Spring 4-30-2013

Control of Neuroendocrine Cell Physiology by a Single Transcription Factor, Drosophila Basic Helix Loop Helix Regulator DIMMED

Tarik Hadzic
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology & Biomedical Sciences
Molecular Genetics and Genomics

Dissertation Examination Committee:
Paul H. Taghert, Chair
Joseph D. Dougherty
Rafi Kopan
Kristen L. Kroll
Jason C. Mills
James B. Skeath

Control of Neuroendocrine Cell Physiology by a Single Transcription Factor, Drosophila Basic Helix Loop Helix Regulator DIMMED

by

Tarik Hadžić

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2013

St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Chapter 3.

List of Abbreviations

ACTH	adrenocorticotropic hormone
AEL	after egg laying
bHLH	basic helix loop helix
ChIP	chromatin immunoprecipitation
ChIP-chip	chromatin immunoprecipitation coupled to microarrays
DA	dopaminergic
EM	electron microscopy
ER	endoplasmic reticulum
FACS	fluorescence activated cell sorting
LEAP	Large, Episodically- Releasing, Amidating Peptide producing cells
LDCV	large dense core vesicle
NE	neuroendocrine
NS	neurosecretory
PDF	Pigment dispersing factor
PHM	peptidylglycine alpha-hydroxylating monooxygenase
REST	RE1-silencing transcription factor
RNAi	ribonucleic acid interference
RNA-Seq	massively parallel (deep) sequencing
RSP	regulated secretory pathway
SV	synaptic vesicles
TGN	trans-Golgi network

Acknowledgements

I would like to thank all my current and past laboratory mates for helpful discussions and camaraderie: Dr. Dongkook Park, Laura Duvall, Dr. Seol Hee Im, Dr. Orie T. Shafer, Neely Williams, Dr. Stacie P. Shepherd, Dr. Dongjo Kim, Dr. Weihua Li, Ms. Jennifer Trigg and Nicole Gong. I am extremely grateful to the Medical Scientist Training Program (MSTP) for funding, as well as the amazing MSTP staff and colleagues here at Washington University. Most of all, I would like to acknowledge Dr. Paul Taghert, for being the most knowledgeable and patient mentor and for allowing me to pursue creative projects and be a good scientist. My work in the Taghert laboratory was supported by an NINDS grant NS21749 (to PHT), as well as by a P30-NS057105 to Washington University. A special thanks to Dr. Sally Elgin's laboratory at the Danforth campus, particularly Mrs. Jo Wuller, for teaching me embryo incubation techniques and letting me use their facility. I would also like to acknowledge Dr. Michael Rosbash and his laboratory, especially Dr. Katharine C. Abruzzi, with whom I worked extensively while pursuing our collaboration project at Brandeis University. Finally, I would like to thank P. Michael Mastrofrancesco, without whom this thesis would never have happened. Thank you for all your love and support.

Dedication

I dedicate this thesis to my late father, Dr. Adem Hadžić, who suddenly passed away in Bosnia in 2010. My dad was one of the most positive and warmhearted people I have ever met. I was only able to move to the United States to pursue education thanks to his immense support and love. A physician for over 30 years, he was liked by patients of all walks of life, as well as by his immediate and large extended family. Additionally, I would also like to dedicate this work to my beloved dog Gidget, who passed away in 2011. Gidget was and still is the most loving dog in the world and I think of her every day.

ABSTRACT OF THE DISSERTATION

Control of Neuroendocrine Cell Physiology by a Single Transcription Factor,
Drosophila Basic Helix Loop Helix Regulator DIMMED

by

Tarik Hadžić

Doctor of Philosophy in Biology & Biomedical Sciences
Molecular Genetics and Genomics

Washington University in St. Louis, 2013

Professor Paul H. Taghert, Chairperson

Neuroendocrine cells feature a large capacity for the processing, accumulation and regulated release of bioactive peptides and peptide hormones. The ultrastructural correlate of this regulated secretory pathway is a specialized organelle, called a dense core vesicle (DCV). DCVs are typically larger than conventional, small synaptic vesicles. Past work has identified intrinsic DCV proteins (non-cargo proteins, like the processing enzyme, carboxypeptidase) or ancillary ones that play a role in DCV trafficking and exocytosis (like CAPS, the Ca2+-dependent activator protein for secretion). Currently, there is a lack of understanding of the developmental and physiological mechanisms that permit neurosecretory cells to coordinate and scale the regulated secretory pathway. In this context, the basic helix-loop-helix transcription factor *dimmed* (*dimm*) is especially important in the fruit fly *Drosophila*, but it is not involved in neuroendocrine cell fate determination.

Neuroendocrine cells require DIMM to accumulate, and process large amounts of secretory peptides, but DIMM does not target individual neuropeptide-encoding genes. Instead, we show that DIMM supports the complete resolution of NE-specific cellular properties by organizing the cellular machinery required to support a highly active RSP. The mouse orthologue Mist1 likewise plays a role in supporting the RSP of serous exocrine cells. This thesis has three goals. First, I evaluated a set of putative DIMM targets obtained by another

scientist in the lab, and ask whether or not these are direct targets of this transcription factor. To accomplish this, I use *in vivo* chromatin immunoprecipitation (ChIP) followed by measuring DIMM binding to putative DIMM enhancers by quantitative Polymerase Chain Reaction (qPCR). This work is described in Chapter 2. Secondly, I extend DIMM ChIP analysis to identify direct DIMM transcriptional targets on a genome-wide level *in vivo* in adult neurons. This was done by DIMM chromatin immunoprecipitation coupled to tiling microarrays (ChIP-chip), and also applying Fluorescence Activated Cell Sorting (FACS) and deep sequencing (RNA-seq) to define the transcriptome of DIMM neuroendocrine cells, as described in Chapter 3.

I then integrate the ChIP-chip and RNA-seq datasets to provide new viewpoints on how DIMM is used to coordinate and appropriately scale the RSP in NE cells. The intersection of the RNA-Seq and ChIP-chip data presents a list of genes that is likely to mediate the bulk of the transcriptional output of DIMM – i.e., its molecular "mechanism". In order to conduct a functional assay and thus validate the list of intersected genes, I conducted a behavioral genetic screen. DIMM-expressing cells have previously been shown to regulate sleep amount in flies. I conducted an RNA interference-based screen, in which expression of individual DIMM target genes was knocked down in DIMM neurons and the effects of this manipulation on sleep were quantified. This in vivo validation provides an important filter with which to ascribe single gene functions and gives further insights into the general mechanisms by which DIMM operates.

Chapter 1.

Introduction:

Molecular and cellular biology of neuroendocrine cells

Neuroendocrine Cells

Neuroendocrine (NE) cells are a specialized class of neurons dedicated to the production, storage and release of large amounts of peptide hormones (Park and Taghert 2009; Burgoyne and Morgan 2003). The Polish biologist Kopeć produced the first indication anywhere in the animal kingdom that the nervous system is capable of producing hormones (Kopeć 1917; Kopec 1922; Scharrer B 1987a). Kopeć found the existence of a “pupation hormone,” which is produced in the brain of the gypsy moth *Lymantria dispar* in order for the animal to execute pupation (Scharrer B 1987a). Speidel was the first to describe NE cells in the spinal cord of elasmonbranchs and bony fishes as "gland-cells of internal secretion" (Speidel 1919; Scharrer B 1987a). In addition to Kopec, Speidel, and Ernst and Bertha Scharrer, one other report identified a protein, called "Substance P," extractable from mammalian gut and the nervous tissue (Scharrer B 1987a; von Euler and Gaddum 1931).

