with various coactivators including cAMP binding protein (CBP) and CREB-regulated transcription coactivator (CRTC). Upon complex formation, CREB binds to cAMP response elements (CRE) within the genome (Hirano, 2016; Montminy et al., 1986, Smolik et al., 1992). CREs are usually located within enhancer or promoter regions of genes and often act to increase transcription. Only LTM has been shown to require CREB-dependent transcription and the targets of CREB remain of great importance to understanding the processes behind LTM (Frank and Greenberg, 1994).

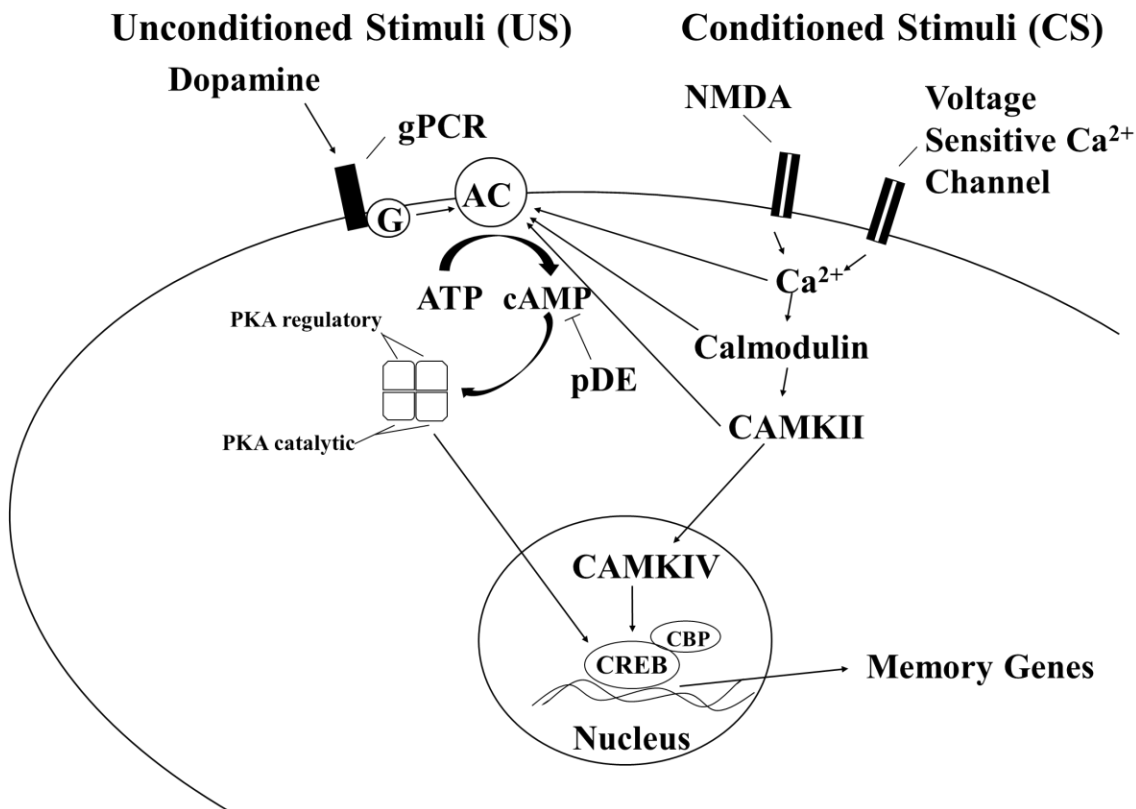


Figure 1.1: The Canonical Molecular Pathway for Memory Formation. Diagram illustrating the canonical molecular pathway for associative memory formation. Associative memory can be formed by repeated exposure to two environmental signals: the biologically neutral, conditioned stimulus (CS), and the unconditioned stimulus (US), which elicits an involuntary biological response. At a molecular level, US act upon g-protein coupled receptors (gPCR), whereas CS act on calcium-effecting receptors, like N-methyl-D-aspartate (NMDA). These two signals converge to activate adenylyl cyclase (AC), which is known to be required for both short term memory (STM), as well as long term memory (LTM) and acts to convert adenosine triphosphate (ATP) to cyclic adenosine monophosphate (cAMP). Inhibition of cAMP through phosphodiesterase's (PDE) have also shown to be required for STM. Downstream of cAMP, the tetrameric enzyme protein kinase A (PKA) disassociates and acts to phosphorylate cAMP-response element binding protein (CREB), a transcription factor known to be required for LTM. Adapted from Bolduc & Tully, 2014.

1.2 The cellular correlates of memory

At the cellular level, learning and memory can be correlated to both structural alterations within neuronal networks and changes in synaptic strength, also known as synaptic plasticity (Lisman, 1994). Synapses are junctions between neurons which act to pass electrical or chemical signals between one another. The neuronal networks created by synaptic connections show remodeling in response to environmental inputs that induce the experience-dependent learning circuit (Holtmaat and Svoboda, 2009). Remodeling of neural circuits is often represented by an increase in dendritic branching and length, dendritic spine growth and stabilization, and the formation of new synaptic contacts (Bourne and Harris, 2011; De Roo, Klausner & Muller., 2008; Trachtenberg et al., 2002). The structural changes that occur during neuronal remodeling are initially transient, with most existing only for a short period; however, some will be stabilized to become functional synapses within existing networks (Hill and Zito, 2013). Taken together, the evidence that neural networks undergo structural changes in response to environmental input highlights the dynamic nature of these networks.

Critical to associative learning and memory is the presence of two environmental signals that converge to alter synaptic strength (Lee, 2015). Long-term potentiation (LTP) is a form of synaptic plasticity that involves the persistent strengthening of synapses in response to two distinct environmental inputs (Shors and Matzel, 1997). LTP is the best candidate for being the cellular correlate of associative LTM as it has features advantageous to memory storage (Sigurdsson et al., 2007). First, and most obvious to LTM, is that LTP can enact a lasting increase in synaptic strength. Second, LTP is input-specific, with only stimulated synapses being activated, not spreading to other synapses connected to the same neuron (Andersen et al., 1980). This is an important feature as synapses individually strengthened in response to environmental inputs would display a larger storage capacity than if general changes occurred over the dendritic tree. Finally, LTP is both cooperative and associative, requiring multiple inputs, to become potentiated (Barrionuevo and Brown, 1983). Taken together with evidence showing that the cAMP pathway is a modulator of synaptic strength, it is clear that LTP offers the best candidate for being the cellular correlate of associative LTM (Frey et al., 1993).

LTP can be divided into two separate phases, distinct both temporally and mechanistically. Early phase LTP can be induced with a single stimulation and in rat hippocampal slices can last

between one and two hours (Huang & Kandel, 1994). Early phase LTP is independent of protein synthesis, instead depending on modifying existing proteins (Andersen et al., 1980). These modifications include the phosphorylation of postsynaptic AMPA receptors (AMPA) to increase their activity or by trafficking existing non-synaptic AMPARs into the postsynaptic membrane (Malinow & Malenka, 2002). The increased activity and number of AMPARs in the postsynaptic membrane is crucial as it allows future excitatory stimuli to evoke a greater response. This can be contrasted with late-phase LTP. Late phase LTP is induced by repeated stimulation and has the potential to last for days. Perhaps the most important differentiating factor between the two phases of LTP is that late-phase LTP is dependent on gene transcription (Barrionuevo and Brown, 1983; Huang & Kandel, 1994). This reliance on the expression of genes in response to environmental input is key as it indicates that to generate lasting changes to synaptic strength, ultimately, the synthesis of new proteins is required. Of the different forms of associative memory only LTM has been shown to require gene transcription. Thus, revealing the genes differentially expressed during LTM and the proteins they encode may offer new insight into the molecular mechanisms of learning and memory (Baranodes and Jarvik, 1964; Montarolo et al., 1986).

1.3 *Drosophila melanogaster* as a model organism for learning and memory processes

Drosophila melanogaster, commonly referred to as the fruit fly and hereafter referred to as *Drosophila*, is an organism that is commonly used to provide insight into the genes underlying biological processes. *Drosophila* offers a flexible model for study as it is both easy to culture and quick to breed, with each successive generation taking about ten days to develop from egg to adulthood (Roote and Prokop, 2017). However, perhaps most important to *Drosophila*'s use in genetic study is that, while structurally different from humans, there is considerable genetic homology. It has been estimated that 75% of human disease genes have a recognisable match within the fruit fly genome (Reiter et al., 2001). Available for use in *Drosophila* are many genetic tools which can be used to study these disease genes and further our understanding of the normal molecular pathways disrupted in disease.

One genetic tool available for *Drosophila* is the UAS/GAL4 system which acts to direct the expression of genes within specific cell populations (Brand and Perrimon, 1993). The UAS/GAL4 system utilizes the yeast transcriptional activator GAL4 to activate the expression of

transgenes under the control of the GAL4-specific enhancer, UAS. Tissue-specific expression of the UAS controlled transgenes is achieved by expressing GAL4 under the control of one or more transcriptional enhancers (Jennett et al., 2012). By utilizing specific enhancers to drive GAL4, its expression pattern is both predictable and reproducible, and allows for UAS-target gene expression in specific cell populations (Pfeiffer et al., 2008). In *Drosophila*, the UAS/GAL4 system has been combined with several other genetic tools to further study the role of specific genes in specific cell populations. One example of a genetic tool used in combination with the UAS/GAL4 is the targeted expression of a fluorescent reporter protein, like that used in this study, green fluorescent protein (GFP). As GAL4 alone is not visible, the use of a fluorescent cellular tag is necessary for observation of UAS/GAL4 expression by microscopy and can also be used to isolate specific fluorescently-tagged cell populations for molecular profiling (Pfeiffer et al., 2008; Henry et al., 2012). While capable of morphological and molecular profiling of normal flies, the UAS/GAL4 system can also be combined with gene knockdown tools, like RNA interference (RNAi), to study the effects that specific-gene loss has biologically. Thus, taken together, the UAS/GAL4 system offers a valuable tool for studying the role specific genes play in specific cell populations.

Drosophila is commonly used as a model organism to study learning and memory. Using olfactory shock-avoidance conditioning for training, many of the first learning and memory genes were identified in flies including *dunce*, *rutabaga*, *radish*, *cabbage* and *turnip* (Quinn et al., 1974; Dudai et al., 1976; Livingstone et al., 1984; Folkers et al., 1993; Aceves-Piña and Quinn, 1979; Choi et al., 1991). These fly mutants helped to establish the role of the cAMP pathway in memory. Further studies on *Drosophila* have shown that it also has the other required molecular components for memory that are also important in mammals, including NMDA, calmodulin and CAMKII (Lee, 2015; Malik & Hodge 2014). Like other species used to study memory, *Drosophila* shows distinct phases of memory differentiated by distinct cellular and molecular properties. Of these phases only LTM requires gene transcription (Tully, 2003; McBride et al., 1999). Thus, *Drosophila* offers an excellent model for studying learning and memory, and by using the available genetic tools, the molecular components of memory can be further dissected.

1.4 The mushroom body

The mushroom body (MB) is a region of the fly brain crucial to normal learning and memory. Chemical ablation of this structure impairs both STM and LTM in various learning paradigms including olfactory and courtship conditioning (deBelle and Heisenberg, 1994; McBride et al., 1999). Physically, this structure appears as a pair of neuropils, synaptically dense and containing multiple distinct anatomical domains (**Figure 1.2**). Overall, there are approximately 2200 neurons which have synaptic connections into the MB (Aso et al., 2014). These neurons can be broadly divided into two categories: intrinsic and extrinsic. Intrinsic neurons of the MB are located within the dorsal protocerebrum and consist of 2000 Kenyon cells (KC) (Heisenberg, 1998). KC dendrites, which cluster to form the calyx, receive olfactory input from projection neurons and send their outputs through axons that form the peduncle. The peduncle extends anterior within the brain and segregates into five terminal lobes: α , α' , β , β' and γ (Crittenden et al., 1998). It is thought that KC's which innervate each lobe play a distinct role in learning and memory processes. Specifically, γ KC's being required for STM, α/β KCs playing a role in LTM and α'/β' for memory consolidation (Krashes et al., 2007; Trannoy et al., 2011; Montague and Baker, 2016).

Extrinsic neurons of the MB include MB output neurons (MBON), dopaminergic neurons (DAN) and dorsal-anterior-lateral neurons (DAL). MBON, which number no greater than 34, have dendrites which connect to the MB lobes, forming 15 discrete compartments which receive input from KC's. Conversely, there are approximately 100 DANs, which have axons innervating MB lobes and converge upon KC-MBON compartments. This convergence on KC-MBON synapses from DAN's may be the basic computational unit of learning, acting to transform unstructured KC olfactory signal input to an ordered MBON output, encoding the basis of behavioral modification (Aso et al., 2014). Finally, DAL neurons establish synaptic contacts with α/β neurons in the frontal domain of the mushroom body calyx and are thought to act as an extra-MB memory circuit involved in LTM retrieval (Chen et al., 2012).

On a molecular level, elements of the cAMP signaling pathway are highly expressed in the MB (Blum et al., 2009). Among these are the previously mentioned proteins rutabaga, an AC, and the PDE, dunce, as well as a fly CREB homologue, *CREB2-b* (Dudai et al., 1976; Livingstone et al., 1984; Zhang, 2015). Taken together, it is clear the MB is a complex structure, composed of

varying cell-types which play significant roles in various aspects of memory functioning using components of the cAMP pathway. As such, the MB offers the best area of focus to study LTM processes in *Drosophila*.

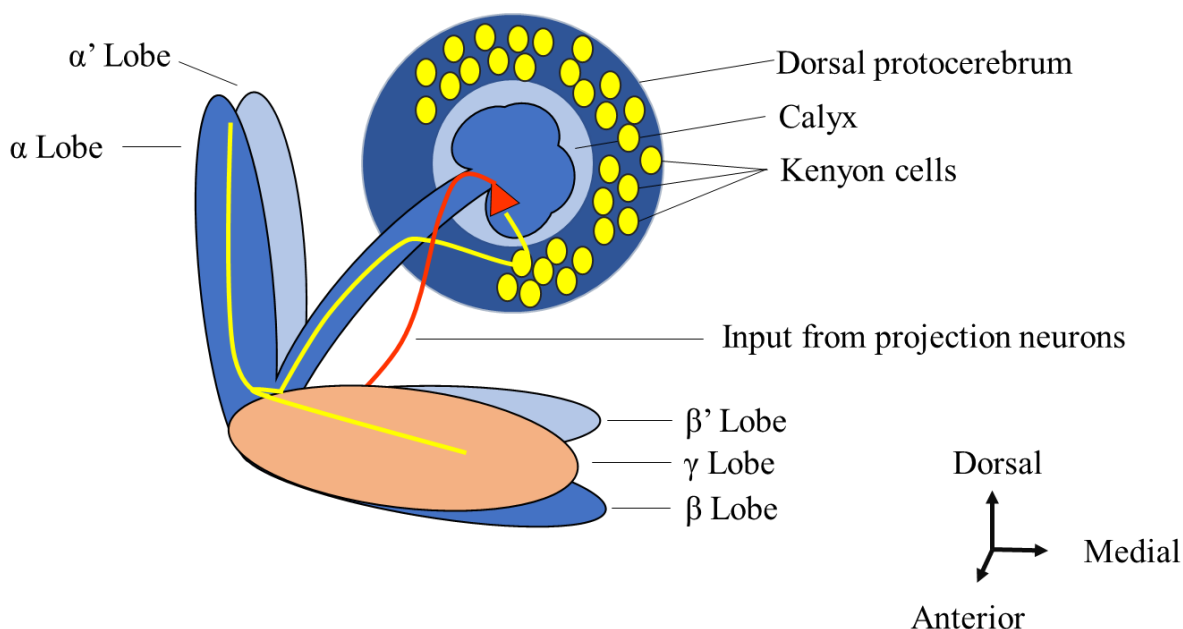


Figure 1.2: Structure of the mushroom body (MB) and transmission of an environmental input through intrinsic MB neurons. The MB is a region of the *Drosophila* brain required for learning and memory. Structurally, the MB is a symmetrical pair of neuropils (only one shown above), densely populated by axons and dendrites, but with relatively few cell bodies. The MB contains three main elements: Kenyon cells (KC), calyces and lobes. KC's are the cell bodies of the MB and are located within the dorsal protocerebrum. Dendrite-like arborizations from KC's extend inwards to form the calyx. Axons from KC's extend anteriorly in parallel, forming the peduncle. Axons forming the peduncle bifurcate and segregate into five different lobes: α , α' , β , β' and γ . The KC's innervating each of these lobes are thought to play differing roles in memory processes. Environmental input to the MB is initially received in the calyx from projection neurons (red). KC dendrites receive this information which is then ultimately relayed to the individual MB lobes (yellow). Adapted from Davis, 2011.

1.5 Courtship conditioning as a learning paradigm for *D. melanogaster*

Courtship conditioning is a commonly used learning paradigm for both STM and LTM. Normal courtship behaviour refers to a set of stereotypical actions that male flies exhibit upon being exposed to a potential mate (Spieth, 1974). Some of these behaviours include orienting towards and pursuing the female, singing a song through wing vibrations and emitting sex-specific pheromones (Burnet et al., 1971; Grillet et al., 2006). These behaviours ultimately lead to an attempt at copulation which the female will respond to either positively, spreading both wings outward to indicate willingness, or negatively, kicking to repel the male (Spieth, 1974). Concurrent to female mate determination, the male makes a similar determination of female suitability and receptiveness through an assessment of various auditory, visual, chemosensory and mechanosensory cues. As males often initiate courtship behaviour with inappropriate targets, including other males, these cues are crucial in determining if the male should continue or terminate courting behaviour (Manning, 1959).

Courtship conditioning relies on male sexual behaviour being modifiable in response to prior experience (Siegal and Hall, 1979). In the courtship conditioning assay, a newly-eclosed male is isolated for five days, remaining socially naïve to the mating behaviour of female flies. After this isolation period, the naïve male is placed with a single pre-mated female, which will not re-mate after prior copulation. During this training period, the male attempts to court the female, however, the female is unreceptive to the male fly's advances. As a response to the failed mating attempts, the male fly will suppress future courting attempts towards the female. Critical to this training period is that the male learns to associate the failed copulation attempts with an olfactory cue, the pheromone profile of the pre-mated female, and will continue to suppress courting behaviours upon re-exposure to the same olfactory stimuli. Thus, in this learning paradigm, courtship suppression acts as a measure of learning and the retention of this behaviour is a representation of memory. By extending the training period between naïve male and pre-mated female, both STM and LTM can be formed (McBride et al., 1999; Griffith and Ejima, 2009). For STM to be formed, a one-hour training is required, whereas in LTM, a five to seven-hour training is required. To observe memory formation, males are re-isolated after training and paired with a new pre-mated female after either one hour for STM or 24 hours for LTM. The time spent courting

by the trained male towards the new pre-mated female can be measured and compared to that of a naïve male, to confirm induction of memory formation.

As a form of associative learning, courtship conditioning is similar to other olfactory conditioning paradigms. In classic olfactory conditioning, flies are trained to modify their behaviours in response to the pairing of either a shock (aversive conditioning) or sucrose (appetitive conditioning), representing the US, with a specific odor, the CS. In courtship conditioning, it is thought that during training the male pheromone cis-vaccenyl acetate (cVA), acts as the CS (Ejima et al., 2007). During mating, the male deposits cVA on the female and this acts to distinguish virgin from mated females to other males. Similar to appetitive conditioning, cVA acts to provide input to the γ lobe of the mushroom body through dopamine receptors (Keleman et al., 2012; Montague and Baker, 2016). Additionally, while cVA naturally suppresses courting behaviour, this effect is amplified upon the pairing of cVA with unsuccessful copulation. Therefore, in courtship conditioning, this rejection acts as the US (Ejima et al., 2007).

While similar to other olfactory conditioning paradigms, courtship conditioning contains two main distinctions. Practically, courtship conditioning benefits from being capable of inducing LTM using a single mass training period of five to seven-hours (McBride et al., 1999). This is unlike other olfactory conditioning assays which require repetitive, spaced CS/US pairing to induce LTM (Tully et al., 1994). Continual, mass training in courtship conditioning is possible because males naturally space their mating attempts, eliminating the requirement for manual separation during the training period (McBride et al., 1999). However, perhaps the most important distinction from other olfactory conditioning assays, is that courtship conditioning manipulates a naturally occurring behaviour, courtship suppression, requiring minimal external input for the formation of memory. As such, courtship conditioning may reflect a more biologically-relevant form of LTM for the study of learning and memory processes.