In 1928, Ernst Scharrer reported discovering gland-like nerve cells in the hypothalamus of teleost fish (Scharrer B 1987a; Scharrer E 1952). At this time, a central dogma of neurobiology was that neuronal cells communicated strictly through electric signals (Scharrer B 1987a). Ernst Scharrer’s discovery was striking because this small group of hypothalamic cells contained impressive amounts of secretory material comparable to that of a pancreatic endocrine cell (Scharrer B 1987a; Scharrer E 1952). Therefore, the term NE neuron was coined, denoting the neural and endocrine capacities of this cell type (Scharrer E 1952; Scharrer B 1987a). These few early reports mark the early days of studying NE cells, before neuroendocrinology was a formal discipline. Ernst Scharrer was the first to formally propose the concept of neurosecretion in 1934 (Burbach 2011; Scharrer E and Scharrer B 1937).

One of the first classes of neuroendocrine neurons described were those neurons found in the supraoptic and paraventricular nuclei of the hypothalamus (Zupanc 1996). These neurons, also known as magnocellular neurons synthesize the posterior pituitary hormones oxytocin and

vasopressin (Zupanc 1996; Leng and Ludwig 2008). Oxytocin and vasopressin are transported along the axons of the hypothalamic-hypophysial tract, and end up stored in nerve endings in the posterior lobe of the pituitary gland, where they are finally exocytosed in response to stimuli (Zupanc 1996; Leng and Ludwig 2008).

The idea that neurons may be capable of dispatching blood-borne signals was initially met with great resistance (Scharrer B 1987a). In 1921, Loewi proposed the theory of the chemical transmission of the nervous impulse, which provided no tangible support for the concept of neurosecretion (Scharrer B 1987a; Loewi 1921). In the early studies, neurohormones appeared to be distinct from classic neurotransmitters in their proteinaceous character, the large amount of material produced and the extracellular pathway from site of release to site of action (Scharrer B 1987a).

Founding work by Ernst and Berta Scharrer in vertebrates and invertebrates, respectively, has showed the virtually universal occurrence of distinctive peptide-producing NE cell groups throughout the animal kingdom (Scharrer B 1987a). This work established that the vertebrate hypothalamic-hypophysial system was remarkably analogous to the insect brain-corpus cardiacum-corpus allatum system (Scharrer B 1987a; Scharrer B and Scharrer E 1944). The workhorse of both of these types of systems is the NE cell. NE cells, while they may express different neuropeptides, and control completely different physiologic processes, are nevertheless “wired” in the same way to accommodate a large secretory capacity.

Within the technical limitations of their time, extensive cytophysiologic studies by Ernst and Berta Scharrer provided the first evidence for a functional role of NE products in the control of developmental, reproductive, and metabolic functions in animals (Scharrer B 1987a). The age of electron microscopy allowed Palay, Bargmann, Bodian and the Scharrers to precisely characterize the sites of synthesis and release of peptidergic NE products (Scharrer B 1987a). This work showed that the NE material consisted of strikingly electron-dense, membrane-

bounded granules that were easily identified and localized (Scharrer B 1987a). Further demonstration that the Golgi apparatus was the site of dense core granule formation in packaged form, as well as findings of their axonal release by exocytosis provided evidence in support of a close similarity with the products of other protein-secreting gland cells (Scharrer B 1987a; Scharrer E and Brown 1961).

Protein secretion systems

Cells deliver proteins to the cell membrane or secrete them to the extracellular environment in two general ways: classic secretory systems and non-conventional systems that traffic proteins with special characteristics (Nickel 2010). The common entry point for proteins meant for extracellular release is the endoplasmic reticulum (ER) (Arvan and Castle 1998). Proteins destined for secretion are folded and processed in the ER and the Golgi apparatus, similar to all other proteins (Burgess and Kelly 1987; Mellman and Simons 1992; Zupanc 1996). After the initial processing in the cis-Golgi apparatus, secretory proteins enter the trans-Golgi network (TGN), where they are segregated from lysosomal enzymes and trafficked to the cell membrane in various types of organelles (Burgess and Kelly 1987; Zupanc 1996).

The TGN portion of the Golgi has come to be recognized as a major branch point from which distinct classic anterograde membrane traffic pathways emanate (Arvan and Castle 1998; De Matteis and Luini 2008; Tooze et al. 2001). These classic pathways include constitutive traffic via small vesicles to the plasma membrane, lysosomal biogenesis via the endosomal system and the regulated secretory pathway (RSP) via large dense core granules (LDCVs), used by specialized cell types (Arvan and Castle 1998; Morvan and Tooze 2008). Amongst classic anterograde secretory systems, the most commonly employed is the constitutive secretory pathway, believed to be utilized by all cells (Schmidt and Stephens 2010; Thiele and Huttner 1998; Tooze and Stinchcombe 1992). Constitutive secretion is the ubiquitous process by which proteins of general importance, such as those of the extracellular matrix, are

continuously released from the cell (Meldolesi et al. 2004). In addition to constitutive secretion, various specialized secretory systems have been described: constitutive-like secretion, regulated secretion, secretion of small permeable mediators by diffusion (nitric oxide, endocannabinoids) as well as non-conventional secretion (Arvan and Castle 1998; Nickel 2010; Südhof 2007; Vazquez-Martinez et al. 2011).

NE cells package peptide hormones, neuropeptides and specific neurotrophins into secretory vesicles for release in a regulated manner upon stimulation - also known as the regulated secretory pathway (RSP) (Park and Loh 2008). Constitutive-like secretion is a specialized secretory pathway first detected in exocrine tissues as a pathway of protein discharge attributable to neither constitutive nor regulated exocytosis (Arvan and Castle 1998). During maturation of the secretory vesicle, the RSP proteins are retained in the maturing vesicle by binding to a retention receptor (Park and Loh 2008). At the same time, constitutively secreted proteins are removed from the vesicle by a clathrin-mediated budding mechanism to yield constitutive-like vesicles for secretion (Park and Loh 2008). Non-conventional secretion examples include the export of cytoplasmic proteins such as fibroblast growth factor 2 mediated by direct translocation across plasma membranes, or export involving intracellular transport intermediates as shown for acyl-CoA binding protein (Nickel 2010). Gases such as nitric oxide and small lipophilic molecules are also known to diffuse directly out of the cell (Südhof 2007).

The neuroendocrine system

NE cells are capable of storing high concentrations of secretory proteins (Arvan and Castle 1992; Tooze and Stinchcombe 1992). These proteins are stored in a specialized type of vesicles with dense cores, the LDCVs (Tooze and Stinchcombe 1992; Scharrer E and Brown 1961). Their cargo is released only upon stimulation by a secretagogue, thereby giving rise to the name “regulated” secretory pathway (Burgess and Kelly 1987). Like other cells, NE cells engage in constitutive secretion, but they have also evolved the RSP. The primary function of

this pathway is the secretion of substances (mainly neuropeptides) that produce modulatory changes in target neurons and other cells (De Camilli and Jahn 1990; Hökfelt et al. 1986). In most cases these changes appear to involve the generation of intracellular second messengers (De Camilli and Jahn 1990; Winkler and Fischer-Colbrie 1998).

In order to execute a complicated secretory program that uses the RSP, NE cells have distinct cell biologic organelles and properties. These individual components, together, have allowed NE cells to assume important roles in the survival and evolution of metazoan organisms. NE cells control critical processes such as: growth of cells and organisms, homeostatic adaptation to environmental changes and stress, reproduction, circadian rhythms and sleep, metabolism and nutrient storage, defense against predators and other survival mechanisms (Kalra and Kalra 2010; Tsigos and Chrousos 2002).

Neurosecretory control over these processes is achieved through precise triggers, controlled modulation with on and off switches and with finely tuned timing. One excellent example of how precise and cued the neuroendocrine system is comes from insect molting. A remarkable cascade of endocrine factors regulates this seemingly simple task of shedding the cuticle called ecdysis (Mesce and Fahrbach 2002). Nevertheless, ecdysis involves a complex sequence of behaviors, called the ecdysis sequence (Truman 2005). A large body of work by Riddiford, Truman and others has identified the several hormones implicated in this process: Ecdysis Triggering Hormone (ETH), Eclosion Hormone (EH), Crustacean Cardioactive Peptide (CCAP), Bursicon and Corazonin (Truman 2005; Horodyski et al. 1989; Reynolds et al. 1979; Horodyski et al. 1993; Truman and Riddiford 1970; Taghert and Truman 1982).

Although the ecdysis mechanism is still subject to debate, it is widely accepted that EH and ETH regulate the timing of ecdysis and that they form an endocrine positive feedback system (Ewer 2005; Truman 2005). EH released into the circulation causes ETH release, which in turn feeds back on the CNS, to cause further release of EH (Ewer 2005). The positive

feedback between EH and ETH causes a massive surge of both peptides during which the stores of both are completely exhausted (Truman 2005). This provides an all‐or‐nothing signal that irreversibly commits the insect to an ecdysis attempt (Truman 2005). Occurrence of ecdysis behavior is strictly confined to the end of the molt (Ewer 2005). The EH/ETH loop is non-functional prior to this time, thereby restricting the occurrence of ecdysis behavior to the appropriate developmental time (Ewer 2005). An early attempt or a delayed attempt at ecdysis would have major consequences for the animal's metabolism and its ability to successfully shed cuticle. This example shows that without sophisticated control over neurosecretion, species would likely be impaired in their ability to withstand evolutionary pressures.

Similar to insects and other metazoans, human development and survival depend on NE cells, as judged by the multitude of pediatric and adult endocrine disorders. Diabetes mellitus, a disease that arises from destruction of the pancreatic beta cell is one of the most rapidly developing chronic diseases of the twenty-first century and is now one of the main threats to human health (Naser et al. 2006; Zimmet et al. 2001). The other increasingly common problem that frequently occurs together with diabetes is central and childhood obesity, both of which are now of epidemic proportions (Havas et al. 2009; Knecht et al. 2008; Levin 2009). Neuroendocrine signals originating from hypothalamus in the brain, the gastrointestinal system and adipose tissue play major roles in the pathogenesis of obesity, as they provide cues for cessation of meals and the coordination of caloric intake with maintenance of stable body weight (de Kloet and Woods 2010).

It is now well established that the hypothalamus plays a critical role in the regulation of energy balance through multiple output pathways (Abizaid and Horvath 2008). NE cells that respond to hormones leptin, insulin and ghrelin from the periphery and integrate various outputs into behavioral and neurohumoral cues are required to maintain body weight and adiposity

within relatively narrow limits in many individuals (Levin 2009). While diabetes and obesity are two of the gravest, there are a multitude of other neuroendocrine disorders with harmful effects on human health, many of which lack proper treatments. Therefore, to treat neuroendocrine diseases properly, it is important to understand the molecular basis of neuroendocrine system's function.

Invertebrate systems have played a major role in our understanding of how hormones adjust the functioning of the nervous system (Truman 2005). The episodic events, including the molting cycle and metamorphic transformations that lead to the emergence of adult insects, are programmed with greater precision than the developmental steps leading to maturity in most vertebrates (Scharrer B 1987b). The fruitfly *Drosophila melanogaster* is an excellent model organism for the study of NE cells. It is estimated that roughly 75% of fly genes have homologous sequences in the human genome (Bier 2005). Therefore, conclusions reached in flies are readily applicable to questions in human biology, as well as disease conditions. Flies are a workhorse of modern genetics with an experimental history exceeding one hundred years. Their rapid development, small size, fewer paralogous and redundant genes than humans, and remarkable functional conservation make *Drosophila* an excellent model system for the study of neurosecretion. With regards to NE cells, flies are a great model system for two reasons: i) the fly neuroendocrine system is highly analogous to the human neuroendocrine system ii) in addition to the principles on which the NE system is built, the building blocks themselves (genes encoding processing enzymes, LDCV constituents, RSP factors and neuropeptides) are highly conserved with human genes (Jiang et al. 2000; Han et al. 2004; Hwang et al. 2000; Kolhekar et al. 1997; Littleton 2000; Nässel 2002; Rayburn et al. 2003; Sidyelyeva et al. 2006; Siekhaus and Fuller 1999).

Cell biology of NE cells

NE cells possess a robust RSP, which acts through the subcellular compartment consisting of: i)LDCVs), ii) cellular processing machinery necessary for synthesis, processing and storage of large amounts of neuropeptides and iii) the ability to release cargos only in response to precise stimuli (Arvan and Castle 1998; Burgess and Kelly 1987). It is believed that NE cells, like all other cells, possess a constitutive secretory pathway alongside the RSP (Tooze and Stinchcombe 1992; Bauerfeind and Huttner 1993). These two pathways, while not mutually exclusive, are thought to utilize different mechanisms of protein trafficking.

The “two pathway hypothesis” was originally proposed by Kelly in 1985, based on a set of elegant experiments in a pituitary cell line (Gumbiner and Kelly, 1982). This work showed that AtT-20 pituitary-derived cells use two remarkably different cellular pathways for the delivery of two types of cargo: a viral protein versus the mature form of the adrenocorticotropic hormone (ACTH) (Gumbiner and Kelly, 1982). The viral protein, as well as the ACTH precursor were delivered to the cell surface and released soon after synthesis in a constitutive-like manner. On the other hand, the mature form of ACTH was stored in special vesicles and not readily released. Secretion of the matured ACTH occurred in response to a secretagogue, whereas the same secretagogue had little effect on the release of the viral protein or the immature form of ACTH (Gumbiner and Kelly, 1982). These experiments solidified the concept of distinct post-Golgi pathways for proteins targeted for constitutive versus regulated secretion (Morvan and Tooze 2008). It is now known that there are several other types of secretion systems (both conventional, such as the constitutive-like secretory pathway) and the non-conventional systems that accomplish protein secretion (Nickel 2010).

LDCVs are the main subcellular compartment used by cells to execute the RSP (Arvan and Castle 1998; Tooze and Stinchcombe 1992; De Camilli and Jahn 1990). NE cells can achieve a remarkable concentration and condensation of secretory products into LDCVs (Arvan and Castle 1992; Tooze and Stinchcombe 1992; Salpeter and Farquhar 1981). NE cells can

concentrate secretory proteins from the ER to the LDCVs by a factor as high as 200-fold (Salpeter and Farquhar, 1981). In contrast, a constitutively releasing cell, such as a plasma cell shows at most a two-fold concentration factor (Burgess and Kelly 1987; Hearn et al. 1985). This likely also reflects the temporal dynamics of these two different types of secretory systems. One (constitutive) is meant for constant trafficking, delivery and release of proteins into the extracellular environment or the cell membrane, thereby making cargo concentration unnecessary. The other (regulated) system works over longer time periods and its cargos are super-concentrated inside LDCVs, awaiting release at a precise moment in time (Arvan and Castle 1998).