1.6 Transcriptome analyses of LTM in *D. melanogaster*

Forward genetic screens, which aim to identify the genetic basis of behavioral phenotypes, have been an approach used by many studies to identify the molecular components of memory. Early studies using this approach, using chemical mutagenesis to induce single-gene mutants, identified much of what we know about LTM, including the importance of the cAMP pathway (Quinn et al., 1974; Dudai et al., 1976; Livingstone et al., 1984; Folkers et al., 1993; Aceves-Piña and Quinn, 1979; Choi et al., 1991). With advances in the genetic tools available for *Drosophila*, including the UAS/GAL4 system, the number of genes that can be screened simultaneously have been greatly expanded upon. One recent example of this is a study by Walkinshaw et al. (2015). Using a central-nervous system specific GAL4 driver, *nsyb-gal4*, Walkinshaw et al. screened 3655 single gene UAS-RNAi lines to identify 3h post-training memory defects in olfactory aversion conditioned flies. Overall, >500 genes with reduced memory function and >40 genes that enhance memory were identified. While a large-scale RNAi screen benefits from directly observing memory perturbations *in vivo*, one drawback to the approach used by Walkinshaw et al. is that it potentially limits its search for candidate genes by only including RNAi lines for genes specific to neuronal processes.

Transcriptome-wide profiling of gene expression, using technologies like RNA-sequencing and microarray, is an approach for identifying candidate genes without the potential selection bias introduced by large scale RNAi screens. Transcriptome profiling is a particularly effective approach for the study of LTM, as it is the only phase of memory which requires gene transcription. With CREB acting as the primary transcription factor required for LTM formation, identifying the genes differentially expressed during LTM may also help elucidate the downstream targets of CREB, which have not been fully established. Currently, few studies have profiled transcriptome changes during LTM in flies and include those by Dubnau et al. (2003) and Winbush et al (2012). Using microarray, Dubnau et al. (2003) profiled whole fly-heads 0, 6 and 24h post olfactory avoidance training to identify 42 transcriptionally regulated candidate genes. Mutants of some of these candidate genes were found to yield defective memory including *staufen*, *pumilo* and *oskar*, which have mRNA localization and translational regulation roles. Conversely, Winbush et al. (2013) used RNA-sequencing to profile whole fly-heads 24h post-training in courtship conditioned flies. This approach identified 91 differentially expressed genes including *fruitless*,

which is involved with the sexual differentiation of male neural circuits, and *orb2*, which functions to maintain activity-dependent synaptic changes. One drawback to these studies by Dubnau et al. and Winbush et al. is that by using whole-fly heads for analysis, a significant fraction of neurons non-specific to memory are profiled. This increases unwanted biological variance and could prevent the identification of LTM candidate genes. Thus, one area of focus for future transcriptome-wide studies is to only profile neurons altered by memory formation.

Current literature has begun to shift to reflect this need for increased biological resolution. In a recent study by Crocket et al. (2016), patch-clamp pipets were used to harvest RNA 30 minutes post olfactory avoidance training from specific MB cells types. Using this approach Crocker et al. revealed that MB cell type could be determined by the expression of certain cell surface receptors, as well as also identifying several differentially expressed genes in 3 types of MB extrinsic neurons, including the light-sensing genes *NinaC*, *pinta*, *Rh3* and *Rh4*. Interestingly, Crocker et al. did not identify differential expression in α/β or γ KC's, which they attributed to their approach for sample pooling.

While Crocker et al. offer the next step for observing cell-specific gene expression during LTM, the limitations of their methodology highlight the challenges presented in isolating pure samples of individual cell types. Techniques like patch-clamp pipetting and fluorescence-activated cell sorting (FACS), are limiting in that they require extensive tissue manipulation and handling, which often introduces artifacts, and yield minimal biological material. One method which looks to improve upon the challenges of these methods is the isolation of nuclei tagged in a specific cell-type (INTACT). Originally described in *A. thaliana* and later extended to *C. elegans* and *D. melanogaster*, the INTACT method isolates nuclei marked with a genetically encoded tag (Deal and Henikoff, 2010; Steiner et al., 2012). Specifically, using the UAS/GAL4 system, desired nuclei are tagged with *unc84-GFP*, a nuclear membrane protein fused to the fluorescent tag GFP (Henry et al., 2012). These tagged nuclei are then purified from non-tagged nuclei using anti-GFP bound beads (**Figure 1.3**). With a wide selection of GAL4 lines to drive expression of *unc84-GFP* in desired cell-types and using a procedure which requires minimal handling, INTACT offers a powerful tool for eavesdropping on the molecular processes of LTM in the nucleus.

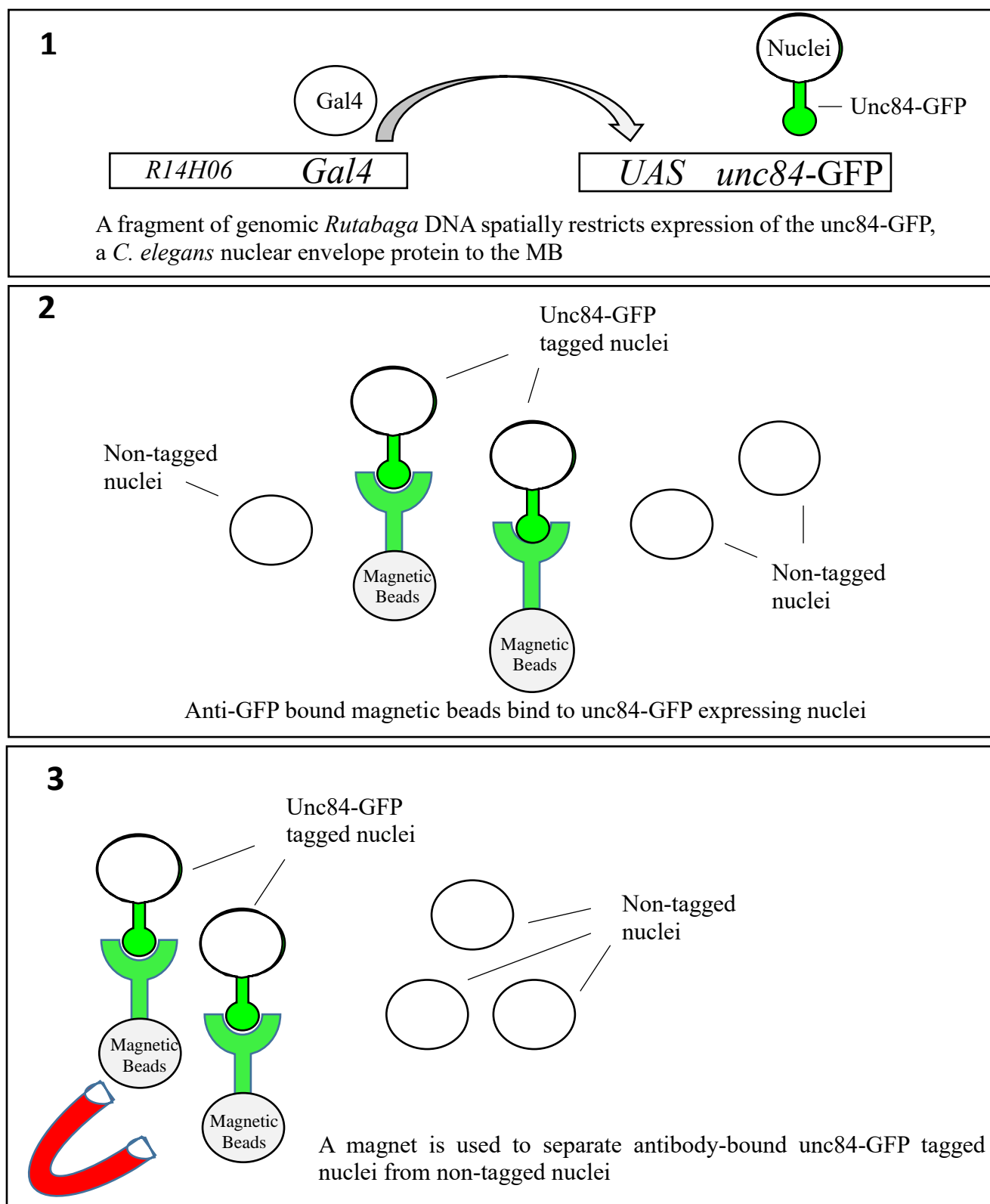


Figure 1.3: Isolation of Nuclei in a Specific Cell Type (INTACT). Schematic illustrating the isolation of MB nuclei using the INTACT method.

1.7 Study objective

LTM requires CREB-mediated gene transcription and the synthesis of new proteins to establish a persistent cellular and molecular footprint. However, very little is known about which genes are transcriptionally regulated during LTM. One approach used to identify candidate genes is through transcriptome-wide genetic screens profiling gene expression changes induced by LTM. Previous studies of LTM-induced gene expression have predominantly profiled whole fly-heads, which contain a significant fraction of non-neuronal tissue and can increase biological variance. Current literature has shifted to focus on cell-type specific profiling; however, these studies have only profiled one time-point and have had technical limitations. Thus, we hypothesize that currently identified memory-regulated genes only reflect a subset of those involved in LTM.

This study, using advances in isolating specific cell types, looks to expand upon the literature by characterizing gene expression changes in a memory-specific neuronal subset over a time course of LTM formation and maintenance induced through manipulation of a biologically relevant behaviour. Thus, using *Drosophila melanogaster* as a model organism, the objective of my study is to identify differentially expressed genes in the mushroom body during a time course of LTM formation and maintenance. It is expected that our results will provide a list of candidate genes which will generate novel hypotheses and studies which will help further our understanding of the molecular mechanisms underlying memory.

Chapter 2: Methods

2.1 Fly stocks

All *Drosophila* strains were cultured at 25° C and 70% humidity on a 12:12 light:dark cycle. Cultures were raised on a standard medium (cornmeal-sucrose-yeast-agar) supplemented by the mold inhibitors methyl paraben and propanoic acid (Koemans et al., 2017). To utilize the UAS/GAL4 expression system flies containing the MB-specific GAL4 line R14H06-Gal4 (Bloomington Stock #48667) (R14H06Gal4) were crossed to flies with UAS_ *unc84*-2XGFP (*unc84*-GFP), which encodes a *C. elegans*-derived nuclear tag combined with green fluorescent protein (GFP). R14H06Gal4 flies were generated by Janelia Farm Research Campus and obtained from Bloomington stock center and *unc84*-GFP flies were donated by Gilbert L. Henry, Janelia Farm Research Campus (Jennett et al., 2012; Henry et al. 2012). Flies used for transcript analysis were heterozygotes generated by crossing *unc84*-GFP;R14H06Gal4 flies to P{CaryP}attP2 flies also obtained from the Bloomington. This cross generated flies for downstream analysis with the genotype *unc84*-GFP/+;R14H06Gal4/+ (MB-UNC84). Courtship conditioning was performed using pre-mated, wild-type females with a Canton-S:Oregon-R genetic background (generated by J. Kramer).

2.2 LTM induction using courtship conditioning

Long-term memory was induced using a modified version of the courtship conditioning assay (Siegal and Hall, 1979; McBride et al., 1999; Koemans et al., 2017). Newly eclosed MB-UNC84 males were collected and individually held in an isolation chamber for approximately five days. Males were then trained by introducing a single pre-mated female into the isolation chamber for a period of six to seven hours. After training, males were separated from females and isolated. Flies being used for RNA-seq analysis were collected one-hour post-training (trained), to represent LTM formation, and 24-hours post-training (trained and naïve), to represent LTM maintenance (**Figure 2.1**). For each day of training, a subset of naïve and trained males was tested for LTM induction by being transferred to a 1 cm diameter chamber, re-introduced to a new pre-mated female and filmed for 10 minutes. For each male, a courtship index (CI) was calculated by manual visual analysis. CI is the percentage of time spent by a male fly engaging in courtship behaviour during the 10-minute period. The CI of trained flies was then compared to the CI of naïve flies to

calculate a learning index (LI), which is the percent reduction in courtship behaviour due to training.

$$LI = \left(\frac{CI \text{ Naive} - CI \text{ Trained}}{CI \text{ Naive}} \right) \times 100$$

Statistical significance of courtship suppression was evaluated using a Mann-Whitney *U*-test with critical P-values set to 0.05 or less.

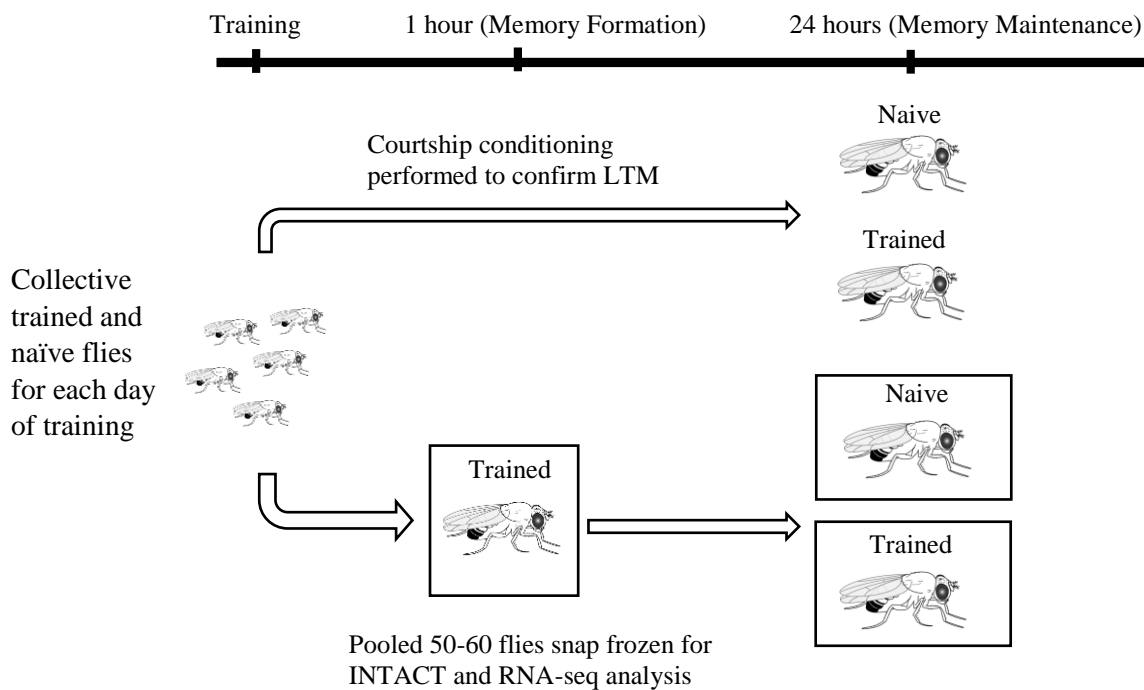


Figure 2.1: Schematic showing approach used for sample collection and LTM validation.

Newly eclosed MB-UNC84 males were isolated for five days and then trained by being paired with an unreceptive, pre-mated female. After training, males were re-isolated and collected either 1h post-training or 24h post-training. These time points were used to represent LTM formation and LTM maintenance, respectively. In parallel, for each day of training a subset of naïve and trained flies were tested for LTM induction 24h post-training. LTM induction was tested by pairing with a different pre-mated female for 10 minutes and courting behaviour measured. Boxes represent groups that were collected for INTACT and RNA-seq analysis.

2.3 Isolation of nuclei tagged in a specific cell-type (INTACT)

To isolate the mushroom body for downstream transcriptome analysis, a modified version of the INTACT method was utilized (Henry et al., 2012) (**Figure 1.2**). Antibody-bound magnetic beads were freshly prepared for each immunopurification by absorbing 5 μ g of anti-GFP antibody (Invitrogen: G10362) to 300 μ l of Protein G Dynabeads (Invitrogen: 10004D) in 200 μ l PBS/0.1% Tween 20 for 10 minutes at room temperature. Beads were then isolated and re-suspended in 300 μ l of PBS/0.1% Tween 20. Non-specific binding beads were prepared simultaneously using the same procedure without the addition of anti-GFP.

Samples of approximately 50-60 adult male flies (**Figure 2.1**) were anesthetized with CO₂ and flash frozen in liquid N₂. Fly heads were isolated from the abdomen, wings and legs by vortexing followed quickly by separation through a series of sieves. Heads were then suspended in 30 ml of a homogenization buffer (25 mM KCl, 5 mM MgCl₂, 20 mM tricine, 0.15 mM spermine, 0.5 mM spermidine, 10 mM β -glycerophosphate, 0.25 mM sucrose, 1X protease inhibitors (Invitrogen: A32965), pH 7.8) and blended for approximately one minute. To disrupt the cell membrane and release nuclei into solution, first, NP40 was added to the homogenate to an end concentration of 0.3%. This homogenate was then transferred to a 40 mL Dounce homogenizer and cells physically disrupted by plunging six times (tight-pestle B). The homogenate was then filtered using a 40 μ m cell strainer into a new 50 ml falcon tube, at which point a 1 ml input fraction was taken. This input fraction is representative of the whole head, containing both MB-specific GFP nuclei untagged non-MB nuclei. Input fractions were then centrifuged at 4000 xg for 10 minutes (4 °C) and the supernatant discarded, to generate a nuclear pellet and stored on ice. To reduce non-specific binding of GFP-negative nuclei and proteins, the homogenate was pre-cleared by adding 300 μ l of beads with no anti-GFP and incubated for 10 minutes at 4°C with rotation. Beads were then collected on a magnet, the supernatant extracted and recovered into a new 50 ml falcon tube. Next, 300 μ l of anti-GFP bound beads were added to the supernatant and incubated for 30 minutes at 4°C with rotation. Beads were then collected using a magnet, the supernatant removed and washed in 10 ml of homogenization buffer for 10 minutes at 4°C with rotation. After washing, the beads were collected using a magnet, the supernatant extracted, the beads re-suspended in 1 ml of homogenization buffer and then transferred to a new 1.5 ml Eppendorf tube. The beads were then once again collected on a magnet and the supernatant carefully removed using

multiple pipetting steps. This remaining bead-bound nuclear isolate represented the enriched fraction, containing MB-specific GFP nuclei, which was used for downstream transcript profiling.

2.4 RNA isolation and RNA-sequencing sample preparation

RNA was isolated using a PicoPure RNA Isolation Kit (Invitrogen: KIT0204) for both the input and enriched fractions according to the manufacturers instructions complemented with on-the-column DNAase treatment (Qiagen: 79254). Nuclear RNA was then converted to complementary DNA (cDNA) using a Nugen Ovation Drosophila RNA-Seq System 1-16 (Nugen: NU035032). cDNA was then sheared to a target size between 200-300 bp using a Covaris S2 sonicator according to the manufacturers protocol. Library synthesis steps were performed according to the manufacturers protocol for the Nugen Ovation Drosophila RNA-Seq System 1-16, and included a *Drosophila*-specific rRNA depletion step, as well as library amplification step, guided by real-time quantitative polymerase chain reaction (qPCR). Completed libraries were then sequenced on an Illumina NextSeq500 to 75 bp read length with single-end reads at London Regional Genomics Centre.

2.5 INTACT validation by qPCR

To determine specificity of the INTACT protocol, real time quantitative polymerase chain reaction (qPCR) was performed on RNA samples obtained in parallel with the samples used for RNA-seq analysis. Primers were designed using FlyPrimerBank to detect MB-enriched transcripts (*dac*, *oamb*, and *unc84*), MB-depleted transcripts (*repo*), and reference transcripts (*betacop*, *eif2b*, *polIII*, *Rac1*, *act5c*) (Hu et al, 2013) (**Table 2.1**). Primer amplification efficiency was validated through serial dilutions and were included if they had an efficiency of 100% +/- 10. RNA isolated from INTACT was converted to cDNA using the recommended protocol from the SensiFAST cDNA Synthesis Kit (Bioline: BIO-65053). qPCR was performed using a SensiFAST SYBR No-ROX kit (Bioline: BIO-98020) with a final reaction volume of 10 μ l on a Bio-Rad CFX-384 Touch Real-Time PCR Detection System. Quantification cycle and melt curve analysis was determined using Bio-Rad CFX Manager. Log₂ fold change values were then calculated between enriched and input samples for reference normalized MB-specific and MB-depleted genes.

Table 2.1: Primers used to validate MB-Specificity of INTACT. Forward and reverse sequences for primers obtained using FlyPrimerBank to determine MB-enriched profiles on samples obtained simultaneously to those used for downstream transcriptome analysis.