It has been shown that NE cells store neuropeptide-filled LDCVs within their cytoplasm for long periods of time, thus acquiring large pools of intracellular mature secretory product (Burgess and Kelly 1987; Zupanc 1996; Burbach 2011). Cells that employ a robust RSP are therefore designed to synthesize and store one or a few secretory products and to discharge rapidly a large fraction of these products in response to physiologically specific stimulation, even in the absence of new protein synthesis (Burgess and Kelly 1987; Arvan and Castle 1998; Zupanc 1996).

LDCVs

LDCVs are membrane-enclosed spherical organelles with diameters of up to several hundred nanometers (Borgonovo et al. 2006). LDCVs have electron-opaque content that forms the “dense core” seen on electron microscopy (EM, Morvan and Tooze 2008; De Camilli and Jahn 1990; Peters et al. 1991). In NE cells, this core is separated from the membrane by a halo of “space” (Fuller et al. 1985; Geuze et al. 1983; Griffiths et al. 1981; Griffiths et al. 1984; Pothos et al. 2002). Little is known about the structure or assembly of this material itself (Burgess and Kelly 1987). The dense cores are stable, since empty unloaded cores have been observed even after exocytosis (Tooze and Tooze 1986). Additionally, when detergent is used to remove the

LDCV membrane, secretory products remain bound to the dense cores (Zanini et al. 1980).

Therefore, it can be generalized that cells that employ an RSP condense their secretory products, whereas those lacking an RSP do not (Kelly 1985). Constitutive secretion can also occur by means of various transport granules, which have distinct appearance and characteristics compared to LDCVs. Such constitutively secreted granules readily fuse with the cell membrane. Thus, in constitutively secreting cells, newly synthesized protein is not stored but leaves the Golgi apparatus in short-lived membrane vesicles that fuse immediately with the plasma membrane in the absence of any extracellular signal (Moore and Kelly 1985). In contrast, LDCVs are prevented from fusing with the plasma membrane until an intracellular messenger, such as calcium, undergoes a significant change in levels (Kelly 1985).

In addition to the rate of exocytosis, NE cells can also concentrate the sites of exocytosis. Some secretory cells, such as mast cells or neutrophils are not polarized and they traffic and release their products towards any part of the plasma membrane that is stimulated (Kelly 1985). NE cells, on the other hand, appear to be highly polarized. Hypothalamic neurons, for example, release neuropeptides exclusively from terminals located in the posterior pituitary gland (Kelly 1985; Leng and Ludwig 2008).

Secretory cargo sorting in the TGN

The TGN is the key exit point not only for RSP proteins and cargo, but for all other proteins that are destined for various locations, such as the plasma membrane, constitutive secretory vesicles, endosomes and immature secretory granules (Morvan and Tooze 2008). The mechanism of RSP sorting has been a matter of controversy. Two dominant hypotheses have been put forth: sorting-at-entry, and sorting-by-retention (Dikeakos and Reudelhuber 2007). There are several lines of evidence supporting both hypotheses, so it is possible that they are not mutually exclusive. In the sorting-at-entry mechanism, cargo proteins aggregate together in the acidic, calcium-enriched environment of the TGN (Park and Loh 2008). Certain

RSP cargoes such as the secretogranins are thought to be more prone to aggregation, and they in turn “seed” precipitation of other regulated cargoes (Borgonovo et al. 2006). Further support for this hypothesis comes from results showing complete spatial segregation of different pituitary hormones inside LDCVs found in single acidophilic bovine pituitary cells (Hashimoto et al. 1987). Thus, this study demonstrates the sorting of three different hormones to distinct LDCVs with separate storage within distinct granule types.

In the sorting-by-retention model, on the other hand, immature granules contain proteins meant for both constitutive and regulated secretion (Arvan and Castle 1998). As vesicles mature, proteins destined for constitutive secretion are extruded in low-density vesicles, ultimately leaving only the “correct” RSP cargo in the mature LDCV (Dikeakos and Reudelhuber 2007). Thus, in this model, non-secretory proteins are removed due to receptor-mediated sorting and/or their inefficient retention properties such as the inability to form aggregates (Vazquez-Martinez et al. 2011).

Enzymatic processing of neuropeptide precursors

Certain molecular characteristics are hallmarks of neuropeptides. Although neuropeptides are small protein molecules, they are still about 50 times larger than low-molecular-weight classical neurotransmitters (Salio et al. 2006). As a consequence, neuropeptides possess several more recognition sites for receptors than smaller neurotransmitter molecules and are capable of eliciting a biological effect when released in lower quantities (Salio et al. 2006). There are three commonly used criteria to denote neuropeptides: gene expression and biosynthesis by neurons, storage and regulated release upon demand, and the ability to modulate neural functioning directly through neural receptors (Burbach 2011).

Neuropeptides are generally synthesized as proteins with a pre and a pro region at the N-terminal (Docherty and Steiner 1982). The pre-region signal peptide is the 15-30 amino acid

long amino acid extension at the N-terminal of the precursor, the pre-propeptide (Zimmermann et al. 2011; Burbach 2011). The pre-region signal is required for import into the ER, and therefore is key for entry into secretion routes (Zimmermann et al. 2011; Burbach 2011). After ER entry, the pre-region is cleaved and the proneuropeptide can undergo appropriate biochemical processing. Since peptide precursors contain basic amino acids separating the various bioactive or spacer regions, it was hypothesized that a trypsin-like endopeptidase initially cleaved the precursors, thereby generating intermediates with C-terminal basic residues (Fricker 2005; Docherty and Steiner 1982). These basic residues would then be removed by a carboxypeptidase B-like enzyme, in many cases producing the final peptide (Fricker 2005).

Indeed, two such trypsin-like endopeptidases, prohormone convertase 1 and 2 (PC1 and PC2) are involved in the production of neuropeptides (Fricker 2005; Zhou et al. 1999; Seidah and Prat 2002). PC1 and PC2 were identified based on their similarity to the yeast enzyme Kex2p, the endoprotease responsible for cleaving pro-alpha-mating factor at paired basic amino acid sites (Eipper et al. 1993; Seidah and Chrétien 1999; Seidah et al. 1991; Smeekens and Steiner 1990). PCs cleave at the C-terminal side of dibasic residues, leaving peptides with a pair of basic amino acid residues (Lysine (K) and Arginine (R)) extending from their C-terminus (Perello and Nillni 2007; Fricker 2005).

The intermediates produced by PC1 and PC2 are subsequently cleaved by exopeptidases, such as carboxipeptidase E, which removes the C-terminal basic amino acids (Perello and Nillni 2007; Fricker and Snyder 1982). Cleavage of the neuropeptide precursor leads to the generation of specific peptides with different activities and functions, as well as non-functional byproducts (Burbach 2011). Cell-specific differences in PCs lead to different sets of peptides from the same precursor (Burbach 2011). While immature LDCVs are initially formed at the TGN, subsequent maturation steps occur beyond this compartment (Burgoyne and Morgan 2003; Tooze and Huttner 1990; Arvan and Castle 1987).

Peptide amidation

In addition to cleavage, many neuropeptides undergo additional posttranslational processing steps such as C-terminal amidation, N-terminal acetylation, or other modifications such as glycosylation, sulfation, and phosphorylation (Fricker 2005; Eipper et al. 1992; Burbach 2011). These extra steps may provide longer stability and stronger activity. C-terminal amidation is especially common (Eipper et al. 1993). At least 50% of all examined neuropeptides regardless of species have an alpha-amide moiety at their carboxyl end (Eipper et al. 1992). More than 90% of *Drosophila* neuropeptides are predicted to be amidated (Park and Taghert 2009). For many neuropeptides, the presence of the alpha-amide moiety is essential for biologic activity (Eipper et al. 1992; Tazi et al. 1987). Other speculations about the role of amidation have included a role in protecting peptides from enzymatic degradation (half-life) and increasing binding affinity (Fuhlendorff et al. 1990; Kreil 1985).