Primer Name	Forward Sequence (5' to 3')	Reverse Sequence (3' to 5')
<i>dac</i>	CCAAGGTCGTACAACCTCACCG	AGAGCATCGTTTCGTTGCTAA
<i>Oamb</i>	TGGCAACTGCCTCGTTGCTAA	GGCCACAGCTAGGTTGACAATA
<i>repo</i>	TCGCCCAACTATGTGACCAAG	CGGCGCACTAATGTACTCG
<i>unc-84</i>	AACTTCCACGCCTTTGTTCC	TGGTCAGCTTCATGTAGGCA
<i>Act5C</i>	AAGCTGTGCTATGTTGCCCT	ATTCCCAAGAACGAGGGCTG
<i>βCOP</i>	AGCGGGTAATCAAGTTGCTG	GGCAGGACGAAGCGTATGA
<i>Pol2</i>	CTGCGAAATCTAACTTACTCCGC	GAAAGTCTTTTGATGCTGCGTT
<i>eIF2Bβ</i>	CAGACCCTTAACTTTAGCTCCG	GATGGTCAAATCTGAGACCTGG
<i>Rac1</i>	GGAAATCGAACCATGCAGGC	GTCGAACACGGTGGGTATGT

2.6 RNA-seq data analysis

Raw sequence reads were trimmed using Prinseq quality trimming to a minimum base quality score of 30 (error probability of 1 in 1,000 base calls) (Schmieder and Edwards; 2011). Read quality was then visualised using FastQC (Andrews, 2010). Trimmed reads were then aligned to an annotated *D. melanogaster* reference genome (Ensembl release 88) using STAR aligner (Dobin et al., 2013; Aken et al., 2016). To look at MB-enrichment the *C. elegans*-derived nuclear tag *unc-84* (accession: NC_003284.9:13584780-13589496) was added to the reference genome for alignment. Mapped reads that uniquely aligned to one locus with a maximum of four mismatches were then used by HTSeq-count using the default union setting to generate counts of reads mapping to genic regions (Anders et al., 2015). All gene features (including introns and exons) were selected to generate gene count tables because nuclear RNA was sequenced, which includes pre-spliced features. Reads mapping to *Drosophila* ribosomal genes were quantified and then removed from count tables prior to differential expression (DE) analysis. *Drosophila* rRNA assessment was performed to ensure the effectiveness of the rRNA depletion step of library preparation. Samples that had >5 million genic non-rRNA reads, a cut-off selected to optimize coverage depth and number of replicates, were then used in R for DE analysis using DESeq2 (R Core Team., 2015; Love, Huber & Anders, 2014).

To determine the MB-specificity of sequenced samples, count tables were normalized for size factors for genes which on average had a coverage of ≥ 1 count between samples (11714 genes, 67% of annotated genes). Normalized counts for each enriched sample were then compared to the geometric mean of four sequenced input samples for a selection of genes known to be MB-enriched MB-enriched (*dac*, *oamb*, *sNPF*, *ey*, *toy*) or depleted (glia-specific *repo*), as well as *unc-84*. To determine the consistency of MB-enrichment between samples and experimental conditions the percent relative deviation was determined for each gene. To further visualize MB-enrichment of sequenced samples principal components analysis (PCA) was performed on log transformed values of the normalized counts using the *plotPCA* function in DESeq2 with *blind* set to “false”.

To determine genes differentially expressed during LTM, count tables for enriched samples were normalized for size factors after eliminating genes that on average had less than 100 counts across samples. Highly represented genes were utilized for analysis as low-count genes can

decrease the power of detection by affecting the multiple testing correction used to calculate the false discovery rate (Conesa et al., 2016). This left 6986 genes, representing 40% of all annotated genes, with sufficient coverage for subsequent analysis. DE analysis was then performed for each potential comparison between experimental conditions and genes deemed significant if they had a q value of <0.2 and a fold difference of 1.3 up or down.

2.7 GO and motif enrichment analysis

Gene ontology (GO) analysis was performed using the DAVID Bioinformatics Resources 6.8 (Huang, Sherman & Lempicki., 2009). Gene lists were uploaded to DAVID and compared to a manual input background list which included all genes found to be represented by at least 2 counts across samples (11714 genes). GO terms were identified for biological processes, molecular functions, as well as cellular components and declared significant if they had an uncorrected p-value of < 0.05 . Further functional analysis of the individual genes associated with each enriched term was provided by FlyBase (Gramates et al., 2017)

Identification of the CRE motif and *de novo* motifs within the DE gene lists was performed using Hypergeometric Optimization of Motif EnRichment (HOMER) (Heinz et al., 2010). For both the identification of CRE motifs and *de novo* motifs, HOMER was set to search 2 kb upstream and downstream of the TSS within promoter regions of the DE gene lists. Statistical significance of *de novo* motifs was calculated in HOMER by comparing enrichment of identified motifs with a length of either 8, 10 or 12 bp to their presence in the promoter region of all fly genes. Once an enriched *de novo* motif is found it is then compared to known motifs to associate the found motif to a potentially biologically relevant transcription factor.

Chapter 3: Results

This study aims to profile transcriptome changes during LTM formation and maintenance induced by courtship conditioning. To collect pooled samples of approximately 50-60 flies for analysis, courtship conditioning was performed multiple times over a period of months, with individual samples sometimes consisting of flies trained from different days. For each day flies were trained, a subset of naïve and trained male flies were tested for LTM induction 24 hours post-training. Observing LTM in these proxy flies for each day of training was necessary, not only to provide evidence that MB-UNC84 flies could form LTM, but as flies were collected from multiple crosses cultured over time, the consistency of this assays ability to induce LTM was essential. While flies used for transcriptome profiling themselves were not tested for LTM due to practical and logistical reasons, testing proxy flies for courtship suppression acts to support that the flies they were trained along with would display similar behavioural alterations. Additionally, by confirming LTM for each day of training it allowed for the removal of flies from transcriptome analysis if courtship suppression was not seen in their concurrently trained siblings.

3.1 MB-UNC84 males show normal LTM

Overall, tested proxy trained males showed reduced courting behaviour in comparison to naïve males (**Figure 3.1 A**; $P < 0.0001$ Mann-Whitney *U*-test), indicating the successful induction of LTM. While some variation was seen between training days in both the base courting of naïve flies and relative courtship suppression seen in trained flies, this is to be expected due to the normal variability in courtship behaviour and the relatively lower numbers of flies tested on each individual day (**Figure 3.1 B**). Regardless, courtship suppression in trained flies was significant on each individual day where proxy flies were tested, indicating consistency in the courtship conditioning paradigm and giving strong evidence for LTM induction in flies utilized for transcriptome analysis.

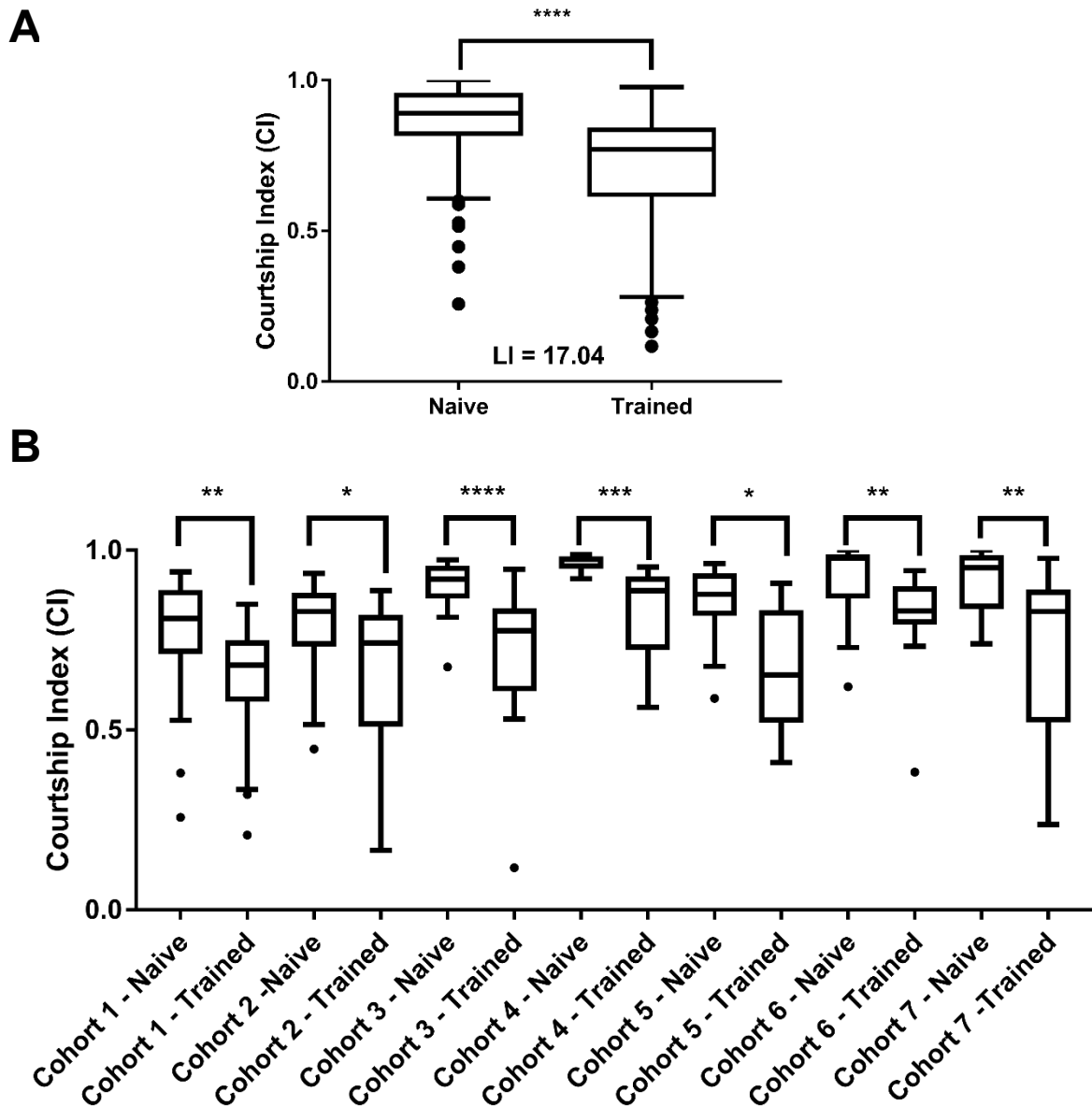


Figure 3.1: Long term courtship memory is intact in MB-UNC84 flies. Naïve male flies when paired with pre-mated female trainer flies for 7 hours show reduced courting behaviour 24 hours post-training **A**) Boxplot showing courtship indices for MB-UNC84 flies tested, $n = 109$ and 115 , for naïve and trained males, respectively; **** $P \leq 0.0001$ in Mann-Whitney U -test. **B**) Boxplot indicating courtship indices for proxy flies from individual days where flies were utilized in downstream analyses. $n = 22/25, 20/18, 22/22, 8/11, 11/7, 10/16$ and $16/16$, respectively, for each naïve/trained pair; * ≤ 0.05 , ** $P \leq 0.01$, *** $P \leq 0.001$ and **** $P \leq 0.0001$ in Mann-Whitney U -test.

3.2 Validation of MB-enrichment with INTACT

To validate that the INTACT method was capable of enriching for MB nuclei, RNA was isolated from MB-UNC84 flies for both the input (whole head) and enriched (mushroom body nuclei) fractions and converted to cDNA. qPCR was then performed using a selection of primers for genes known to be MB-enriched (*dac*, *oamb*) or depleted (*repo*), as well as *unc84*, which is expressed exclusively in the GAL4 targeted MB neurons. Across samples, modest enrichment of *dac* and *oamb* was seen in MB-enriched fractions, with fold enrichments of 2.25 and 1.7, respectively. While dramatically less than the enrichment of *unc84*, which had a fold enrichment of 25.6, this level of enrichment for *dac* and *oamb* was expected as they are not solely expressed in the MB. In addition to the consistent depletion of the glial-cell specific *repo*, with a fold of 0.2, taken together, this observed expression pattern gave a strong indication that INTACT is capable of enriching for MB nuclei (**Figure 3.2**).

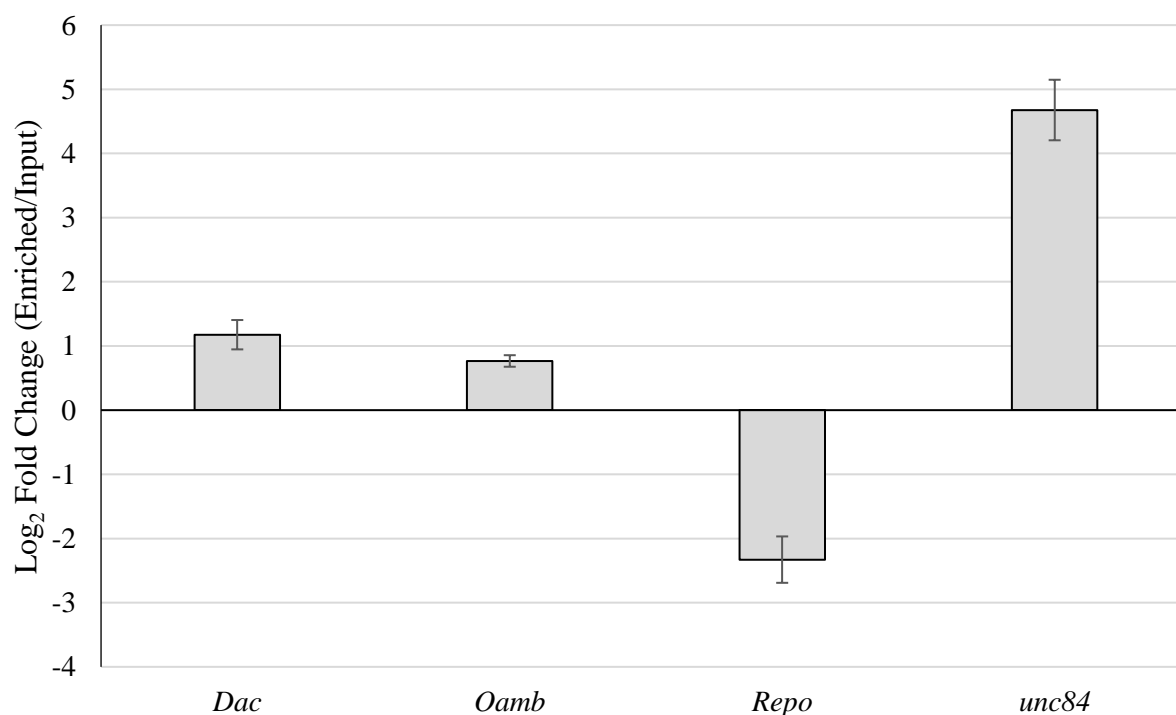


Figure 3.2: qPCR Validates MB-Specificity of INTACT. To confirm MB-enrichment of nuclei obtained from INTACT, RNA was isolated from whole head, input, and enriched fractions. MB-specificity was confirmed by observing enrichment of the genes *dac* (n=3), *Oamb* (n=3) or *unc-84* (n=2) and depletion of the glial-specific marker *repo* (n=5). Log₂ fold changes represent the enriched fraction relative to the input fraction.

3.3 Analysis and quality control of MB-UNC84 RNA-seq data

After LTM induction had been confirmed within proxy flies, INTACT was used to extract MB-nuclei from samples, followed by RNA isolation and library preparation for RNA-seq. RNA-seq libraries were obtained for four biological replicates of the whole head input fraction (I) and three biological replicates of each mushroom body enriched experimental condition: naïve (EN), one-hour post-training (E1), and 24-hour post-training (E24). To assess both the quantity and the presence of consistent fragment sizes within the 200-300 base pair range, each RNA-seq library was then run on an Agilent Bioanalyzer using a high sensitivity DNA assay kit. Libraries were then sequenced and reads processed using a bioinformatics pipeline that included the removal of low quality reads, reads with >4 mismatches, and reads which mapped to *Drosophila* rRNA genes (**Table 3.1 and Table 3.2**). The average alignment efficiency across all samples was 48.5% for high quality reads (**Figure 3.3 A**). Among the aligned reads, an average of 87.7% mapped to genic features (**Figure 3.3 B**). One EN sample was then removed for not meeting the minimum inclusion criteria of >5 million non-rRNA genic reads. This left four I, two EN, three E1 and three E24 samples for differential expression analysis.

Critical to downstream differential expression analysis is relatively consistent MB-enrichment between samples. Comparisons between samples with varying levels of MB-enrichment could potentially lead to DE of genes required for MB function and not specific to learning and memory function. As such eliminating samples with inconsistent MB-enrichment from analysis is crucial. To investigate MB enrichment, relative expression levels, compared to the input, were calculated for a selection of genes known to be MB-enriched (*dac*, *oamb*, *sNPF*, *ey*, *toy*) or depleted (glia-specific *repo*), as well as *unc-84*. Overall, 7 samples displayed MB-enriched profiles, with enrichment of *unc84*, *dac*, *oamb*, *sNPF*, *ey* and *toy*, as well as depletion of *repo* (**Figure 3.4 A**). One sample was removed from subsequent DE analysis after it was determined *repo* was not depleted (fold change 0.99) and there was low *unc-84* enrichment (fold change 8.9, compared to the average fold change of 39.2). To determine the consistency of MB-enrichment for the remaining 7 samples, percent relative deviation was calculated for each gene (**Table 3.3**). Percent relative deviation is a measure of the variation found between samples relative to the mean. Overall, relative deviation between samples was lowest for *unc-84* at 5.2%. As *unc-84* is the transcript encoding the nuclear tag used for INTACT, this suggested consistent MB-

enrichment. This was further supported by consistency for the MB-enriched genes, with relative deviations between 8 and 13%. The glial-specific *repo* had the greatest relative deviation at 59.6%, however, this was primarily driven by variation in the EN samples. While this variation of *repo* could indicate some variability in MB-enrichment between samples, overall, the reduced relative deviation of *unc-84* and MB-enriched genes strongly suggested that our samples are consistently MB-enriched.

To further support MB-enrichment of our samples, principal component analysis (PCA) was performed, revealing distinct separation of input and enriched samples (**Figure 3.4 B**). Clustering was not observed between experimental conditions, with greater variance between EN samples than that seen in E1 and E24 samples. To explore the main sources of variance contributing to sample separation, component scores were obtained for the top 10 variable genes (**Table 3.4**). Principal component 1 (PC1) accounted for 66% of the variance and contributed to the separation of input and enriched samples. This variance was correlated to gene expression of a subset of expected MB-enriched genes including the nuclear tag *unc-84* and *pvt*. This gave strong evidence that all samples were MB-enriched. Principal component 2 (PC2), which accounted for 12% of the variance, contributed to the separation of samples by experimental condition and was correlated to the expression of mitochondrial genes. This suggested that non-specific binding of biological material may be binding to the beads during INTACT immunopurification. It should be noted that the variance in PC2 is primarily limited to one EN sample, which could indicate it had more non-specific binding than other samples. However, as PC2 accounted for less variance than PC1, non-specific binding was not expected to contribute greatly to our analysis. It should also be noted that known learning and memory genes did not prominently contribute to the separation of samples by experimental condition suggesting that LTM induced gene expression is subtle. Taken together, this evidence suggested that while our samples are consistently MB-enriched, additional biological replicates may be required to reduce intra-condition variability and improve downstream analysis.

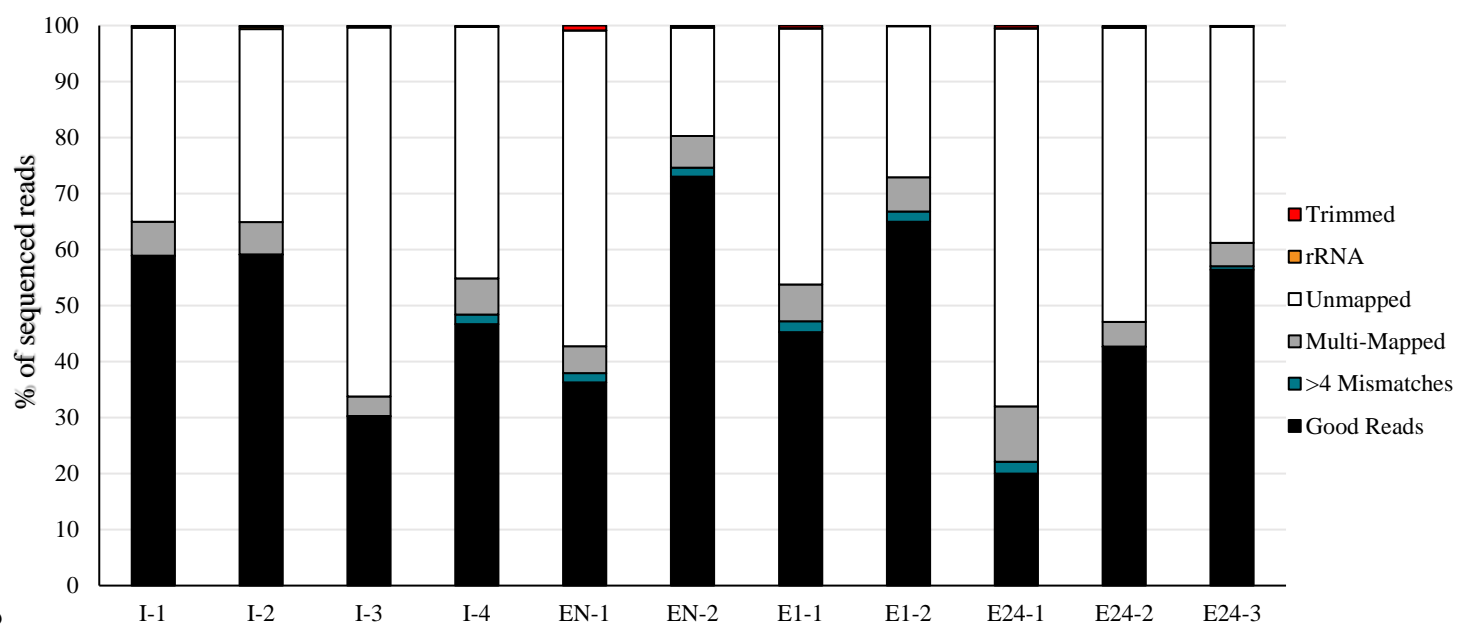
Table 3.1: Raw read distribution of RNA-sequencing data. Distribution of reads after processing by the bioinformatics pipeline for whole he input (I), naïve (EN), 1h post-training (E1) and 24h post-training (E24) samples used in the downstream DE analysis. Sequenced reads represent raw total reads generated for each sequenced sample. Trimmed represents the amount of reads which had a quality score greater than 30. Reads mapping to fly ribosomal genes are indicated as rRNA reads. Unmapped represents reads which did not align to the *Drosophila* genome. Multi-mapped represents reads which aligned to multiple loci within the *Drosophila* genome. Uniquely mapped represents reads which aligned to one loci. Reads which aligned to one loci within the *Drosophila* genome with greater than four mismatches are indicated by >4 mismatches. Good reads indicate aligned non-rRNA reads which were used to generate count tables for genic features.

Sample Name	Sequenced Reads	Trimmed	rRNA Reads	Non-rRNA	Unmapped	Multi-Mapped	Uniquely Mapped	>4 Mismatch	Good Reads
EN-1	83,977,393	83,248,350	8,052	83,240,298	47,349,315	4,042,461	31,848,522	1,387,767	30,460,755
EN-2	86,016,347	85,732,655	13,865	85,718,790	16,673,566	4,871,895	64,173,329	1,353,809	62,819,520
E1-1	122,847,123	122,301,140	63,699	122,237,441	56,163,357	8,088,944	57,985,140	2,425,119	55,560,021
E1-2	129,223,570	129,107,989	8,093	129,099,896	34,905,600	7,864,813	86,329,483	2,405,528	83,923,955
E24-1	41,294,960	41,092,458	13,307	41,079,151	27,869,785	4,088,758	9,120,608	870,039	8,250,569
E24-2	19,079,368	19,038,708	38,840	18,999,868	10,019,227	836,721	8,143,920	11,068	8,132,852
E24-3	48,625,213	48,592,466	80,256	48,512,210	18,764,087	2,024,875	27,723,248	302,603	27,420,645
I-1	19,975,383	19,937,286	40,395	19,896,891	6,918,458	1,206,946	11,771,487	12,849	11,758,638
I-2	19,835,062	19,785,715	73,175	19,712,540	6,835,367	1,142,375	11,734,798	11,528	11,723,270
I-3	19,943,350	19,920,404	49,635	19,870,769	13,138,108	695,575	6,037,086	8,709	6,028,377
I-4	65,219,971	65,092,537	17,403	65,075,134	29,316,589	4,189,123	31,569,422	1,139,597	30,429,825

Table 3.2: Raw count data for RNA-sequencing results. Distribution of count data for aligned, non-rRNA good reads for whole head input (I), naïve (EN), 1h post-training (E1) and 24h post-training (E24) samples used in the downstream DE analysis as processed by HTSeq. Reads mapping to no feature are those that could not be assigned to any feature. Ambiguous counts indicate where multiple features could be assigned for a single read and thus were excluded from DE analysis. Genic counts indicate reads mapped to introns and exons.

Sample Name	Genic	No Feature	Ambiguous
EN-1	26,986,636	1,717,518	1,756,601
EN-2	57,375,106	1,271,704	4,172,710
E1-1	48,055,978	4,279,110	3,224,933
E1-2	76,181,873	2,345,826	5,396,256
E24-1	6,029,168	1,942,996	278,405
E24-2	7,359,203	215,066	558,583
E24-3	25,025,546	572,531	1,822,568
I-1	10,331,447	556,794	870,397
I-2	10,466,265	362,148	894,857
I-3	5,308,654	240,615	479,108
I-4	26,597,216	1,596,212	2,236,397

A



B

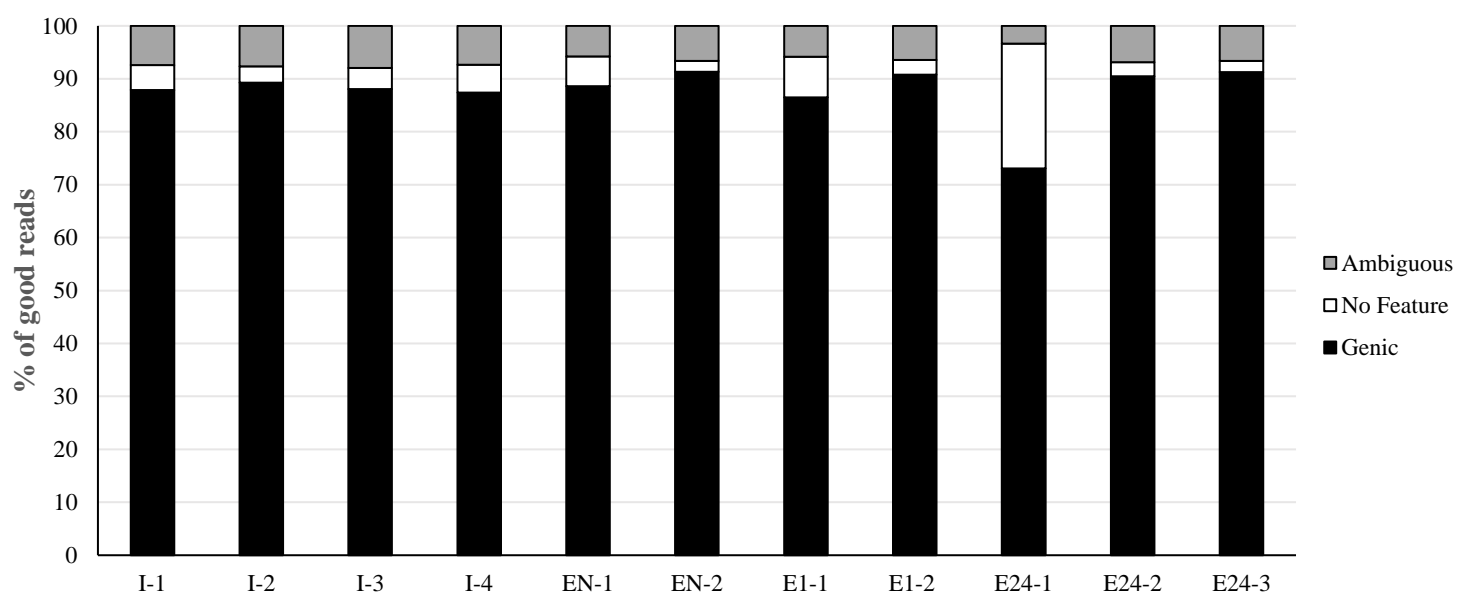


Figure 3.3: Alignment efficiency and association of reads with genomic features for INTACT-obtained RNA-sequencing results. Processing, alignment and count results for INTACT-derived sequencing data for whole head input (I), naïve (EN), 1h post-training (E1) and 24h post-training (E24) samples used in downstream DE analysis. **A)** Distribution of reads after processing by the bioinformatics pipeline represented as the percentage of total sequenced reads for each sample. Trimmed represents reads with a quality score less than 30. Reads mapping to fly ribosomal genes are indicated as rRNA. Unmapped indicates reads that did not align to the *Drosophila* genome. Reads which mapped to multiple loci or had greater than four mismatches are indicated by multi-mapped and >4 mismatches, respectively. Good reads indicate reads which were used to generate gene count tables. **B)** Distribution of counts for genic features (introns and exons) as processed by HTSeq, represented as a percentage of the total good reads for each sample. Reads mapping to no feature are those that could not be assigned to any feature. Ambiguous reads indicate where multiple features could be assigned for a single read and thus were excluded from DE analysis.

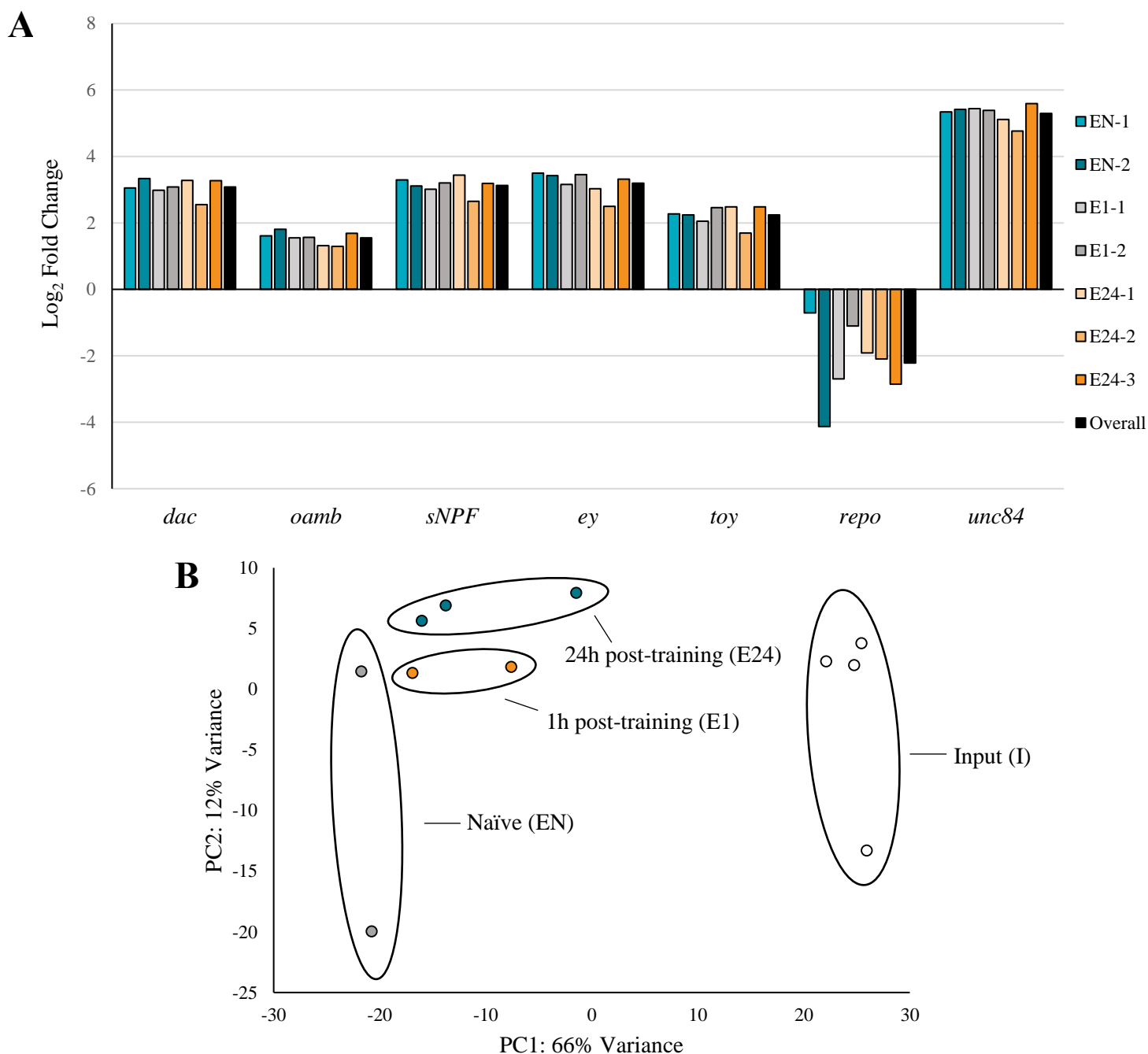


Figure 3.4: Sequencing data for RNA isolated from INTACT-obtained nuclei shows MB-enrichment. Sequencing data for naïve (EN), 1h post-training (E1) and 24h post-training (24) samples. **A)** Normalized counts were compared between enriched and input samples showing that the relative expression of the nuclear tag *unc84*, as well as a selection of MB-specific and depleted genes, indicates a MB-enriched profile. **B)** Principal component analysis was performed on transformed count data using the *plotPCA* function within DESeq2. The resulting clusters show distinction between enriched and input samples.

Table 3.3: Consistency of MB-enrichment between samples. Standard deviation and mean were calculated using \log_2 fold change data between enriched and input fractions for each gene used to determine MB-enrichment. Relative deviation was calculated by dividing the standard deviation by the absolute value of the mean. Consistency was greatest for the nuclear tag used for INTACT, *unc-84*, followed by MB-enriched genes (*dac*, *oamb*, *sNPF*, *ey*, *toy*). Depletion of glial-specific *repo* had the greatest variability.

Gene	Mean	Standard Deviation	Relative Deviation (%)
<i>dac</i>	3.07	0.27	8.76
<i>oamb</i>	1.54	0.19	12.07
<i>sNPF</i>	3.12	0.25	8.03
<i>ey</i>	3.18	0.35	11.07
<i>toy</i>	2.22	0.29	13.10
<i>repo</i>	-2.01	1.15	59.59
<i>Unc-84</i>	5.29	0.27	5.16

Table 3.4: Genes contributing the greatest source of variance for the first two principal component. To explore the main sources of variance contributing to the separation of samples in the principal cluster analysis (PCA) (**Figure 3.4**) component scores were obtained for the top 10 variable genes. Principal component 1, which accounted for 66% of the variance, was correlated to gene expression of a subset of expected MB-enriched genes including the nuclear tag *unc-84* and *prt*. Principal component 2, which accounted for 12 % of the variance, contributed to the separation of samples by experimental condition and was correlated primarily to mitochondrial genes.

Flybase ID	Gene Name	PC1	Flybase ID	Gene Name	PC2
FBgn0013278	<i>Hsp70Bb</i>	-0.13037	FBgn0013686	<i>mt:lrRNA</i>	-0.31501
N/A	<i>Unc-84</i>	-0.11797	FBgn0013688	<i>mt:srRNA</i>	-0.23588
FBgn0000053	<i>Ade3</i>	0.096551	FBgn0005391	<i>yp2</i>	0.185416
FBgn0002563	<i>Lsp1beta</i>	0.089866	FBgn0004047	<i>Yp3</i>	0.148978
FBgn0001258	<i>ImpL3</i>	-0.08788	FBgn0030334	<i>Karl</i>	0.147092
FBgn0030334	<i>Karl</i>	0.08773	FBgn0046323	<i>ORY</i>	-0.14501
FBgn0004102	<i>Oc</i>	0.082712	FBgn0002563	<i>Lsp1beta</i>	0.133368
FBgn0043005	<i>Prt</i>	-0.08253	FBgn0028982	<i>Spt6</i>	0.110728
FBgn0000052	<i>Ade2</i>	0.081802	FBgn0037107	<i>CG7166</i>	0.109477
FBgn0001263	<i>inaD</i>	0.081098	FBgn0013672	<i>mt:ATPase6</i>	0.109043

3.4 DE analysis reveals a list of candidate genes differentially expressed 24h post-training

To identify a list of candidate genes involved in LTM formation and maintenance, DE analysis was performed for each potential comparison between experimental conditions- E1 v N, E24 v N and E24 v E1 - and genes deemed differentially expressed if they had a q value of <0.2 and a fold difference of 1.3 up or down. Between comparisons, this analysis identified 85 upregulated and 28 downregulated genes between E24 v E1, 21 upregulated and 11 downregulated genes between E24 v EN and no DE genes between E1 v EN (**Figure 3.5 A-C; for full list see Appendix A, Supplementary Tables 1-4**). These gene expression changes observed between comparison were subtle, as predicted from PCA (**Figure 3.4**), with a median fold change of 1.45. After removing duplicates, a total of 90 unique upregulated and 30 unique downregulated genes between the different comparisons were identified to be differentially expressed 24h post-training (**for full list see Appendix A, Supplementary Tables 5-6**). As no differentially expressed genes were identified 1h post-training, potentially due to one less biological replicate and greater intra-variation between naïve fly samples, the identified genes in this study only reflect a subset of genes transcriptionally regulated during early LTM maintenance.