Peptide amidation is carried out by a bifunctional enzyme that can only act after PCs and carboxypeptidases have acted (Eipper et al. 1993). The precursors to alpha-amidated peptides always contain a Gly residue to the COOH-terminal side of the residue to be alpha-amidated (Eipper et al. 1992). The first step of the two-part amidation reaction is carried out by a peptidylglycine alpha-hydroxylating monooxygenase (PHM), which catalyzes the copper, ascorbate, and molecular oxygen dependent conversion of peptidylglycine substrates into an alpha-hydroxyglycine intermediate (Eipper et al. 1991; Eipper et al. 1992; Eipper et al. 1993). A peptidyl-alpha-hydroxyglycine alpha-amidating lyase (PAL) activity catalyzes the conversion of these intermediates into alpha-amidated products along with the production of glyoxylate (Eipper et al. 1993; Katopodis et al. 1991; Kato et al. 1990). PAL and PHM are contained within a single bifunctional enzyme in mammals, whereas in *Drosophila*, these enzymes are encoded by one or more separate genes (Kolhekar et al. 1997; Jiang et al. 2000; Han et al. 2004).

Newly synthesized neuropeptides move through the RSP while the various endoproteolytic cleavages are occur (Eipper et al. 1993). The subcellular location in which each enzyme functions is not fully clear, but it is known that although the TGN marks the beginning of processing, the majority of the cleavages occur in LDCVs (Eipper et al. 1993; Schnabel et al. 1989; Glembotski 1981). Finally, neuropeptides can undergo various other chemical modifications including N- and O-glycosylation, phosphorylation, sulfation, and acetylation, all of which have various biological effects (Burbach 2011).

While Work by Kelly and subsequent work established the basis for the RSP, the mechanism by which the RSP operates from cargo's entry into the pathway to its delivery to the extracellular environment is still largely unknown. Furthermore, we currently lack extensive knowledge of the mechanism that establishes the RSP inside developing NE cells, and then maintains it in mature NE cells. What is clearly known is that the RSP appears to be conserved across species and genera. NE cells are commonly studied in worms, flies and other insects, frogs, fish and various mammalian species.

Fast neurotransmission versus slow neuropeptide-mediated neurotransmission

Although neuroendocrine signaling is very important for the nervous system, quantitatively, synaptic vesicle-mediated synaptic transmission is the dominant form of communication between neurons (Südhof 2007). Nevertheless, this does not mean that synaptic transmission is more important than the other types of neurotransmission: the two principally different signaling pathways play distinct roles in information processing by the brain, and both are essential for brain function (Südhof 2007).

Synaptic vesicles (SVs) have a characteristic cross-sectional diameter of 30-40 nanometers, appear clear by EM and contain classical neurotransmitters, such as acetylcholine, gamma-amino-butyric acid (GABA) or glutamate (Craige et al. 2004; Edwards 1998; Südhof 2007). SVs and LDCVs can be readily distinguished by EM: unlike LDCVs, SVs have a clearly

electron-lucent appearance, which is due to the absence of protein or peptides in their lumen (Craige et al. 2004). LDCVs and SVs are further distinguished by the mode of stimulation required for release, the type of calcium channels involved in the exocytotic process, and the time course of recovery after stimulation (Zupanc 1996).

In addition to morphology and composition, the two vesicle types are also quite different functionally: SVs cluster over synaptic active zones, respond to high micromolar calcium concentrations and release their contents rapidly, within a millisecond (Edwards 1998; Katz 1969; Südhof 2007). In addition to the rapid pace of release, neurons release SVs in a highly localized manner, restricted to an area of less than a square micrometer (Südhof 2004; Südhof 2007). SV release occurs mainly at the presynaptic terminal into the synaptic cleft (Ludwig and Leng 2006).

As with all phenomena in biology, there are exceptions to all rules. In most neurons for any single event, a single vesicle per synapse fuses with the membrane following action potential invasion, and this exocytotic event is limited to the ultrastructurally defined presynaptic active zone (Biro et al. 2005; Matsui and Jahr 2006; Sargent et al. 2005; Silver et al. 2003). Nevertheless, at certain synapses, more than a single vesicle can be released per action potential, and in rare circumstances neuronal exocytosis can occur from sites that are unremarkable in electron micrographs (Ceccarelli et al. 1979; Matsui and Jahr 2006; Zenisek et al. 2000). It is important to note that the above-mentioned example is an exception and not a rule.

When an action potential depolarizes a terminal, calcium enters into the terminal through voltage-gated channels at the “active zone” and causes SV exocytosis with resultant release of classic neurotransmitters into the synapse (Ludwig and Leng 2006; Owald and Sigrist 2009; Südhof 2004). Calcium concentration upon entry in the vicinity of the channels can be very high, but it does not penetrate far due to rapid sequestration by high-capacity buffers (Ludwig and

Leng 2006). SV exocytosis involves sensors with a low affinity for calcium, likely due to high calcium concentrations found close to the site of calcium entry (Ludwig and Leng 2006). Since LDCVs are frequently located deeper inside the cell, at sites distant from active zones and are not exposed to such calcium high concentrations, LDCV exocytotic machinery has a high calcium affinity and so might be activated by calcium entry when a high rate of entry is induced by intense activation (Ludwig and Leng 2006; Zupanc 1996).

In contrast to SVs, LDCVs have a diameter of 80–200 nanometers and exhibit a characteristic electron-dense core that contains neuropeptides (Edwards 1998; Crivellato et al. 2005; Crivellato et al. 2006). SVs have dedicated release sites that are associated mainly with synapses, but LDCVs are released from all parts of a neuron, including the axon, soma and, especially, the dendrites (Ludwig and Leng 2006; Morris and Pow 1991; Salio et al. 2006; Südhof 2007). When LDCVs pool in in neuronal cell bodies, dendrites as well as the axonal terminals, they await a proper signal that would instruct their release (Arvan and Castle 1998).

Once triggered, LDCVs release their contents relatively slowly, taking more than 50 milliseconds and they do so in response to low micromolar calcium concentrations (Edwards 1998; Martin 2003). Compared to SVs, LDCV exocytosis requires more sustained calcium transients, but is slower and longer lasting than SV exocytosis (Südhof 2007; Stjarne 2000; Martin 2003). For example, calcium-initiated chromaffin granule exocytosis is 10-fold slower than SV exocytosis, but otherwise the two systems are similar (Südhof 2007). Though fast, SV-mediated and slow, LDCV-mediated neurotransmission are built on many of the same principles, NE cells are known to have evolved special adaptations for LDCV-mediated slow neurotransmission (Fukuda et al. 2004).

Similar to exceptions to the rule seen for SVs, there are such exceptions for LDCVs as well. Occasionally, LDCVs have been observed undergoing exocytosis at typical active zones of nerve terminals in stimulated synapses of frog sympathetic ganglia and of rat trigeminal

subnucleus caudalis (Pécot-Dechavassine and Brouard 1997). In the majority of examined cases, classic neurotransmitters and peptides are not stored together: the former are stored in SVs, which are densely clustered in synaptic terminals, whereas peptides reside in LDCVs, which are more rare in terminals (Ludwig and Leng 2006).