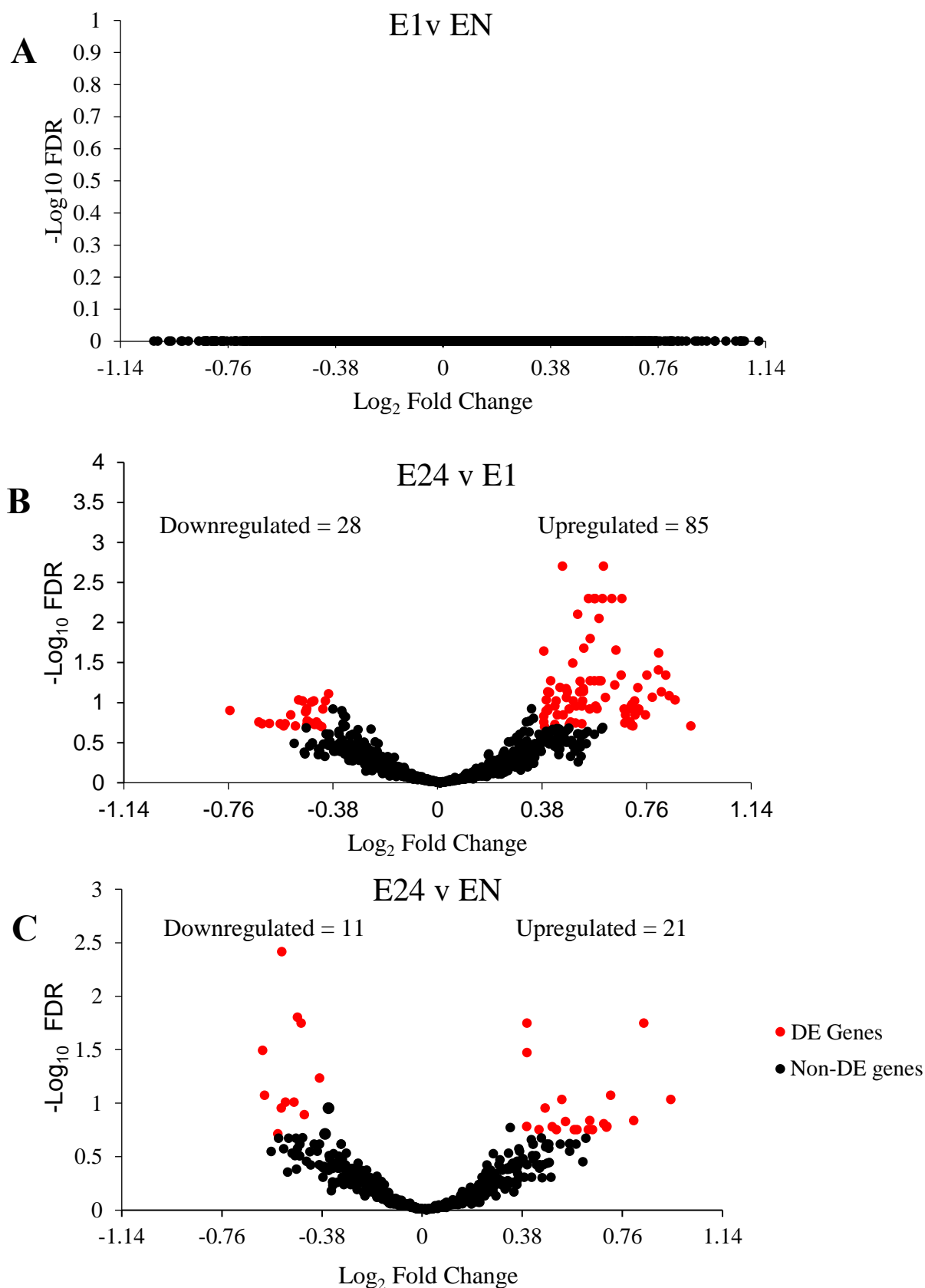


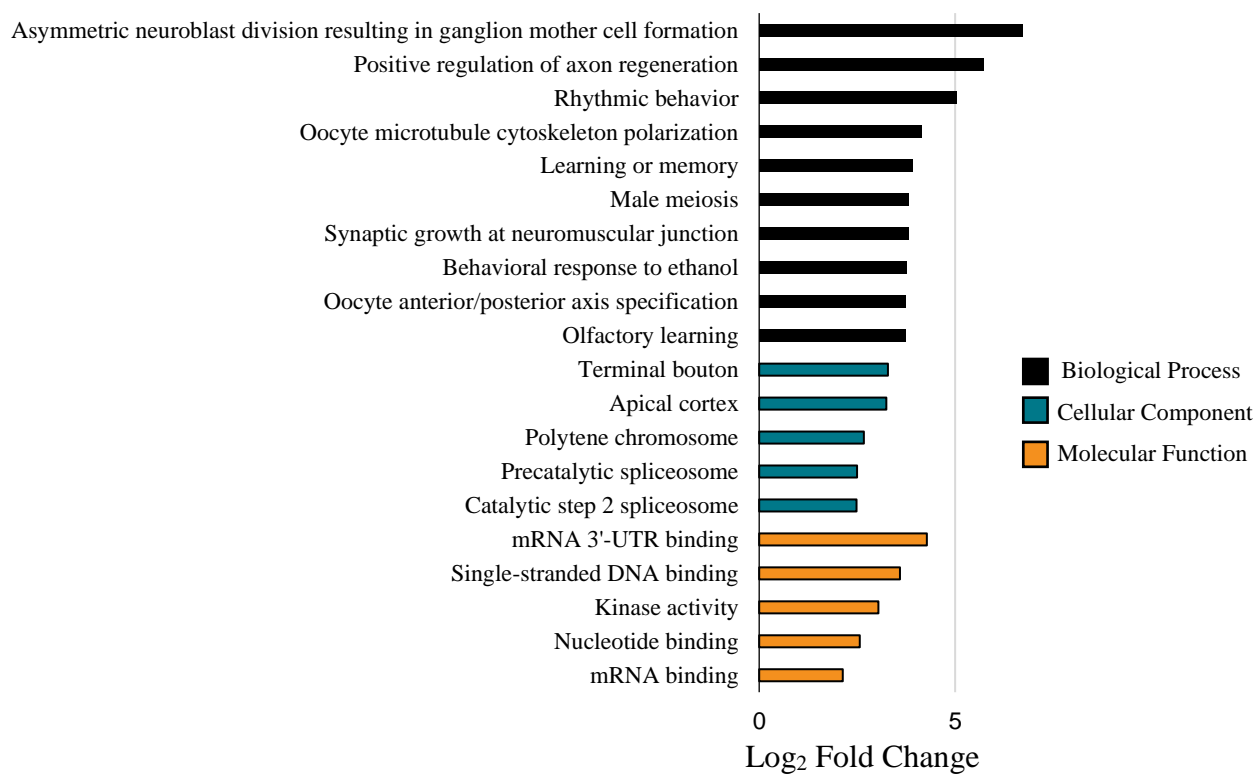
Figure 3.5: Volcano plots displaying genes identified as differentially expressed between experimental conditions. Volcano plots for each experimental comparison showing the results of the DE analysis by plotting genes using corresponding \log_2 fold change and $-\log_{10}$ FDR values. DE genes indicate $q < 0.2$, fold change > 1.3 . A) 1h post training (E1) compared to naïve (EN) B) 24h post-training (E24) compared to 1h post-training C) 24h post-training compared to naïve.

3.5 GO analysis of DE genes reveals terms enriched for learning and memory processes

To identify potentially important biological pathways and processes within the upregulated and downregulated DE gene lists, GO analysis was performed for biological processes, cellular components and molecular functions and terms declared significant if they had an unadjusted p-value < 0.05. Several enriched biological processes were related to learning and memory such as “long-term memory”, “olfactory learning” and “learning or memory” (**Figure 3.6 A; for full list see Appendix A, Supplementary Tables 7-9**). In total, 15 of 90 upregulated genes were identified to have been previously associated with biological processes relevant to courtship behaviour, courtship conditioning or memory (**Table 3.5**). Other GO terms that were enriched among the upregulated genes included: “asymmetric neuroblast division resulting in ganglion mother cell formation” (most enriched biological process), “oocyte microtubule cytoskeleton polarization”, “oocyte anterior/posterior axis specification” and “mRNA 3’-UTR binding”.

GO analysis of the highly expressed downregulated DE candidate genes revealed a limited number of enriched terms, likely due to the small number of genes (30) (**Figure 3.6 B; for full list see Appendix A, Supplementary Tables 10-11**). These terms primarily were linked to two genes encoding voltage-gated potassium channels (*elk*, *shawl*) and two genes with serine-peptidase activity (*CG11319*, *CG17684*).

A



B

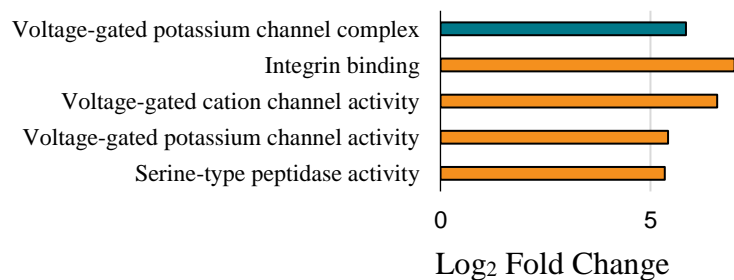


Figure 3.6: GO results for upregulated and downregulated DE genes. Significant GO terms (unadjusted p-value <0.05) for DE analysis results with enrichment shown as log₂ fold change. A) GO analysis for upregulated DE genes B) GO results for downregulated DE genes shown

Table 3.5: DE gene results reveal a list of genes previously associated with learning and memory. DE genes previously identified to be involved with: learning and memory processes or within the canonical learning pathway (L), long-term memory (LTM), olfactory learning (O), olfactory behaviour (OB), courtship behaviour (C).

Flybase ID	Gene	Associated Category	Source
FBgn0004907	<i>14-3-3 ζ</i>	L, O	Philip, Acevedo & Skoulakis, 2001
FBgn0000253	<i>Cam</i>	L	Pang et al., 2010
FBgn0261934	<i>dikar</i>	L, LTM, O	Dubnau et al., 2003; Alkalal et al., 2011
FBgn0086675	<i>fne</i>	C	Zanini et al., 2012
FBgn0011661	<i>Moe</i>	OB, LTM	Dubnau et al., 2003; Sambandan et al., 2006; Freymuth & Fitzsimons, 2017
FBgn0037705	<i>Mura</i>	L, LTM, O	Dubnau et al., 2003; Alkalal et al., 2011
FBgn0261710	<i>nocte</i>	L	Winbush et al., 2012
FBgn0000273	<i>Pka-C1</i>	L, O	Sokolowski, 2001
FBgn0022382	<i>Pka-R2</i>	L	Muller, 1997
FBgn0003093	<i>pkc98E</i>	LTM	Zhang et al., 2013
FBgn0004103	<i>Pp1-87B</i>	L, O	Sokolowski, 2001
FBgn0004595	<i>Pros</i>	C	Grosjean et al., 2007
FBgn0003371	<i>Sgg</i>	O	Wolf et al., 2007
FBgn0045823	<i>Vsg</i>	L, LTM, O	Dubnau et al., 2003; Alkalal et al., 2011
FBgn0261113	<i>xrpl</i>	OB	Sambandan et al., 2006

3.6 Identification of known and *de novo* motifs within the promoters of DE genes


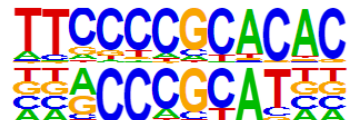



To determine if CREB may be involved in the regulation of identified DE genes, HOMER was used to locate putative CRE (5'-TGACGTCA-3') 2 kb upstream and downstream of the transcriptional start site (TSS) (Zhang et al., 2005). CREs were identified in six upregulated genes (*CG13055*, *CG43347*, *csw*, *ctp*, *Hk*, *Lk6*) and one downregulated gene (*cv-c*) (**Table 3.6**).

To identify potential novel transcription factors involved in LTM, HOMER was utilized to identify *de novo* motifs within the DE gene list. Several enriched *de novo* motifs were found, however, the motif 5'-TCTCTCTCTCTC-3', which was found in 58.26% of DE genes is of interest as it displayed the highest correlation to a known transcription factor binding site, with a 93% match to the binding site for trl (**Table 3.7**).

Table 3.6: Homer identifies CRE motifs within promoter regions of DE genes: Genes found with CRE from the upregulated (italic) and downregulated (bold italic) DE gene list and their respective position to the transcriptional start site (TSS).

Gene name	Distance from TSS
<i>CG43347</i>	-1612
<i>Hk</i>	912
<i>Ctp</i>	666
<i>Csw</i>	1412
<i>Cv-c</i>	-1596/-1482
<i>Lk6</i>	1727
<i>CG13055</i>	-1470

Table 3.7: Homer identifies *de novo* motifs within promoter regions of DE genes.
 Top 5 enriched *de novo* motifs identified among unique DE genes and their similarity to known transcription factor binding site motifs.

<i>De novo</i> Motif/ Best Match Motif (5'- 3')	P- value	% of targets	% of background	Average distance in base pairs from TSS in targets (background)	Transcription factor best match (similarity)
	1 e-12	23.48%	4.4 %	927.3 (1143.3)	Byn (0.496)
	1 e-12	23.48%	4.41%	1015.8 (1273.4)	Gcm2 (0.574)
	1 e-11	38.26%	12.17%	732.9(1262.7)	Cf1-II (0.668)
	1 e-11	32.17%	8.98%	938.1 (1183.1)	E-box (0.615)
	1 e-10	58.26%	28.12%	1055.9(1195.5)	Trl (0.929)

Chapter 4: Discussion

In this study, I have profiled the *Drosophila* transcriptome in a specific subset of MB neurons over a time course of LTM formation and maintenance. Specifically, this study offers a novel use of the INTACT method during LTM to isolate MB-nuclei, which form the learning center of the fly brain, within a hypothesis-generating RNA-sequencing experiment. By using a cell-specific approach to profile tissue specific to LTM changes, this study improves upon previous fly LTM transcriptome-wide studies where whole fly heads were profiled, which can introduce biological variability. As very few studies have profiled LTM-induced transcriptome changes, the results of this study provide a rich list of candidate genes, which through biological validation and further study, can expand our understanding of learning and memory processes.

4.1 Genes with greater transcript abundance 24h post-training

This study identified 90 genes differentially upregulated ($q < 0.2$; fold change > 1.3) 24h after LTM induction by courtship conditioning. Using GO analysis to guide the functional profiling of our upregulated DE gene list, we identified 15 genes that have been previously associated with learning and memory (**Table 3.5**). These genes encode proteins with a wide array of functions required during LTM including the cAMP-dependent protein kinase subunits Pka-C1 and Pka-R2, as well as 14-3-3 ζ and jeb, which have roles in the Ras/MAPK cascade, a pathway that ultimately converges to activate CREB (Michael et al., 1998). Only two genes identified by this study coincide with DE genes found by previous LTM transcriptome-wide studies, specifically those conducted by Dubnau et al. (2003) and Winbush et al. (2012). These studies both profiled whole fly heads, with Winbush et al. investigating courtship conditioned flies 24h post-training and Dubnau et al. using olfactory shock avoidance to observe gene expression changes 0, 6 and 24h post-training. Overlapping DE genes we identified include the cytoskeletal functioning *moesin* (Dubnau et al. 2003) and the circadian entrainment gene *nocte* (Winbush et al. 2012). This minimal overlap is to be expected as we profiled a different set of tissue than the whole-heads profiled by Dubnau et al. and Winbush et al., and used a different memory assay than Dubnau et al. No genes were found to overlap with a MB cell-type specific study conducted by Crocker et al. (2016), where patch clamp pipets were used to harvest MB neurons 30 minutes after olfactory shock avoidance training. This was also expected as we investigated a different time-point and

their differentially expressed genes were identified solely from extrinsic MB neurons, whereas our study predominantly profiled Kenyon cells. In summary, while this study does not share many genes with previous LTM transcriptome-wide studies, there was significant overlap of our upregulated DE genes with those previously associated with learning and memory. This provides evidence strongly supporting that our cell-type specific study profiled LTM induced gene expression changes. Thus, we believe that the DE genes identified by this study, but not yet associated with LTM, represent a list of novel candidate genes for further study and biological validation in learning and memory processes.

Multiple GO terms associated with our upregulated DE genes were found to be enriched with no direct link to learning and memory. These terms included “asymmetric neuroblast division resulting in ganglion mother cell formation”, “oocyte microtubule cytoskeleton polarization”, “oocyte anterior/posterior axis specification” and “mRNA 3’-UTR binding”. Interestingly, 13 of the upregulated genes associated with these terms are known to interact with the proteins *pumilo*, *staufer* and *oskar*, primarily through protein-protein or RNA-protein interactions (**Table 4.1**). *Pumilo*, *staufer* and *oskar*, which have functions related to mRNA localization and translational regulation, were previously shown to be required for LTM by Dubnau et al. (2003). Specifically, differential upregulation of *staufer* and *pumilo* was observed 6h post-training, and individual fly mutants for the three genes were shown to yield defective LTM. Dubnau et al. suggested that *pumilo*, *staufer* and *oskar* provide a molecular mechanism for the synapse-specific delivery of gene products during LTM, a hypothesis which has been supported by further study (Heraud-Farlow & Kiebler, 2014). As the genes identified by our study are differentially expressed predominantly 24h post-training, I hypothesize that we have identified downstream targets of *oskar*, *staufer* and *pumilo* for RNA localization which could have significant roles in LTM. As 3 of these proposed downstream targets, *Act5C*, *14-3-3 ζ*, and *pros* are known learning and memory genes, this strongly suggests that the other proposed targets we have identified will be as well. This could be further explored using adult-specific RNAi knockdown at specific timepoints to fully determine the role of these proposed downstream targets of *oskar*, *staufer* and *pumilo* in the persistence of LTM.

Also among the upregulated DE candidate genes were several with known functions related to the epigenetic regulation of chromatin. Chromatin regulation directly impacts gene expression

by altering the accessibility of DNA to transcription by changing between the relaxed form, euchromatin, to the more tightly packed, heterochromatin state. Chromatin structure can be regulated by: ATP-remodeling complexes to manipulate nucleosome positioning, post-translational modification of histones and replacing canonical histones with variants (Taniguchi & Moore, 2014). Within our upregulated DE gene list, we identified the increased expression of *CtBP*, a subunit of the ATP-dependent chromatin complex ToRC (Emelyanov et al., 2012). Genes with post-translational histone modifying functions such as acetylation (*E(Pc)*), and methylation (*Hmt4-20*, *Ncoa6*) were also seen to be upregulated. *E(Pc)* is of particular interest as mutants have dendrite mistargeting and expression of the histone acetyltransferase complex it composes part of, Tip60, has shown to be required for LTM maintenance past 24-hours post training (Taniguchi & Moore, 2014; Hirano et al., 2016). Two genes implicated to the replacement of canonical histone variants were also identified: *CG8677* and *his3.3b*. *CG8677* is known to form part of the chromatin remodeling factor RSF which contributes to histone H2Av replacement to aid in heterochromatin formation, potentially with aid from the Tip60 complex (Hanai et al., 2008). Histone H2av replacement has significance to learning and memory as it has been suggested that it has the capacity to mediate molecular stability required for memory retention in mice (Zovkic et al., 2014). *His3.3b* is thought to function similarly to histone H2av, and potentially plays a role in both active and bivalent promoters (Santoro and Dulac, 2015). As the epigenetic regulation of chromatin has the potential to induce sustained differences in neural networks which may be critical during later phases of LTM processes, we believe these identified upregulated DE genes are of significant interest (Zovkic, Guzman-Karlsson & Sweatt, 2013). With our results profiling gene expression changes 24h after LTM induction, I hypothesize that the upregulated DE genes we identified with known chromatin regulation functions may alter the expression of genes that are involved downstream in later LTM maintenance. This could be further studied using chromatin immunoprecipitation sequencing (ChIP-seq) to identify regions of the genome epigenetically regulated during LTM.

Table 4.1: Genes with known physical interactions to pumilo, oskar and stauften within the upregulated DE gene list. FlyBase was used to identify genes within our upregulated DE gene list that have known protein-protein or RNA-protein interactions with the known memory genes *pumilo*, *oskar* and *stauften*. For each identified DE gene, an example of a known associated function is provided.

Flybase ID	Gene Name	Interaction (Type)	Known Function
FBgn0000042	<i>Act5C</i>	Oskar (protein-protein)	Cytoskeletal/chromatin remodeling
FBgn0004907	<i>14-3-3 ζ</i>	Oskar (protein-protein)	Ras/MAPK cascade
FBgn0010300	<i>Brat</i>	Pumilo (protein-protein), stauften (RNA-protein, protein-protein)	Protein translation
FBgn0052767	<i>CG32767</i>	Staufen (RNA-protein)	Nucleic acid binding
FBgn0041605	<i>Cpx</i>	Staufen (RNA-protein)	Synaptic transmission
FBgn0004838	<i>Hrb27C</i>	Oskar (RNA-protein)	Protein translation
FBgn0285926	<i>Imp</i>	Oskar (RNA-protein)	Protein translation
FBgn0261618	<i>Larp</i>	Oskar (protein-protein)	Male meiosis
FBgn0026206	<i>Mei-P26</i>	Pumilo (RNA-protein)	Protein ubiquitination
FBgn0265297	<i>pAbp</i>	Oskar (RNA-protein), Pumilo (protein-protein)	Protein translation
FBgn0004595	<i>Pros</i>	Staufen (RNA-protein)	Neural differentiation
FBgn0004636	<i>Rapl</i>	Staufen (RNA-protein)	Small GTPase
FBgn0038826	<i>Syp</i>	Oskar (RNA-protein)	mRNA binding

4.2 Genes with lower transcript abundance 24h post-training

This study identified 30 genes differentially downregulated ($q < 0.2$, fold change > 1.3) 24h after LTM induction. GO analysis of the downregulated DE candidate gene list revealed enrichment of terms associated with voltage-gated potassium channels, which fits with what we currently know about LTM. In excitatory neurons, potassium ion channels are often expressed concurrently with sodium and calcium channels to repolarize cells after action potential firing (Shah, Hammond & Hoffman, 2010). By allowing potassium to efflux into the post-synaptic terminal after activation, these channels have the potential to inhibit LTP. Indeed, it has been shown that during normal LTP induction both slow-conductance calcium-activated potassium channels and voltage-gated potassium channels are internalized to prevent repolarization (Shah, Hammond & Hoffman, 2010). Thus, while *elk* and *shawl* have not previously been associated with LTM, I hypothesize that their downregulation in the context of LTP is a requirement for proper LTM maintenance. This could be tested using a similar adult-specific gene knockdown approach like that suggested for our proposed *oskar*, *staufen* and *pumilo* downstream targets.