After mature LDCVs are formed by budding from the TGN, they are transported from the cell body to axons or dendrites or they can remain in the cell body (Südhof 2007). A calcium-based signal triggers the translocation and fusion of LDCVs with the plasma membrane outside of the active zone (Rutter and Tsuboi 2004; Südhof 2007). There still exists a certain amount of uncertainty and controversy regarding the precise mechanism of LDCV fusion with the membrane. After exocytosis, empty LDCVs recycle and refill by transport to the cell body and recycling via the Golgi complex (Südhof 2007).

The mechanism of coordination and maintenance of NE cell properties

Many cell-intrinsic regulatory mechanisms that help establish specific transmitter phenotypes have been discovered (Park and Taghert 2009; Ding et al. 2003; Cheng et al. 2004). Nevertheless, what is not known is how NE cells develop their mature properties (Park and Taghert 2009). A host of transcription factors such as Mash1, Otp, Brn2, Sim1 and Sim2 have been shown to regulate the early differentiation of hypothalamic neuroendocrine centers by their expression in neuronal progenitors and in pre-migratory neurons (Acampora et al. 1999; Goshu et al. 2004; Hosoya et al. 2001; Michaud et al. 1998; Nakai et al. 1995; Park and Taghert 2009; Schonemann et al. 1995; Wang and Lufkin 2000).

The above mentioned factors are important early in specification of their corresponding NE cell lineages, as well as in proliferation and migration of NE precursors (Park and Taghert 2009). Very little is known about factors acting in post-mitotic cells, once they are restricted to their appropriate lineages. This is a time point in the life of a newly born neuron when the neuron needs to start developing properties of the mature cell type it is destined to become. In

the case of NE cells, a central question is: what are the intrinsic regulatory factors that directly organize the maturation of peptidergic cellular properties (Park and Taghert 2009)? Such properties involve proper scaling and modulation of the RSP by which peptides and peptides hormones are packaged and released, as well as active maintenance of the RSP and proper homeostatic response to various environmental and internal perturbations (Park and Taghert 2009; Mills and Taghert 2011).

Although several groups have reported on identification of individual factors that operate in the RSP, the genetic basis of RSP establishment, scaling and maintenance is largely unknown (Arvan and Castle 1998; Beuret et al. 2004; Kim et al. 2001; Park and Taghert 2009). Clonal isolates of particular cell lines (AtT-20 and PC12) have been identified as lacking particular molecular or subcellular features of the RSP (Day et al. 1995; Malosio et al. 1999). Various groups have reported several independently derived neuroendocrine cell variants that lacked regulated secretory organelles or failed to express genes known to be functionally important for secretory cells (Pance et al. 2006; Matsuuchi and Kelly 1991; Borgonovo et al. 1998; Leoni et al. 1999; Pance et al. 1999).

The transcription factor RE1-silencing transcription factor (REST), is the only known transcription factor known to play a role in the mammalian RSP (Pance et al. 2006). For example, the A35C PC12 clone lacks markers for both SVs and LDCVs at the protein and mRNA level (Pance et al. 1999). Hybrid cells created by fusing A35C with normal PC12 also failed to accumulate LDCVs, suggesting the existence of a transcriptional repressor that blocks the existence of a functional RSP (Pance et al. 1999; Pance et al. 2006). This further indicates that neurosecretion could perhaps be controlled in a modular way (Borgonovo et al. 1998; Pance et al. 2006). REST is necessary but not sufficient to suppress the NE phenotype in A35C cells, which implies that other transcription factors or signaling proteins may need to act together with REST to repress the RSP (Pance et al. 2006). It is therefore possible that RSP

regulation in other systems also requires factors acting transcriptionally, as well as post-transcriptionally.

Work in *Drosophila* has suggested that there is a genetic basis underlying development and maintenance of the NE cell fate (Park and Taghert 2009). A single transcription factor named *dimmed* (DIMM) has been shown to act as a dedicated pro-secretory factor (Hewes et al. 2003; Park and Taghert 2009). The orthologue of DIMM, named Mist1 in mammals, has been shown to be a dedicated pro-secretory factor in the serous exocrine cells (Johnson et al. 2004; Moore et al. 2000; Pin et al. 2000; Ramsey et al. 2007). As in hypothalamic and pituitary development in mammals, there is a whole cascade of transcription factors implicated in the specification of *Drosophila* neuroblasts and their transition to post-mitotic NE cells destined to give rise to mature peptidergic cells (Allan et al. 2005; Benveniste et al. 1998; Baumgardt et al. 2009; Kohwi and Doe 2009; Park et al. 2004). These factors include the transcription factors *apterous*, *squeeze*, *eyes absent*, *grainyhead*, *collier* and other early factors. By employing transcription factor cascades and loops, *Drosophila* neural stem cells generate many distinct neuronal and glial subtypes over time (Kohwi and Doe 2009). Nevertheless, there are only two known *Drosophila* transcription factors involved in development of the mature phenotype of a CNS cell type: the neuroendocrine-specific factor *dimmed* and the pan-glial factor *repo* (Hewes et al. 2003; Freeman et al. 2003).

Past work has demonstrated the existence of transcription factors acting to set up specialized intracellular circuits and pathways that distinguish one cell type over another (Mills and Taghert 2011). Two known cases of transcriptional organizers of the developmental programs of neuronal cell subtypes are factors PET-1 and AST-1 (Park and Taghert 2009; Hendricks et al. 1999; Flames and Hobert 2009). AST-1 is an ETS transcription factor that controls expression of all dopamine pathway genes in all dopaminergic cell types in *Caenorhabditis elegans* (Flames and Hobert 2009). PET-1 is another ETS domain transcription

factor that appears to control expression of serotonin pathway genes in mammalian serotonergic neurons (Hendricks et al. 1999). Both of these factors control major elements of the biosynthetic machinery known to distinguish their respective neuronal subtypes from neurons expressing other biogenic amines or classic neurotransmitters. Neither factor, however, is necessary for the initial fate specification or survival of the respective neuron type that they control (Park and Taghert 2009). There is also no evidence as of yet that either of these factors controls a whole subcellular compartment, such as LDCVs, endosomes or synaptic vesicles.

One known factor that controls a whole subcellular compartment is Transcription Factor EB (TFEB) (Sardiello et al. 2009). TFEB exerts coordinated transcriptional control over most mammalian lysosomal genes (Sardiello et al. 2009). *In vitro*, TFEB induces lysosomal biogenesis and increases degradation of lysosomal complex molecules (Sardiello et al. 2009). Another factor, Peroxisome proliferator–activated receptor Gamma Coactivator-1 (PGC-1), has been shown to control the number of mitochondria inside heart cells, as well as their function in response to energy demands (Finck and Kelly 2007; Lehman et al. 2000). Though it has been convincingly shown that PGC-1 is a master regulator of myocardial energy metabolism, this protein is a not a sequence-specific transcription factor, but rather a co-activator for such factors (Finck and Kelly 2007). The above examples demonstrate master regulators / scaling factors that control the biogenesis, maintenance and homeostasis of lysosomes and mitochondria (Lehman et al. 2000; Sardiello et al. 2009). Based on previous and current work, evidence points to DIMM acting as a scaling factor on the RSP of NE cells, through the morphological correlate of the RSP, the LDCV compartment (Mills and Taghert 2011; Park et al. 2011).

Overview of DIMM biology

DIMM was identified by Hewes et al. in 2003 as a neuroendocrine-specific basic helix loop helix (bHLH) transcription factor in *Drosophila*. The original *dimmed* loss-of-function phenotype was a severe reduction of neuropeptide staining intensity, with normal survival of NE

neurons (Hewes et al. 2003). These early results pointed to a role of DIMM in regulating the differentiation of neuroendocrine lineages (Hewes et al. 2003; Park et al. 2004). Furthermore, DIMM was proposed to be an integral component of a novel mechanism by which diverse neuroendocrine lineages differentiated and maintained their pro-secretory state (Hewes et al. 2003). DIMM is a member of the Atonal subfamily of transcription factors and its bHLH domain displays 78% identity with the mouse Mist1 mammalian protein (Moore et al. 2000; Pin et al. 2000). The predicted DIMM protein consists of a single 390 amino acid open-reading frame with a centrally-positioned bHLH domain (Park and Taghert 2009).

DIMM was originally identified thanks to *Drosophila* P-element based enhancer trap screening that yielded a GAL4-containing P-element named c929-GAL4 inserted 13 kilobases upstream of the DIMM transcription start site (Hewes et al. 2003). c929-GAL4 drives gene expression in a pattern that overlaps to a large extent with DIMM, based on *in situ* hybridization and immunohistochemical analysis (Hewes et al. 2003; Park et al. 2008a). c929-GAL4 and DIMM expression overlap in many peptidergic cells in the CNS, and in peripheral endocrine cells of the corpora cardiaca as well as in the Inka cells of the epitracheal endocrine system (Park and Taghert 2009; Park et al. 2008a). DIMM is essential for survival as *dimm* null animals die late in embryonic stages or rarely as first instar larvae (Park and Taghert 2009; Hewes et al. 2003). This phenotype is strikingly similar to that of animals null for the *Drosophila* PHM gene (Jiang et al. 2000).

The developmental time course of DIMM expression has been explored in detail (Hewes et al. 2003; Allan et al. 2005). Except for maternal DIMM expression and a transient embryonic stage 11 expression, the first time point at which DIMM is stably expressed in the developing embryonic nervous system is in stage 12 embryos (Allan et al. 2005; Hewes et al. 2003; Park and Taghert 2009). Interestingly, there are cases of post-mitotic neurons born in the embryo that do not display characteristics of mature NE cells until metamorphosis (Park et al. 2004). Such

cells lack DIMM expression preceding metamorphosis, but they acquire DIMM expression immediately before differentiation (Park et al. 2004; Park and Taghert 2009). Importantly, once DIMM expression commences in the embryonic CNS, the vast majority of DIMM-expressing neurons maintain persistently high-level DIMM expression throughout all developmental stages and from beginning to the end of their adult life (Park et al. 2008a). In late third instar larvae, there are ~306 DIMM-positive CNS cells, most of which are NE cells (Park et al. 2008a; Park and Taghert 2009).

The expression of DIMM was surveyed against that of 26 different neuropeptide markers (Park et al. 2008a). This study found that 64% of DIMM cells within the larval CNS could be identified by one or more of the 24 CNS peptide gene markers. Furthermore, there was remarkable overlap between NE neurons expressing high levels of PHM and DIMM (Park and Taghert 2009). Outside the CNS, 73% of the 53 prominent DIMM-positive cells in the larva are associated with one or more neuropeptide markers (Park et al. 2008a). Therefore, most (if not all) DIMM-expressing neurons are peptidergic (Park et al. 2008a). DIMM expression is restricted to NE neurons of the neuroendocrine, as well as interneuronal subtypes (Park and Taghert 2009). On the other hand, motor and sensory neurons do not express appreciable levels of DIMM (Park and Taghert 2009).

Although most DIMM+ cells are peptidergic, the corollary is not necessarily true and varies on a peptide-to-peptide basis (Park and Taghert 2009; Park et al. 2008a). Of the 26 neuropeptide markers, 8 are >90% DIMM-positive, 18 are >40% DIMM-positive and 2 peptide markers do not overlap with DIMM (Park et al. 2008a). There are ~1,000 NE cells in the larval brain that express any of the 26 neuropeptide markers, and ~30 - 40% of them are DIMM-positive (Park et al. 2008a). This selective overlap of DIMM with particular subpopulations of NE cells, even within populations expressing the same neuropeptide, likely provides a clue to DIMM

function. There are several features that unite the subtypes of cells that express DIMM that will be further described below.

Work in this thesis aims to show the mechanism by which DIMM operates to ensure proper development and dynamic maintenance of the NE cell fate. In addition to the homologous roles of DIMM and Mist1 in LDCV biogenesis and maintenance, conclusions reached from studying DIMM could be broadly applicable to other systems due to the high degree of conservation of RSP factors. The majority of processing enzymes, LDCV trafficking, biogenesis and release-implicated genes, as well many neuropeptides are conserved in *Drosophila* and mammals (Han et al. 2004; Hwang et al. 2000; Jiang et al. 2000; Kolhekar et al. 1997; Littleton 2000; Nässel 2002; Rayburn et al. 2003; Sidyelyeva et al. 2006; Siekhaus and Fuller 1999).

In order to decipher the mechanism of DIMM operation, various genomic techniques can be employed to examine the role of DIMM on a genome-wide level. A review of literature shows that there are only a few reports of *in vivo* transcriptional profiling of particular NE cell populations. One such example comes from transcriptional profiling of normal and osmotically stressed rats (Hindmarch et al. 2006; Hindmarch and Murphy 2010). Here, the authors tried to understand the hypothalamo-neurohypophyseal system's response to dehydration. The hypothalamo-neurohypophyseal system is a NE center responsible for the production of the antidiuretic peptide hormone vasopressin (Hindmarch et al. 2006). Dehydration evokes a massive release of the vasopressin from magnocellular neuron axon terminals in the posterior pituitary, which is accompanied by a plethora of changes in the morphology and electrophysiological properties, as well as biosynthetic and secretory activity of the hypothalamo-neurohypophyseal system (Hindmarch et al. 2006). The authors employed microarray technology to globally identify transcripts that were regulated as a consequence of dehydration as well as RNAs that were enriched in specific hypothalamic nuclei under normal

conditions (Hindmarch et al. 2006; Hindmarch and Murphy 2010). Similarly, the role of DIMM in NE cells can probably be best understood with a genome-wide approach aimed at understanding its genomic targets and the NE cell transcriptome.

DIMM and the concept of LEAP cells

Among peptidergic cells, several attributes set DIMM cells apart (Park and Taghert 2009). These properties are: i) the relatively large size of DIMM-expressing NE cells compared to other NE cells (**L**), ii) their pattern of episodic release of peptides (**E**), iii) their specific production of amidated peptides (**AP**) (Park and Taghert 2009). Therefore, the acronym LEAP was coined (Park and Taghert 2009).

Among the peptidergic cells of the fly CNS, LEAP cells have relatively large cell bodies, as well as larger and more complex branching patterns (Park and Taghert 2009). Episodic release of peptides is a property of LEAP cells that is currently largely speculative or indirect, as release has not been documented for many LEAP cells (Park and Taghert 2009). During episodic release, discrete release events have ballistic rising phases and decay phases lasting many minutes, with the majority of stored peptides released (Reynolds et al. 1979). Such release has been documented for several subpopulations of LEAP cells that have important roles in ecdysis (Reynolds et al. 1979). Finally, LEAP cells specialize in the production of amidated peptides. DIMM expression is highly correlated with C-terminal alpha-amidation of secretory peptides (Park and Taghert 2009). Several lines of evidence support this notion: DIMM directly controls the expression of the amidating enzyme PHM and their expression patterns overlap highly (Hewes et al. 2003; Park et al. 2008b).