4.3 Enriched known and *de novo* motifs within the promoter regions of DE genes

Among the promoter regions of the 120 genes found to be differentially expressed ($q < 0.2$, fold change > 1.3) in this study, seven DE genes were identified to have the putative CRE binding site for the transcription factor CREB (**Table 3.7**). While there was no obvious functional connection between these genes, some have previously defined roles which make them of further interest to learning and memory processes, specifically, *ctp*, *lk6* and *hk*. *Ctp*, the *Drosophila* homologue of the dynein light chain has been shown to aid in the facilitation of sensory dendrite pruning through interaction with *Ik2* and *Spn-F* (Lin et al., 2015). *Lk6* is a protein kinase dependent upon the presence of calmodulin, a protein necessary for proper LTM formation, indicating downstream targets of *Lk6* phosphorylation may also play a role in learning and memory (Kidd and Raff, 1997). Finally, *Hk* encodes a beta subunit of voltage-gated potassium channels and interacts with *eag*, an alpha subunit which has known learning and memory implications (Sokolowski, 2001). Voltage-gated potassium channel beta subunits, cannot conduct current on their own but can influence neuronal physiology by modulating the activity of alpha channels. Specifically, when alpha and beta subunits associate with one another channel inactivation has

been shown to occur (Rettig et al., 1994). As previously stated, LTP relies upon the reduction of potassium channel activity to reduce repolarization of neuronal cells, thus, upregulation of *Hk* fits within this context of LTM induction.

While CREB is not the only transcription factor involved in LTM, it is the best characterized across multiple species, however, the genes transcriptionally regulated by CREB have not been fully characterized (Zhang et al., 2005; Alberini 2009). As the presence of CRE can predict CREB-binding, I hypothesize that the DE genes identified with the putative CRE elements represent downstream targets of CREB-mediated transcription. Further study could use adult specific knockdown of CREB during LTM, followed by transcriptome-profiling using INTACT to identify genes with affected expression. This could be cross-referenced with our list of DE genes with CRE to give evidence of CREB-mediated transcription.

Using HOMER, we were also able to identify several *de novo* motifs within the promoter regions of our DE gene list. One motif, 5'-TCTCTCTCTC-3', which was found in 58.26% of DE genes is of significant interest as it displays high correlation (93%) with the transcription factor binding site for trl. Trl, trithorax-like, is a DNA binding protein that binds specifically to GAGAG motifs within promoter regions of genes and has also been shown to interact with a variety of ATP-dependent chromatin remodelers, including the fly SWI/SNF complex (SWItch, Sucrose Non-Fermentable) (Lomaev et al., 2017). Interaction with the SWI/SNF complex is relevant to learning and memory as SWI/SNF components have been shown to be mutated in patients with intellectual disability (Santen, Kriek & Attikum, 2012). In flies, the adult-specific knockdown of SWI/SNF components in the MB has been shown to produce LTM defects, indicating a role for the SWI/SNF complex in memory processes (Stone, 2017). Functionally, trl has been shown to recruit chromatin remodeling complexes to promoter regions of genes to generate a nucleosome free region which can then increase subsequent transcription of nearby genes (Okada & Hirose, 1998). While trl has not been previously implicated in LTM functioning, with its role in downstream transcriptional regulation through chromatin remodeling and enrichment of its binding site seen within the DE gene list, it is a promising candidate for further study.

4.4 Limitations

While this study improves upon the approaches of prior LTM transcriptome studies, including those by Dubnau et al. (2003) and Winbush et al. (2012), it does have multiple limitations. Inherent to RNA-sequencing experiments is the chance that reported DE genes are false positives. As learning and memory induces subtle changes in gene expression, there is a need for even greater biological resolution to be able to discern true positives. Consequently, mitigating false positives was an area of focus during both our sample collection and data analysis approaches. During sample collection, we pooled samples of ~50 flies for each biological replicate. This was a necessity not only to collect enough material for INTACT but was also a means to reduce inter-individual variation, which could impact the DE analysis. To reduce false positives during the DE analysis, our approach included removing low-count genes, as well as using a very stringent statistical methodology employed by DESeq2 (Love et al., 2014). Ultimately, however, the number of biological replicates and sequencing depth are the critical components for determining the true effect of our treatment on gene expression (Conesa et al., 2016; Ching, Huang & Garmire, 2014). Currently, our study has at least two biological replicates for each condition, with lower replicated conditions having a greater sequencing depth (**Table 3.2**). While an increase in sequencing depth does improve power, added depth beyond 10 million aligned reads has diminishing returns and increasing the number of biological replicates is a more effective strategy for reducing false positives (Liu, Zhou & White, 2014). As such, to increase the confidence of our candidate gene lists, the number of biological replicates for each condition should be increased (Conesa et al., 2016). These additional replicates, together with our current-read depth of at least 5 million counts, would improve the power of our study and help reduce intra-condition variability. This intra-condition variability is notable between our naïve fly samples and may have limited this studies ability to adequately assess gene expression changes 1h post-training (**Figure 3.4**). As such, it is believed that with the addition of more biological replicates the scope of this study can be widened to include gene expression changes 1h post-training. However, even with greater biological replicates, biological validation and further study is required to better understand the role identified DE genes play in learning and memory processes.

This study presents a data-set which captures nuclei specific to the MB, the required structure for fly memory, and improves upon the biological resolution seen in prior studies

profiling whole fly heads. However, while our study utilizes a GAL4 line targeted to allow profiling the MB, it is important to note that this driver line is predominantly expressed in KC's of the α , β , and γ lobes, lacking expression in the α' , β' lobes, and has limited expression in extrinsic MB neurons (Jenett et al., 2012). Using our sequencing results, we determined that INTACT could isolate MB nuclei (**Figure 3.3**), however, with each lobe of the MB known to play different roles during learning and memory processes, this study does not fully capture the spatial requirement of the identified DE genes. Regardless, it may be difficult to accurately identify genes required for LTM both temporally and spatially in single cell-types as the engram of memory is dynamic. Thus, by profiling multiple cell-types initially, this allows for the identification of candidate genes at a single time-point that can then be further studied using lobe-specific GAL4 lines to spatially profile our DE genes. As we have already shown that INTACT is capable of profiling LTM changes in specific subsets of tissue, this methodology could be easily applied to future studies profiling LTM in single MB cell-types.

4.5 Genetic tools for further study

This study has been designed as a hypothesis-generating RNA-sequencing experiment, with genes shown to be differentially expressed acting as candidates for further study. As such, I have suggested several hypotheses and approaches to further analyze the biological roles of our identified DE genes. Available for use in *Drosophila* are several genetic tools that could be used in the proposed future studies. These include adult-specific gene knockdown and ChIP-sequencing.

Gene knockdown mediated by RNAi is an approach used to observe biological disruptions caused *in vivo*. Pertinent to this study, candidate gene knockdown could be used to determine if gene loss impairs the courtship suppression seen from courtship conditioning. However, as proteins often have multiple functions, defining the role candidate genes play solely in LTM formation and maintenance is critical. Thus, RNAi knockdown regulated both temporally and spatially is needed to minimize potential unintended effects on developmental processes, which could produce memory perturbations solely due to developmental defects. One such method capable of enacting adult-specific knockdown of genes is the P{Switch} system which carries a RU486-inducible form of the GAL4 transcription factor to manipulate transgene expression in

both time and space (Roman et al., 2001). Recently, this system has been integrated with a MB-specific line to enact knockdown of critical components of LTM during both formation and maintenance (Mao, Roman & Davis, 2004; Hirano et al., 2016). Another genetic tool which can drive adult-specific gene knockdown is GAL80. In yeast, where both GAL4 and GAL80 derive from, GAL80 acts as a transcriptional repressor of GAL4 by binding to GAL4's activating domain (del Valle Rodriguez, Didiano & Desplan, 2013). When spliced together with a temperature-sensitive variant of a yeast-specific vacuolar ATPase subunit, the GAL80 transcriptional repressor can act to temporally restrict the expression of GAL4. Use of the GAL80 or P{Switch} systems could achieve the adult-specific knockdowns we have suggested for several of our candidate DE genes, including the proposed downstream targets of *oskar*, *staufen* and *pumilo*. Specifically, these genetic tools could be used to temporally block the expression of our candidate genes from 1h to 24h post-training, as well as, 24h to 48h post-training. This approach would provide direct biological evidence that our candidate genes play a role in either LTM formation or maintenance *in vivo*.

Among our DE genes we have identified several encoding chromatin regulators, which can affect downstream gene transcription. ChIP-sequencing is a tool that can be used to identify these epigenetically regulated genes during LTM. One specific use of ChIP-sequencing relevant to our results would be to profile the post-translational histone modifications lysine 27 of histone 3 acetylation (H3-K27ac) and lysine 4 of histone 3 mono-methylation (H3-K4me). H3-K4me, a histone modification generated in part by our DE gene *Ncoa6*, and H3-K27ac have been previously associated with active enhancer sites, acting to alter gene regulation of nearby genes (Malik et al. 2014). Additionally, both H3-K27ac and H3-K4me are enriched in enhancer regions in response to neuronal membrane depolarization and regulate activity-dependent transcription of genes critical to memory functioning (Malik et al. 2014, Zhou 2016). It has been shown that ChIP sequencing can be performed on nuclei obtained from INTACT for both H3-K27ac, as well as, H3-K4me histone marks (Henry et al., 2012). By combining ChIP-sequencing results for enhancer marks with our RNA-sequencing data, this could provide evidence for the epigenetic-regulation of some of our DE candidate genes, as well as reveal other genes with functions required for downstream LTM maintenance. Thus, ChIP-sequencing in combination with INTACT offers a compelling avenue to further understand the dynamic epigenetic regulation of chromatin seen during LTM.

4.6 Summary and Conclusions

To summarize, this study presents the first known use of the INTACT method to isolate MB-nuclei for profiling over a time course of LTM formation and maintenance. Through post-sequencing analysis of RNA extracted from whole fly heads with tissue obtained using INTACT, it was determined that INTACT can achieve MB-enriched samples. Overall, DE analysis revealed 120 genes differentially expressed ($q < 0.2$, fold change > 1.3) 24h post-training. Of these, 15 DE genes were identified as having previously been associated with learning and memory functions. This study also identifies multiple DE genes which are potentially novel LTM genes and presents several hypotheses for further validation. These include:

- 13 DE genes with known physical interactions with the previously identified LTM genes *oskar*, *staufen* and *pumilo*. I hypothesize that these DE genes act as downstream targets for RNA localization by *oskar*, *staufen* and *pumilo*, with further study required using adult-specific gene knockdown.
- Several DE genes with known functions for epigenetically regulating chromatin. I hypothesize that these may epigenetically mediate the transcription of genes required for later LTM maintenance. Further study is suggested to include ChIP-sequencing of the enhancer-specific histone modifications H3-K4me and H3-K27ac to discover the identity of these genes.
- Seven DE genes with CRE elements located within 2kb of the TSS. I hypothesize that these genes may be downstream effectors of CREB-mediated transcription. This could be validated by comparing these seven DE genes with the results of transcriptome profiling of adult-specific CREB knockdown during LTM.

In conclusion, this study improves upon previous transcriptome-wide studies by profiling LTM-specific tissue, to provide a rich data-set of transcriptionally-regulated LTM candidate genes for further study.

References

- Aceves-Piña, E. O., & Quinn, W. G. (1979). Learning in normal and mutant *Drosophila* larvae. *Science*, 206(4414), 93–6.
- Akalal, D. G., Yu, D., & Davis, R. L. (2011). The Long-Term Memory Trace Formed in the *Drosophila* Alpha/Beta Mushroom Body Neurons Is Abolished in Long-Term Memory Mutants. *The Journal of Neuroscience*, 31(15), 5643–5647.
- Aken, B. L., Ayling, S., Barrell, D., Clarke, L., Curwen, V., Fairley, S., ... Searle, S. M. J. (2016). The Ensembl Gene Annotation System. *Database: The Journal of Biological Databases and Curation*, 2016, baw093.
- Alberini, C. M. (2014). Transcription Factors in Long-Term Memory and Synaptic Plasticity. *Physiology Review*, 89(1), 1–46.
- Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, 31(2), 166–169.
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Retrieved from <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
- Andersen, P., Sundberg, S. H., Sveen, O., Swann, J. W., & Wigström, H. (1980). Possible mechanisms for long-lasting potentiation of synaptic transmission in hippocampal slices from guinea-pigs. *Journal of Physiology*, 302, 463–82.
- Barondes, S. H., & Jarvik, M. E. (1964). The Influence of Actinomycin-D on Brain RNA Synthesis and on Memory. *Journal of Neurochemistry*, 11(3), 187–195.
- Barrionuevo, G., & Brown, T. H. (1983). Associative long-term potentiation in hippocampal slices. *Proceedings of the National Academy of Sciences of the United States of America*, 80(23), 7347–7351.
- Benton, R., Vannice, K. S., Gomez-diaz, C., & Leslie, B. (2009). Variant ionotropic glutamate receptors as chemosensory receptors in *Drosophila*. *Cell*, 136(1), 149–162.
- Blum, A. L., Li, W., Cressy, M., & Dubnau, J. (2009). Short- and long-term memory in *Drosophila* require cAMP signaling in distinct neuron types. *Current Biology : CB*, 19(16), 1341–50.
- Bolduc, F. V., & Tully, T. (2014). Fruit flies and intellectual disability. *Fly*, 3(1), 91–104.
- Bourne, J. N., & Harris, K. M. (2011). Coordination of size and number of excitatory and inhibitory synapses results in a balanced structural plasticity along mature hippocampal CA1 dendrites during LTP. *Hippocampus*, 21(4), 354–373.

- Brand, A. H., & Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development (Cambridge, England)*, *118*(2), 401–15.
- Brunelli, M., Castellucci, V., & Kandel, E. R. (1976). Synaptic facilitation and behavioral sensitization in *Aplysia*: possible role of serotonin and cyclic AMP. *Science (New York, N.Y.)*, *194*(4270), 1178–81.
- Burnet, B., Connolly K., & Dennis, L. (1971). The function and processing of auditory information in the courtship behaviour of *Drosophila melanogaster*. *Animal Behaviour*, *19*(2), 409-415.
- Chen, C., Wu, J., Lin, H., Pai, T., Fu, T., Wu, C., ... Chiang, A. (2012). Visualizing Long-Term Memory Formation in Two Neurons of the *Drosophila* Brain. *Science*, *335*(February), 678–686.
- Ching, T., Huang, S., & Garmire, L. X. (2014). Power analysis and sample size estimation for RNA-Seq differential expression. *Bioinformatics*, *20*, 1684–1696.
- Choi, K. W., Smith, R. F., Buratowski, R. M., & Quinn, W. G. (1991). Deficient protein kinase C activity in turnip, a *Drosophila* learning mutant. *The Journal of Biological Chemistry*, *266*(24), 15999–606.
- Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., ... Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biology*, *17*(1), 13.
- Crittenden, J. R., Skoulakis, E. M. C., Han, K.-A., Kalderon, D., & Davis, R. L. (1998). Tripartite Mushroom Body Architecture Revealed by Antigenic Markers. *Learning & Memory*, *5*(1), 38–51.
- Crocker, A., Guan, X.-J., Murphy, C. T., & Murthy, M. (2016). Cell Type-Specific Transcriptome Analysis in the *Drosophila* Mushroom Body Reveals Memory-Related Changes in Gene Expression. *Cell Reports*, *15*(7), 1580–1596.
- Davis, R. L. (2011). Review Traces of *Drosophila* Memory. *Neuron*, *70*(1), 8–19.
- Deal, R. B., & Henikoff, S. (2010). A simple method for gene expression and chromatin profiling of individual cell types within a tissue. *Developmental Cell*, *18*(6), 1030–1040.
- de Belle, J. S., & Heisenberg, M. (1994). Associative odor learning in *Drosophila* abolished by chemical ablation of mushroom bodies. *Science*.
- De Roo, M., Klauser, P., & Muller, D. (2008). LTP promotes a selective long-term stabilization and clustering of dendritic spines. *PLoS Biology*, *6*(9), 1850–1860.
- del Valle Rodriguez, A., Didiano, D., & Desplan, C. (2013). Power tools for gene expression and clonal analysis in *Drosophila*. *Nature Methods*, *9*(1), 47–55.

- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15–21.
- Domjan, M. (2005). Pavlovian Conditioning: A Functional Perspective. *Annual Review of Psychology*, 56, 179–206.
- Drain, P., Folkers, E., & Quinn, W. C. (1991). CAMP-Dependent Protein of learning in Transgenic Kinase and the Disruption Flies. *Cell Press*, 6, 71–82.
- Dubnau, J., Chiang, A. S., Grady, L., Barditch, J., Gossweiler, S., McNeil, J., ... Tully, T. (2003). The staufen/pumilio pathway is involved in *drosophila* long-term memory. *Current Biology*, 13(4), 286–296.
- Dudai, Y., Jan, Y. N., Byers, D., Quinn, W. G., & Benzer, S. (1976). dunce, a mutant of *Drosophila* deficient in learning. *Proceedings of the National Academy of Sciences of the United States of America*, 73(5), 1684–8.
- Ejima, A., Smith, B. P. C., Lucas, C., van der Goes van Naters, W., Miller, C. J., Carlson, J. R., ... Griffith, L. C. (2007). Generalization of Courtship Learning in *Drosophila* Is Mediated by cis-Vaccenyl Acetate. *Current Biology*, 17(7), 599–605.
- Emelyanov, A. V, Vershilova, E., Ignatyeva, M. A., Pokrovsky, D. K., Lu, X., Konev, A. Y., & Fyodorov, D. V. (2012). Identification and characterization of ToRC , a novel ISWI-containing ATP-dependent chromatin assembly complex. *Genes and Development*, 26(6), 603–614.
- Folkers, E., Drain, P., & Quinn, W. G. (1993). Radish, a *Drosophila* mutant deficient in consolidated memory. *Proceedings of the National Academy of Sciences of the United States of America*, 90(17), 8123–7.
- Frank, D. A., & Greenberg, M. E. (1994). CREB: A mediator of long-term memory from mollusks to mammals. *Cell*, 79(1), 5–8.
- Frey, U., Huang, Y. Y., & Kandel, E. R. (1993). Effects of cAMP simulate a late stage of LTP in hippocampal CA1 neurons. *Science*, 260(5114), 1661–4.
- Frey, P., Fitzsimons, H. (2017). The ERM protein Moesin is essential for neuronal morphogenesis and long-term memory in *Drosophila*. *Molecular Brain*, 10(41).
- Gramates, L. S., Marygold, S. J., Santos, G., Urbano, J., Antonazzo, G., Matthews, B. B., ... Zhou, P. (2017). FlyBase at 25 : looking to the future. *Nucleic Acids Research*, 45(October 2016), 663–671.
- Griffith, L. C., & Ejima, A. (2009). Courtship learning in *Drosophila melanogaster*: diverse plasticity of a reproductive behavior. *Learning & Memory (Cold Spring Harbor, NY)*, 16(12), 743–750.

- Grillet, M., Dartevielle, L., & Ferveur, J.-F. (2006). A *Drosophila* male pheromone affects female sexual receptivity. *Proceedings. Biological Sciences / The Royal Society*, 273(1584), 315–323.
- Grosjean, Y., Guenin, L., Bardet, H., & Ferveur, J. (2007). Prospero Mutants Induce Precocious Sexual Behavior in *Drosophila* Males. *Behaviour Genetics*, 37(4), 575–584.
- Groves, P., & Thompson, R. (1970). Habituation: a dual-process theory. *Psychological Review*, 77, 419–450.
- Hanai, K., Furuhashi, H., Yamamoto, T., Akasaka, K., & Hirose, S. (2008). RSF Governs Silent Chromatin Formation via Histone H2Av Replacement. *PLoS Genetics*, 4(2).
- Hawkins, R. D., Kandel, E. R., & Bailey, C. H. (2006). Molecular mechanisms of memory storage in *Aplysia*. *The Biological Bulletin*, 210(3), 174–191.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., ... Christopher, K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities, 38(4), 576–589.
- Heisenberg, M. (1998). What do the mushroom bodies do for the insect brain? an introduction. *Learning & Memory*, 5(1-2), 1–10.
- Heraud-Farlow, J & Kiebler, M. (2014). The multifunctional Staufen proteins: conserved roles from neurogenesis to synaptic plasticity. *Trends in Neurosciences*, 37(9), 470-479.
- Hill, T. C., & Zito, K. (2013). LTP-Induced Long-Term Stabilization of Individual Nascent Dendritic Spines. *Journal of Neuroscience*, 33(2), 678–686.
- Hirano, Y., Ihara, K., Masuda, T., Yamamoto, T., Iwata, I., Takahashi, A., ... Saitoe, M. (2016). Shifting transcriptional machinery is required for long-term memory maintenance and modification in *Drosophila* mushroom bodies. *Nature Communications*, 7, 13471.
- Holtmaat, A., & Svoboda, K. (2009). Experience-dependent structural synaptic plasticity in the mammalian brain. *Nature Reviews. Neuroscience*, 10(9), 647–658.
- Hu, Y., Sopko, R., Foos, M., Kelley, C., Flockhart, I., Ammeux, N., ... Mohr, S. E. (2013). FlyPrimerBank: an online database for *Drosophila melanogaster* gene expression analysis and knockdown evaluation of RNAi reagents. *G3*, 3(9), 1607–16.
- Huang, Y. Y., & Kandel, E. R. (1994). Recruitment of long-lasting and protein kinase A-dependent long-term potentiation in the CA1 region of hippocampus requires repeated tetanization. *Learning & Memory*, 1(1), 74–82.
- Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44–57.

- Jenett, A., Rubin, G. M., Ngo, T. T. B., Shepherd, D., Murphy, C., Dionne, H., ... Zugates, C. T. (2012). A GAL4-Driver Line Resource for *Drosophila* Neurobiology. *Cell Reports*, 2(4), 991–1001.
- Kandel, E. R. (2001). The molecular biology of memory storage: A dialogue between gene and synapses. *Science*, 294(5544), 1030–1038.
- Keleman, K., Vrontou, E., Krüttner, S., Yu, J. Y., Kurtovic-Kozaric, A., & Dickson, B. J. (2012). Dopamine neurons modulate pheromone responses in *Drosophila* courtship learning. *Nature*, 489(7414), 145–9.
- Kidd, D., & Raff, J. W. (1997). LK6, a short-lived protein kinase in *Drosophila* that can associate with microtubules and centrosomes. *Journal of Cell Science*, 219, 209–219.
- Koemans, T. S., Oppitz, C., Donders, R. A. T., Bokhoven, H. Van, Schenck, A., Keleman, K., & Kramer, J. M. (2017). *Drosophila* Courtship Conditioning as a Measure of Learning and Memory. *Journal of Visualized Experiments*, 124(June), 1–11.
- Krashes, M. J., Keene, A. C., Leung, B., Armstrong, J. D., & Waddell, S. (2007). Sequential Use of Mushroom Body Neuron Subsets during *Drosophila* Odor Memory Processing. *Neuron*, 53(1), 103–115.
- Lau, H., Timber, T., Mahmoud, R., & Rankin, C. (2013). Genetic dissection of memory for associative and non-associative learning in *Caenorhabditis elegans*. *Genes, Brain and Behavior*, 12, 210–223.
- Lee, D. (2015). Global and local missions of cAMP signaling in neural plasticity, learning, and memory. *Frontiers in Pharmacology*, 6(Aug), 1–7.
- Levin, L. R., Han, P. L., Hwang, P. M., Feinstein, P. G., Davis, R. L., & Reed, R. R. (1992). The *Drosophila* learning and memory gene rutabaga encodes a Ca²⁺ calmodulin-responsive adenylyl cyclase. *Cell*, 68(3), 479–489.
- Lin, T., Pan, P., Lai, Y., Chiang, K., Hsieh, H., & Wu, Y. (2015). Spindle-F Is the Central Mediator of Iκ2 Kinase-Dependent Dendrite Pruning in *Drosophila* Sensory Neurons. *PLOS Genetics*, 5(11), 1–26.
- Liu, Y., Zhou, J., & White, K. P. (2014). RNA-seq differential expression studies: More sequence or more replication? *Bioinformatics*, 30(3), 301–304.
- Lisman, J. (1994). The CaM kinase II hypothesis for the storage of synaptic memory. *Trends in Neurosciences*, 17(10), 406–412.
- Livingstone, M. S., Sziber, P. P., & Quinn, W. G. (1984). Loss of Calcium Calmodulin Responsiveness in Adenylate-Cyclase of Rutabaga, a *Drosophila* Learning Mutant. *Cell*, 37(1), 205–215.

- Lomaev, D., Mikhailova, A., Erokhin, M., Shaposhnikov, A. V., Moresco, J. J., Blokhina, T., ... Chetverina, D. (2017). The GAGA factor regulatory network: Identification of GAGA factor associated proteins. *PLoS ONE*, *12*(3), 1–20.
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(550), 1–21.
- Malinow, R., & Malenka, R. C. (2002). AMPA Receptor Trafficking and Synaptic Plasticity. *Annual Review of Neuroscience*, *25*(1), 103–126.
- Malik, A. N., Vierbuchen, T., Hemberg, M., Rubin, A. A., Ling, E., Couch, C. H., ... Greenberg, M. E. (2014). Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nature Neuroscience*, *17*(10), 1330–1339.
- Malik, B. R., & Hodge, J. J. L. (2014). CASK and CaMKII function in *Drosophila* memory. *Frontiers in Neuroscience*, *8*(June), 1–7.
- Manning, A. (1960). The Sexual Behaviour of Two Sibling *Drosophila* Species. *Behaviour*, *15*(1-2), 123–145.
- Mao, Z., Roman, G., Zong, L., & Davis, R. L. (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*(1), 1–6.
- McBride, S. M., Giuliani, G., Choi, C., Krause, P., Correale, D., Watson, K., ... Siwicki, K. K. (1999). Mushroom Body Ablation Impairs Short-Term Memory and Long-Term Memory of Courtship Conditioning in *Drosophila melanogaster*. *Neuron*, *24*(4), 967–977.
- Michael, D., Martin, K., Seger, R., Ning, M., Baston, R., & Kandel, E. (1998). Repeated pulses of serotonin required for long-term facilitation activate mitogen-activated protein kinase in sensory neurons of *Aplysia*. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(February), 1864–1869.
- Montague, S. A., & Baker, B. S. (2016). Memory elicited by courtship conditioning requires mushroom body neuronal subsets similar to those utilized in appetitive memory. *PLoS ONE*, *11*(10), 1–24.
- Montarolo, P. G., Goelet, P., Castellucci, V. F., Morgan, J., Kandel, E. R., & Schacher, S. (1986). A critical period for macromolecular synthesis in long-term heterosynaptic facilitation in *Aplysia*. *Science*, *234*(4781), 1249–54.
- Montminy, M. R., Sevarino, K. A., Wagner, J. A., Mandel, G., & Goodman, R. H. (1986). Identification of a cyclic-AMP-responsive element within the rat somatostatin gene. *Proceedings of the National Academy of Sciences of the United States of America*, *83*(18), 6682–6. <https://doi.org/10.1073/pnas.83.18.6682>

- Muller, U. (1997). Neuronal cAMP-Dependent Protein Kinase Type II Is Concentrated in Mushroom Bodies of *Drosophila melanogaster* and the Honeybee *Apis mellifera*. *Journal of Neurobiology*, 33(1), 33–44.
- Pang, Z. P., Cao, P., Xu, W., & Su, T. C. (2010). Calmodulin Controls Synaptic Strength via Presynaptic Activation of Calmodulin Kinase II. *The Journal of Neuroscience*, 30(11), 4132–4142.
- Pfeiffer, B. D., Jenett, A., Hammonds, A. S., Ngo, T.-T. B., Misra, S., Murphy, C., ... Rubin, G. M. (2008). Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, 105(28), 9715–9720.
- Philip, N., Acevedo, S. F., & Skoulakis, E. M. C. (2001). Conditional Rescue of Olfactory Learning and Memory Defects in Mutants of the 14-3-3zeta Gene leonardo. *Journal of Neuroscience*, 21(21), 8417–8425.
- Quinn, W. G., Harris, W. A., & Benzer, S. (1974). Conditioned behavior in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences of the United States of America*, 71(3), 708–12.
- R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from <http://www.R-project.org/>
- Rall, T. W., Sutherland, E. W., & Berthet, J. (1956). The Relationship of Epinephrine and Gluagon to Liver Phosphorylase. *J Biol Chem*, 224, 463–475.
- Reiter, L. T., Potocki, L., Chien, S., Gribskov, M., & Bier, E. (2001). A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Research*, 11(6), 1114–25.
- Rettig, J., Heinemann, S., Wunder, F., Lorra, C., Parcej, D., Dolly, J., & Pongs, O. (1994). Inactivation properties of voltage-gated K⁺ channels altered by presence of beta-subunit. *Nature*, 369, 289–294.
- Roman, G., Endo, K., Zong, L., & Davis, R. L. (2001). P {Switch}, a system for spatial and temporal control of gene expression in *Drosophila melanogaster*. *Proceedings of the National Academy of Sciences*, 98(22), 12602–12607.
- Roote, John; Prokop, Andreas (2017): How to design a genetic mating scheme: a basic training package for *Drosophila* genetics. figshare. DOI: 10.6084/m9.figshare.106631.v31
- Sambandan, D., Yamamoto, A., Mackay, T. F. C., & Anholt, R. R. H. (2006). Dynamic Genetic Interactions Determine Odor-Guided Behavior in *Drosophila melanogaster*. *Genetics*, 1363(November), 1349–1363.

- Santen, G. W. E., Kriek, M., & Attikum, H. Van. (2012). SWI/SNF complex in disorder. *Epigenetics*, (November), 1219–1224.
- Santoro, S. W., Dulac, C., & Biology, C. (2015). Histone variants and cellular plasticity. *Trends in Genetics*, 31(9), 516–527.
- Schmieder, R., & Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics*, 27(6), 863–864.
- Shah, M., Hammond, R., & Hoffman, D. (2010). Dendritic Ion Channel Trafficking and Plasticity. *Trends in Neuroscience*, 33(7), 307–316.
- Shors, T. J., & Matzel, L. D. (1997). Long-term potentiation: what's learning got to do with it? *The Behavioral and Brain Sciences*, 20(4), 597–655. DOI: 10.1017/S0140525X97001593
- Siegel, R. W., & Hall, J. C. (1979). Conditioned responses in courtship behavior of normal and mutant *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, 76(7), 3430–3434.
- Sigurdsson, T., Doyère, V., Cain, C. K., & LeDoux, J. E. (2007). Long-term potentiation in the amygdala: A cellular mechanism of fear learning and memory. *Neuropharmacology*, 52(1), 215–227.
- Smolik, S. M., Rose, R. E., & Goodman, R. H. (1992). A cyclic AMP-responsive element-binding transcriptional activator in *Drosophila melanogaster*, dCREB-A, is a member of the leucine zipper family. *Molecular and Cellular Biology*, 12(9), 4123–31.
- Sokolowski, M. B. (2001). *Drosophila*: Genetics meets behaviour. *Nature Reviews Genetics*, 2(November), 8879–890.
- Spieth, H. T. (1974). Courtship Behavior in *Drosophila*. *Annual Review of Entomology*, 19(1), 385–405.
- Steiner, F. A., Talbert, P. B., Kasinathan, S., Deal, R. B., & Henikoff, S. (2012). Cell-type-specific nuclei purification from whole animals for genome-wide expression and chromatin profiling. *Genome Research*, 22(4), 766–777.
- Stone, M. (2017). Regulation of Learning and Memory by the *Drosophila melanogaster* SWI/SNF complex. Electronic Thesis and Dissertation Repository. 4753. <http://ir.lib.uwo.ca/etd/4753>
- Sweatt JD. (2010). Mechanisms of memory. Elsevier, Burlington, MA
- Taniguchi, H., & Academy, P. (2014). Chromatin regulators in neurodevelopment and disease: Analysis of fly neural circuits provides insights. *Bioessays*, 36(9), 872–883.

- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B*, 63(2), 411–423.
- Tomchik, S., & Davis, R. (2009). Dynamics of learning-related cAMP signaling and stimulus integration in the *Drosophila* olfactory pathway. *Neuron*, 64(4), 510–521.
- Trachtenberg, J. T., Chen, B. E., Knott, G. W., Feng, G., Sanes, J. R., Welker, E., & Svoboda, K. (2002). Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, 420(6917), 788–794.
- Trannoy, S., Redt-Clouet, C., Dura, J. M., & Preat, T. (2011). Parallel processing of appetitive short- and long-term memories in *Drosophila*. *Current Biology*, 21(19), 1647–1653.
- Tully, T., Bourtschouladze, R., Scott, R., & Tallman, J. (2003). Targeting the CREB pathway for memory enhancers. *Nature Reviews. Drug Discovery*, 2(4), 267–277. <https://doi.org/10.1038/nrd1061>
- Tully, T., Preat, T., Boynton, S. C., & Del Vecchio, M. (1994). Genetic Dissection in *Drosophila* of Consolidated Memory. *Cell*, 79, 35–47.
- Walkinshaw, E., Gai, Y., Farkas, C., Richter, D., Nicholas, E., Keleman, K., & Davis, R. L. (2015). Identification of genes that promote or inhibit olfactory memory formation in *Drosophila*. *Genetics*, 199(4), 1173–1182.
- Winbush, A., Reed, D., Chang, P. L., Nuzhdin, S. V., Lyons, L. C., & Arbeitman, M. N. (2012). Identification of gene expression changes associated with long-term memory of courtship rejection in *Drosophila* males. *G3*, 2(11), 1437–45.
- Wolf, F. W., Eddison, M., Lee, S., Cho, W., & Heberlein, U. (2007). GSK-3 / Shaggy regulates olfactory habituation in *Drosophila*. *Proceedings of the National Academy of Sciences*, 104(11), 4653–4657.
- Zanini, D., Jallon, J., Rabinow, L., & Samson, M. (2012). Deletion of the *Drosophila* neuronal gene found in neurons disrupts brain anatomy and male courtship. *Gene., Brain and Behavior*, 11(March), 819–827.
- Zhang, J., Little, C. J., Tremmel, D. M., Yin, J. C. P., & Wesley, C. S. (2013). Notch-Inducible Hyperphosphorylated CREB and Its Ultradian Oscillation in Long-Term Memory Formation. *The Journal of Neuroscience*, 33(31), 12825–12834.
- Zhang, J., Tanenhaus, A. K., Davis, J. C., Hanlon, B. M., & Yin, J. C. P. (2015). Spatio-temporal in vivo recording of dCREB2 dynamics in *Drosophila* long-term memory processing. *Neurobiology of Learning and Memory*, 118, 80–88.

- Zhang, X., Odom, D. T., Koo, S., Conkright, M. D., Canettieri, G., Best, J., ... Montminy, M. (2005). Genome-wide analysis of cAMP-response element binding protein occupancy, phosphorylation, and target gene activation in human tissues. *Proceedings of the National Academy of Sciences*, *102*(12), 4459–4464.
- Zhou, J., & Troyanskaya, O. G. (2016). Probabilistic modelling of chromatin code landscape reveals functional diversity of enhancer-like chromatin states. *Nat Communications*, *7*, 1–9.
- Zovkic, I. B., Paulukaitis, B. S., Day, J. J., Etikala, D. M., & Sweatt, J. D. (2014). Histone H2A.Z subunit exchange controls consolidation of recent and remote memory. *Nature*, *515*(7528), 582–586.

Appendix A: Supplementary Tables

Supplementary Table 1: Differentially expressed genes ($q < 0.2$, fold change > 1.3 , < 0.77) between 24h post-training and 1h post-training, sorted by q value. Rounded normalized counts are provided for individual samples.

Flybase ID	Gene Name	Normalized Counts					q value	Fold difference
		E1-1	E1-2	E24-1	E24-2	E24-3		
FBgn0011760	<i>ctp</i>	11805	14058	23244	23961	22675	0.000	1.747
FBgn0013342	<i>nSyb</i>	30098	36927	71205	48174	60298	0.002	1.719
FBgn0027339	<i>jim</i>	34272	40602	57507	53275	58744	0.002	1.488
FBgn0000042	<i>Act5C</i>	15990	16367	36852	23497	25999	0.005	1.695
FBgn0000253	<i>Cam</i>	34399	52546	125212	70205	72814	0.005	1.894
FBgn0004838	<i>Hrb27C</i>	15029	19933	33125	29329	25943	0.005	1.628
FBgn0020238	<i>14-3-3epsilon</i>	11424	15689	30794	20469	24437	0.005	1.759
FBgn0026206	<i>mei-P26</i>	26641	40468	62288	51438	57839	0.005	1.644
FBgn0030758	<i>CanA-14F</i>	11827	16628	25676	23691	23065	0.005	1.637
FBgn0261710	<i>nocte</i>	11973	12430	22830	17747	18626	0.008	1.569
FBgn0021872	<i>Xbp1</i>	4552	4917	8617	7399	8725	0.009	1.661
FBgn0052767	<i>CG32767</i>	7112	9485	15196	12663	14558	0.016	1.630
FBgn0001122	<i>Galphao</i>	12894	16891	30140	20309	25753	0.021	1.626
FBgn0032817	<i>CG10631</i>	2915	3696	6597	6601	6396	0.022	1.793
FBgn0004595	<i>pros</i>	92476	110821	127537	149486	162619	0.023	1.419
FBgn0050361	<i>mtt</i>	9830	8739	5411	5803	6065	0.024	0.643
FBgn0261618	<i>larp</i>	12782	14382	19055	20236	24973	0.032	1.527
FBgn0266100	<i>CG44837</i>	8182	7114	4302	4781	4923	0.039	0.638
FBgn0031835	<i>CG11319</i>	41624	46552	28385	28157	35957	0.045	0.712
FBgn0045823	<i>vsg</i>	4170	3709	10414	5893	7909	0.045	1.821
FBgn0267668	<i>CR46006</i>	11226	11522	7718	6747	7818	0.045	0.673
FBgn0026575	<i>hang</i>	5041	4935	11397	7503	7912	0.053	1.661
FBgn0036583	<i>CG13055</i>	5145	5089	9270	6795	10738	0.053	1.634
FBgn0039808	<i>CG12071</i>	3517	4913	8838	5958	8491	0.053	1.689
FBgn0086675	<i>fne</i>	17459	21598	28678	27438	31908	0.053	1.463
FBgn0262730	<i>dtm</i>	12538	18382	33784	18966	28341	0.053	1.634
FBgn0034570	<i>CG10543</i>	3710	4554	7460	5947	7496	0.054	1.592
FBgn0030328	<i>Amun</i>	4349	3625	11493	5334	7433	0.060	1.778
FBgn0033872	<i>CG6329</i>	6445	8612	11769	12056	10916	0.065	1.486
FBgn0019661	<i>roX1</i>	103861	108937	88294	69416	68689	0.065	0.721
FBgn0035481	<i>CG12605</i>	12163	14493	22794	16138	23159	0.067	1.496

FBgn0262735	<i>Imp</i>	10435	16579	29667	16854	24237	0.067	1.622
FBgn0041094	<i>scyl</i>	7436	10159	16442	15845	13653	0.073	1.622
FBgn0031698	<i>Ncoa6</i>	6073	7291	9623	9522	10634	0.073	1.442
FBgn0037705	<i>mura</i>	12661	19848	33107	20435	29118	0.073	1.588
FBgn0261262	<i>CG42613</i>	7520	9174	15106	11107	12861	0.073	1.503
FBgn0261548	<i>prage</i>	21639	26168	13075	12079	19665	0.073	0.655
FBgn0051140	<i>CG31140</i>	23173	24649	43779	27879	35366	0.075	1.446
FBgn0034789	<i>PIP5K59B</i>	5590	6053	8677	8275	8450	0.078	1.416
FBgn0010247	<i>Parp</i>	23704	24666	19361	12497	15730	0.082	0.679
FBgn0013686	<i>mt:lrRNA</i>	23855	2268	0	0	0	0.086	0.593
FBgn0017581	<i>Lk6</i>	31244	35573	68880	48989	39316	0.086	1.501
FBgn0266019	<i>rudhira</i>	8332	9122	29050	13860	10154	0.086	1.745
FBgn0000581	<i>E(Pc)</i>	5906	7545	9114	10497	10412	0.093	1.439
FBgn0022382	<i>Pka-R2</i>	44429	58809	71744	64146	72286	0.093	1.326
FBgn0264443	<i>CG43861</i>	9171	8957	5806	5166	6887	0.093	0.681
FBgn0000273	<i>Pka-C1</i>	57905	69190	88540	72719	96357	0.096	1.331
FBgn0000557	<i>eEF1alpha2</i>	17666	23710	28601	26514	32773	0.096	1.384
FBgn0003371	<i>sgg</i>	18769	22770	35677	24531	29870	0.096	1.408
FBgn0004401	<i>Pep</i>	7092	6857	14627	9947	9371	0.096	1.529
FBgn0011481	<i>Ssdp</i>	4196	4649	6921	6197	7142	0.096	1.470
FBgn0023388	<i>Dap160</i>	4650	4991	11313	7126	6976	0.096	1.607
FBgn0050158	<i>CG30158</i>	31055	34398	19337	22048	27689	0.096	0.720
FBgn0000382	<i>csw</i>	14878	18181	26575	22753	20914	0.099	1.384
FBgn0058178	<i>CG40178</i>	72459	82812	64853	43500	57726	0.104	0.728
FBgn0004828	<i>His3.3B</i>	6512	9740	19924	10225	12506	0.111	1.601
FBgn0031627	<i>CG15630</i>	11622	9863	7736	6916	7874	0.111	0.719
FBgn0033958	<i>jef</i>	7310	7380	13518	8848	11399	0.111	1.468
FBgn0034802	<i>CNBP</i>	4059	5067	8952	7455	6642	0.111	1.563
FBgn0260995	<i>dpr21</i>	48866	56553	42994	30017	40853	0.111	0.734
FBgn0263198	<i>Acn</i>	3474	4519	8605	9006	4610	0.111	1.646
FBgn0001085	<i>fz</i>	18837	23073	16520	13348	14855	0.120	0.727
FBgn0026577	<i>CG8677</i>	3314	5298	11295	6879	6210	0.120	1.657
FBgn0259994	<i>CG42492</i>	14443	23290	33226	24286	27863	0.120	1.445
FBgn0261113	<i>Xrp1</i>	19049	19628	21829	28722	29220	0.120	1.349
FBgn0004636	<i>Rap1</i>	4451	3029	8838	5489	6062	0.121	1.629
FBgn0004656	<i>fs(1)h</i>	15370	20078	31657	21177	24398	0.121	1.407
FBgn0011589	<i>Elk</i>	21399	28405	18851	14365	18867	0.121	0.717
FBgn0086901	<i>cv-c</i>	14228	13420	10259	10879	9917	0.121	0.760
FBgn0265297	<i>pAbp</i>	12935	14549	19852	30508	15455	0.121	1.510
FBgn0036451	<i>CG9425</i>	8727	9024	15334	10095	14635	0.126	1.444
FBgn0262124	<i>uex</i>	13873	14715	11852	7522	9533	0.126	0.699

FBgn0035253	<i>CG7971</i>	18971	18994	29050	24008	22590	0.126	1.307
FBgn0003093	<i>Pkc98E</i>	11411	13787	23088	14486	19080	0.128	1.439
FBgn0261934	<i>dikar</i>	11421	12749	18725	14390	16637	0.129	1.345
FBgn0264490	<i>Eip93F</i>	28817	37020	43137	44098	48253	0.130	1.344
FBgn0000308	<i>chic</i>	5409	7521	12745	8177	9746	0.142	1.490
FBgn0001218	<i>Hsc70-3</i>	18991	14387	36678	21348	20847	0.142	1.485
FBgn0011828	<i>Pxn</i>	14126	13989	9084	8331	12106	0.142	0.721
FBgn0023179	<i>amon</i>	15608	17427	10318	13208	12772	0.142	0.747
FBgn0035625	<i>Blimp-1</i>	8405	7632	11193	10475	10676	0.142	1.321
FBgn0085395	<i>Shawl</i>	13219	15850	10990	8382	10283	0.142	0.705
FBgn0263396	<i>sqd</i>	11178	17929	25305	19497	20250	0.148	1.431
FBgn0020309	<i>crol</i>	10446	11971	17551	12528	16965	0.170	1.362
FBgn0000008	<i>a</i>	15448	17673	13255	10073	13123	0.175	0.750
FBgn0028704	<i>Nckx30C</i>	32623	46496	56746	47508	53628	0.175	1.308
FBgn0259241	<i>CG42339</i>	20444	25036	18420	14890	16088	0.175	0.743
FBgn0266101	<i>CG44838</i>	4378	6109	8521	6876	8049	0.175	1.424
FBgn0004432	<i>Cyp1</i>	3091	4134	6256	6941	4859	0.175	1.520
FBgn0030243	<i>CG2186</i>	4972	5718	7406	7767	8015	0.175	1.393
FBgn0267431	<i>Myo81F</i>	43373	44279	40489	23317	29791	0.175	0.733
FBgn0000259	<i>CkIIbeta</i>	2777	5508	8755	6258	6607	0.180	1.552
FBgn0025639	<i>Hmt4-20</i>	5873	4901	6885	8087	7940	0.180	1.373
FBgn0036663	<i>CG9674</i>	11972	11972	7988	9438	9089	0.180	0.754
FBgn0041605	<i>cpx</i>	75060	116108	136393	114972	131353	0.183	1.310
FBgn0004103	<i>Pp1-87B</i>	3216	4274	11223	4334	5537	0.185	1.603
FBgn0004907	<i>14-3-3zeta</i>	35321	43693	55931	54208	48027	0.185	1.308
FBgn0011206	<i>bol</i>	19890	25796	35917	27494	29014	0.185	1.321
FBgn0264006	<i>dysc</i>	38380	55226	63439	53694	71069	0.185	1.315
FBgn0265296	<i>Dscam2</i>	13868	20690	27863	19663	26121	0.185	1.371
FBgn0027567	<i>CG8108</i>	6469	9192	19049	12034	8119	0.188	1.515
FBgn0038826	<i>Syp</i>	13756	20359	22249	28281	23064	0.188	1.388
FBgn0085478	<i>CG34449</i>	5962	8178	15112	8129	10075	0.188	1.470
FBgn0263220	<i>Hk</i>	7379	8716	13123	10919	9586	0.190	1.353
FBgn0052000	<i>CG32000</i>	13590	17512	12799	10385	10241	0.191	0.738
FBgn0263072	<i>CG43347</i>	7239	8144	11565	9868	12120	0.191	1.394
FBgn0264693	<i>ens</i>	7183	8018	9743	10949	9930	0.195	1.315
FBgn0010300	<i>brat</i>	13053	16094	20949	16621	21598	0.196	1.322
FBgn0261261	<i>plx</i>	11259	11257	7502	8845	9114	0.196	0.769
FBgn0261403	<i>sxc</i>	5777	7573	3937	4475	4642	0.196	0.691
FBgn0263780	<i>CG17684</i>	110741	127423	109225	54477	83710	0.196	0.720
FBgn0015558	<i>tty</i>	9331	12499	19702	11593	16639	0.200	1.399
FBgn0021800	<i>Reph</i>	6185	6631	13219	7447	8545	0.200	1.433

Supplementary Table 2: Differentially expressed genes ($q < 0.2$, fold change > 1.3 , < 0.77) between 24h post-training and naive, sorted by q value. Rounded normalized counts are provided for individual samples.

Flybase ID	Gene Name	Normalized Counts					q value	Fold difference
		E1-1	E1-2	E24-1	E24-2	E24-3		
FBgn0011760	<i>ctp</i>	15093	12854	23244	23961	22675	0.004	1.625
FBgn0261262	<i>CG42613</i>	6482	7626	15106	11107	12861	0.016	1.745
FBgn0000253	<i>Cam</i>	48284	44482	125212	70205	72814	0.018	1.792
FBgn0019661	<i>roX1</i>	132454	102337	88294	69416	68689	0.018	0.658
FBgn0086677	<i>jeb</i>	16943	19870	29379	26775	25872	0.018	1.458
FBgn0030758	<i>CanA-14F</i>	14317	15906	25676	23691	23065	0.032	1.549
FBgn0266410	<i>CG45050</i>	10226	10825	16197	14522	15786	0.034	1.444
FBgn0026206	<i>mei-P26</i>	35834	38145	62288	51438	57839	0.058	1.503
FBgn0020238	<i>14-3-3epsilon</i>	16112	15029	30794	20469	24437	0.084	1.555
FBgn0260995	<i>dpr21</i>	59073	53383	42994	30017	40853	0.084	0.691
FBgn0011828	<i>Pxn</i>	16093	14849	9084	8331	12106	0.092	0.661
FBgn0027339	<i>jim</i>	43432	41602	57507	53275	58744	0.092	1.317
FBgn0020496	<i>CtBP</i>	6086	9265	16838	10132	13960	0.098	1.643
FBgn0265297	<i>pAbp</i>	12282	12890	19852	30508	15455	0.098	1.627
FBgn0013342	<i>nSyb</i>	39051	43188	71205	48174	60298	0.111	1.424
FBgn0041094	<i>scyl</i>	6437	11280	16442	15845	13653	0.111	1.613
FBgn0011661	<i>Moe</i>	6169	3915	23531	14379	3935	0.128	1.924
FBgn0052183	<i>Ccn</i>	13189	9775	6987	7537	8201	0.145	0.684
FBgn0085414	<i>dpr12</i>	47189	38333	28768	27073	35667	0.145	0.728
FBgn0004595	<i>pros</i>	102551	117190	127537	149486	162619	0.148	1.318
FBgn0001085	<i>fz</i>	23824	18954	16520	13348	14855	0.156	0.714
FBgn0263396	<i>sqd</i>	12364	15373	25305	19497	20250	0.165	1.493
FBgn0264443	<i>CG43861</i>	9442	8426	5806	5166	6887	0.165	0.690
FBgn0086675	<i>fne</i>	17770	23754	28678	27438	31908	0.166	1.383
FBgn0010247	<i>Parp</i>	26102	20802	19361	12497	15730	0.168	0.698
FBgn0004838	<i>Hrb27C</i>	21595	20901	33125	29329	25943	0.177	1.361
FBgn0031453	<i>Bacc</i>	13573	15881	20619	18146	20339	0.177	1.318
FBgn0031835	<i>CG11319</i>	42202	39637	28385	28157	35957	0.177	0.764
FBgn0050158	<i>CG30158</i>	30235	33838	19337	22048	27689	0.177	0.734
FBgn0050361	<i>mtt</i>	8554	7835	5411	5803	6065	0.177	0.721
FBgn0052767	<i>CG32767</i>	10478	9046	15196	12663	14558	0.177	1.409
FBgn0004828	<i>His3.3B</i>	9177	7529	19924	10225	12506	0.193	1.566

Supplementary Table 3: Enriched GO terms (unadjusted $p < 0.05$) for unique upregulated DE genes for biological processes.

Term	Associated Genes	Fold Enrichment
GO:0055060~asymmetric neuroblast division resulting in ganglion mother cell formation	<i>PROS, BRAT</i>	106.066
GO:0048680~positive regulation of axon regeneration	<i>CHIC, IMP</i>	53.033
GO:0007622~rhythmic behavior	<i>SGG, CKIIBETA, PKA-R2, PKA-C1</i>	32.636
GO:0008103~oocyte microtubule cytoskeleton polarization	<i>14-3-3ZETA, PKA-C1, 14-3-3EPSILON</i>	17.678
GO:0007611~learning or memory	<i>DIKAR, PPI-87B, 14-3-3ZETA, VSG, PKA-C1, MURA</i>	15.152
GO:0007140~male meiosis	<i>HIS3.3B, BOL, PABP, LARP</i>	14.142
GO:0051124~synaptic growth at neuromuscular junction	<i>IMP, SGG, DAPI60, CPX, JEB</i>	13.956
GO:0048149~behavioral response to ethanol	<i>VSG, BACC, HANG, PKA-R2, PKA-C1, MURA</i>	13.540
GO:0008355~olfactory learning	<i>DIKAR, SGG, PPI-87B, 14-3-3ZETA, VSG, PKA-C1, MURA</i>	13.258
GO:0007314~oocyte anterior/posterior axis specification	<i>SQD, MOE, PKA-C1</i>	13.258
GO:0072499~photoreceptor cell axon guidance	<i>CAM, MOE, DYSC</i>	12.728
GO:0008356~asymmetric cell division	<i>GALPHAO, PROS, BRAT</i>	12.728
GO:0045451~pole plasm oskar mRNA localization	<i>CHIC, SQD, HRB27C</i>	10.972
GO:0008285~negative regulation of cell proliferation	<i>PKC98E, PROS, BRAT, XRP1</i>	10.348
GO:0070374~positive regulation of ERK1 and ERK2 cascade	<i>PPI-87B, 14-3-3ZETA, 14-3-3EPSILON</i>	9.642
GO:0007420~brain development	<i>CHIC, PROS, BRAT</i>	9.359
GO:0042052~rhabdomere development	<i>CAM, MOE, DYSC</i>	9.359
GO:0045475~locomotor rhythm	<i>SGG, CKIIBETA, PKA-R2, PKA-C1, DYSC</i>	9.144
GO:0008582~regulation of synaptic growth at neuromuscular junction	<i>PIP5K59B, BRAT, DYSC</i>	8.839
GO:0046579~positive regulation of Ras protein signal transduction	<i>PPI-87B, 14-3-3ZETA, 14-3-3EPSILON</i>	8.600
GO:0035071~salivary gland cell autophagic cell death	<i>CYP1, CAM, EIP93F, CTP, LARP</i>	7.915
GO:0022416~chaeta development	<i>AMUN, XBPI, CTP, CTBP</i>	7.071
GO:0035220~wing disc development	<i>AMUN, XBPI, SSDP, CTP, CTBP</i>	6.629
GO:0016055~Wnt signaling pathway	<i>SGG, GALPHAO, CKIIBETA, CTBP</i>	6.527
GO:0007616~long-term memory	<i>DIKAR, PKC98E, VSG, MURA</i>	6.428
GO:0007411~axon guidance	<i>CHIC, PPI-87B, PKA-R2, PROS, BRAT, JEB, HRB27C, 14-3-3EPSILON</i>	4.849
GO:0007283~spermatogenesis	<i>CHIC, IMP, BOL, PABP, CTP</i>	4.383
GO:0000398~mRNA splicing, via spliceosome	<i>SYP, IMP, PABP, SQD, CG7971, ACN, PEP</i>	3.908
GO:0007476~imaginal disc-derived wing morphogenesis	<i>BOL, VSG, FS(1)H, CROL, CTP, PKA-C1, XRP1</i>	3.827
GO:0046331~lateral inhibition	<i>CAM, BOL, XBPI, CG31140, CROL, HRB27C</i>	3.459
GO:0048477~oogenesis	<i>SGG, PPI-87B, PABP, CTP, SQD, BRAT, PKA-C1</i>	3.406
GO:0006468~protein phosphorylation	<i>LK6, SGG, PKC98E, CAM, CKIIBETA, PKA-R2, PKA-C1</i>	3.242

Supplementary Table 4: Enriched GO terms (unadjusted $p < 0.05$) for unique upregulated DE genes for cellular components.

Term	Associated Genes	Fold Enrichment
GO:0043195~terminal bouton	<i>NSYB, CPX, JEB</i>	9.749
GO:0045179~apical cortex	<i>DAP160, PROS, BRAT</i>	9.478
GO:0005700~polytene chromosome	<i>CG8677, HIS3.3B, PPI-87B, E(PC), EIP93F, CTBP, HMT4-20</i>	6.369
GO:0071011~precatalytic spliceosome	<i>SYP, IMP, PABP, SQD, CG7971, ACN, PEP</i>	5.647
GO:0071013~catalytic step 2 spliceosome	<i>IMP, PABP, SQD, CG7971, ACN, PEP</i>	5.594
GO:0005654~nucleoplasm	<i>SGG, CAM, BACC, HRB27C, MURA, 14-3-3EPSILON</i>	5.290
GO:0005813~centrosome	<i>LK6, SGG, CAM, 14-3-3EPSILON</i>	5.290
GO:0005938~cell cortex	<i>CHIC, SGG, MOE, PROS</i>	4.892
GO:0005875~microtubule associated complex	<i>ACT5C, CYP1, CAM, PABP, 14-3-3ZETA, PROS, HSC70-3, PEP, 14-3-3EPSILON</i>	3.604
GO:0005737~cytoplasm	<i>HK, CHIC, CAM, 14-3-3ZETA, CG31140, RUDHIRA, FNE, MOE, PKA-C1, ACT5C, BOL, DAP160, PKA-R2, BRAT, LARP, ACN, HRB27C, 14-3-3EPSILON, LK6, CG34449, SGG, CYP1, CSW, PABP, SQD, IMP, MEI-P26, CKIIBETA, CTP, SCYL, MURA</i>	2.598
GO:0005829~cytosol	<i>LK6, SGG, CYP1, CAM, PABP, CKIIBETA, CPX, RUDHIRA, MOE, LARP</i>	2.430
GO:0005634~nucleus	<i>CG8677, 14-3-3ZETA, CG31140, CG10631, MOE, HMT4-20, SYP, BLIMP-1, HIS3.3B, BOL, XBPI, EIP93F, BACC, PROS, ACN, HRB27C, 14-3-3EPSILON, AMUN, SGG, CYP1, PABP, CROL, FS(1)H, CG43347, SQD, CTBP, E(PC), CKIIBETA, SSDP, HANG, MURA</i>	2.216
GO:0005886~plasma membrane	<i>CAM, PKC98E, 14-3-3ZETA, MOE, PKA-C1, NSYB, GALPHAO, DAP160, RAPI, TTY, PKA-R2, PROS, 14-3-3EPSILON</i>	2.197

Supplementary Table 5: Enriched GO terms (unadjusted $p < 0.05$) for unique upregulated DE genes for molecular functions.

Term	Associated Genes	Fold Enrichment
GO:0003730~mRNA 3'-UTR binding	<i>IMP, BOL, PABP, SQD, HRB27C</i>	19.408
GO:0003697~single-stranded DNA binding	<i>SSDP, HRB27C, PEP</i>	12.061
GO:0016301~kinase activity	<i>LK6, SGG, CKIIBETA</i>	8.237
GO:0000166~nucleotide binding	<i>SYP, IMP, BOL, PABP, DAP160, CG31140, FNE, SQD, ACN, HRB27C</i>	5.925
GO:0003729~mRNA binding	<i>SYP, BOL, PABP, FNE, SQD, HRB27C</i>	4.386
GO:0004674~protein serine/threonine kinase activity	<i>LK6, SGG, PKC98E, CKIIBETA, PKA-C1</i>	4.108
GO:0003676~nucleic acid binding	<i>BLIMP-1, BOL, CG32767, CG12071, CG12605, CROL, CG10543, CG10631, HANG, CG43347, SQD, JIM, HRB27C</i>	3.686
GO:0005515~protein binding	<i>SGG, CHIC, CSW, CAM, PABP, 14-3-3ZETA, MOE, PKA-C1, CTBP, GALPHAO, DAP160, RAPI, CTP, PKA-R2, BRAT, ACN, HRB27C, 14-3-3EPSILON</i>	3.423
GO:0046872~metal ion binding	<i>PPI-87B, PKC98E, CROL, CG31140, CG10543, CG10631, CG43347, JIM, PEP, BLIMP-1, CG32767, GALPHAO, CG12071, CG12605, CG9425</i>	2.921

Supplementary Table 6: Enriched GO terms (unadjusted $p < 0.05$) for unique downregulated DE genes for cellular components

Term	Associated Genes	Fold Enrichment
GO:0008076~voltage-gated potassium channel complex	<i>ELK, SHAWL</i>	57.495

Supplementary Table 7: Enriched GO terms (unadjusted $p < 0.05$) for unique downregulated DE genes for molecular functions.

Term	Associated Genes	Fold Enrichment
GO:0008236~serine-type peptidase activity	<i>CG11319, CG17684</i>	40.537
GO:0005249~voltage-gated potassium channel activity	<i>ELK, SHAWL</i>	42.789
GO:0022843~voltage-gated cation channel activity	<i>ELK, SHAWL</i>	96.276
GO:0005178~integrin binding	<i>CCN, PLX</i>	128.368

Curriculum Vitae

Name: Spencer Jones

Post-secondary Education and Degrees: McMaster University
Hamilton, Ontario, Canada
2011-2015 B.H.Sc.

Honours and Awards: Dean's Honour List (McMaster University)
2012-2013, 2015

Provost's Honour Roll (McMaster University)
2015

Department of Biology Graduate Student Teaching Award
(Western University)
2017

Related Work Experience Teaching Assistant (Biology 2001A/1002B/1201A/1202B)
The University of Western Ontario
2015-2017

Publications:

Jones, S.G. (2014). Potential Therapies for Cystic Fibrosis: A Look at Cystic Fibrosis Transmembrane Conductance Regulator Correctors and Potentiators. *The Meducator*, 1(25): 23-26.

Conference Presentations:

Jones, SG, Nixon KCJ & Kramer, JM (2017). Gene expression changes in the mushroom body of *Drosophila melanogaster* during a time course of long-term memory formation. Canadian *Drosophila* Research Conference. Banff, AB, Canada. [Poster].