Mechanism of DIMM action

DIMM loss-of-function mutants show reduced neuropeptide staining, but also reduced levels of neuropeptide processing enzymes (Hewes et al. 2003; Park et al. 2008b). On the other hand, DIMM loss appears to have no effect on the survival or arborization or morphology of NE

neurons that normally express DIMM (Hewes et al. 2003). Selective DIMM misexpression in the CNS can increase the steady state peptide accumulation in NE cells (Hewes et al. 2003; Allan et al. 2005; Miguel-Aliaga et al. 2008). Furthermore, misexpression of DIMM confers high-level PHM expression onto all neurons (Allan et al. 2005). This suggests that DIMM acts as a "master regulator" of the peptidergic cell fate. On the other hand, work in the *Drosophila* embryo and larvae suggests that DIMM also participates in combinatorial codes with various other transcriptional regulators to promote unique subtypes of NE cells (Allan et al. 2005; Baumgardt et al. 2009; Miguel-Aliaga et al. 2008). These two modes of operation are not mutually exclusive and can be reconciled with each other, even in the same cell.

DIMM is a class 2 bHLH protein in the Atonal superfamily (Moore et al. 2000). Class 2 bHLH transcription factors function mostly as heterodimers with a class 1 ubiquitous bHLH protein such as E12/E47 in mammals and *daughterless* in flies (Massari and Murre 2000; Park and Taghert 2009). DIMM appears to favor functioning as a homodimer as opposed to forming heterodimers with DAUGHTERLESS, as other members of its superfamily (Allan et al. 2005). The simplest prediction for how DIMM functions is that as a transcription factor, it binds to enhancers in the genome and activates expression of target genes (Powell and Jarman 2008). Prior to work presented in Chapters 2 and 3 of this thesis, there was a single known direct DIMM target - the *Drosophila* amidation enzyme PHM (Park et al. 2008b).

The function of DIMM and PHM is tightly linked, both *in vitro* and *in vivo* (Park and Taghert 2009). Careful analysis has identified precise enhancers in the first intron of PHM that DIMM binds to and transactivates PHM expression (Park et al. 2008b). *In vitro*, DIMM controls PHM transcription by binding to each of three adjacent E boxes within the first PHM intron (Park et al. 2008b; Park and Taghert 2009). The three identified E-boxes also contribute equally to proper expression of a PHM transgenic reporter (Park et al. 2008b). Therefore, prior to this

thesis work, PHM was the single *bona fide* direct DIMM target, *in vitro* and *in vivo* (Park et al. 2008b; Park and Taghert 2009).

More recent studies have shown additional evidence in support of DIMM as a master regulator of peptidergic cell fate (Hamanaka et al. 2010; Park et al. 2011). When DIMM is misexpressed in non-peptidergic neurons such as photoreceptors, the morphology and physiology of these neurons changes dramatically (Hamanaka et al. 2010). Normally, non-peptidergic neurons fail to accumulate ectopically expressed neuropeptides (Hamanaka et al. 2010; Helfrich-Förster et al. 2000; Rao et al. 2001). Nevertheless, when non-peptidergic neurons are forced to express a neuropeptide and DIMM, they can successfully express high levels of neuropeptides (Hamanaka et al. 2010). Furthermore, when DIMM is co-misexpressed with a neuropeptide, photoreceptors can successfully make bioactive peptides, as measured by mass spectrometry (Hamanaka et al. 2010). Ultrastructural analysis of photoreceptors misexpressing DIMM and/or neuropeptide showed a dramatic appearance of an LDCVs inside photoreceptors (Hamanaka et al. 2010). When DIMM and a neuropeptide are co-misexpressed, the resultant novel LDCVs package PDF inside ectopic LDCVs, as shown by immunogold EM (Hamanaka et al. 2010). Normally, photoreceptors package histamine inside SVs and do not produce any LDCVs (Borycz et al. 2005; Hamanaka et al. 2010).

In addition to DIMM causing LDCV biogenesis when expressed ectopically, it also appears to repress features required for classic histamine neurotramission in photoreceptors, including T bar ribbons, sites for vesicle exocytosis, and capitate projections, sites for vesicle endocytosis, as well the normal architecture of the terminal itself (Hamanaka et al. 2010). Although classic neurotransmitters and neuropeptides coexist in most neurons, these results suggest the possibility that DIMM effects an extreme peptidergic phenotype with a near exclusive dedication of the cell to the accumulation and release of neuropeptides (Hamanaka et al. 2010).

Exploring DIMM's mechanism of action

A precise interplay between cis-acting elements and trans-acting factors allows signaling pathways to activate or repress the expression of specific genes and to maintain these expression patterns in differentiated tissues (Carrera and Treisman 2008). In the eukaryotic genome, RNA polymerase II transcribes most genes, which requires assembly of a pre-initiation complex, consisting of general transcription factors (Carrera and Treisman 2008; Fuda et al. 2009). In addition to general transcription factors, there are many sequence specific transcription factors that function in cell-specific and tissue-specific ways (Macquarrie et al. 2011). Such transcription factors enable Pol II to gain access to promoters and to initiate RNA synthesis at transcription start sites (Fuda et al. 2009). This is accomplished by transcription factor binding in a sequence-specific manner to cis-regulatory modules (enhancers, promoters) found throughout the non-coding portions of the genome (Biggin and Tjian 2001; Levine and Tjian 2003; Li et al. 2008).

There are several techniques for studying how transcription factors interact with the genome (Southall and Brand 2007; Nègre et al. 2006; Sandmann et al. 2007). Chromatin immunoprecipitation (ChIP) has become the technique of choice to investigate protein–DNA interactions inside the cell (Collas 2009; Lee et al. 2006; Sandmann et al. 2007). During ChIP, a protein of interest is selectively immunoprecipitated from a chromatin preparation to determine the DNA sequences associated with it (Collas 2009; Gilmour and Lis 1984). ChIP was developed independently by Gilmour and Lis to study the distribution of RNA Polymerase and histones in bacteria and flies, and by Solomon and Varshavsky to study the binding of *Drosophila* heat shock protein 70 to DNA (Gilmour and Lis 1984; Gilmour and Lis 1987; Solomon et al. 1988). Since then, ChIP has been applied to sequence specific transcription factors and other chromosome-associated proteins (Collas 2009).

For ChIP, DNA and proteins are reversibly cross-linked with formaldehyde (which is heat-reversible) to covalently attach proteins to target DNA sequences (Collas 2009). A specific antibody is then used to immunoprecipitate the protein of interest (Nègre et al. 2006). Formaldehyde crosslinks proteins and DNA molecules within 2 Angstroms of each other, and thus is suitable for crosslinking proteins which directly bind DNA (Collas 2009; Toth and Biggin 2000). Traditionally, DNA sequences isolated by ChIP were quantified or identified by Southern blotting or quantitative polymerase chain reaction (qPCR) (Mukhopadhyay et al. 2008; Toth and Biggin 2000; Nègre et al. 2006).

In the past decade, ChIP has been coupled to microarray hybridization (ChIP-chip) and to high-throughput sequencing (ChIP-Seq, Park 2009). This has enabled determination of the precise, genome-wide distribution of transcription factor binding sites (Macquarrie et al. 2011). These advances have allowed an unbiased detection of protein–DNA interactions without having to “select” certain candidate binding and control regions for interrogation (Sandmann et al. 2007).

Although individual genomic data sets are highly informative, integrating them together with global profiles of transcription or protein interactions offers the exciting potential to answer many long-standing questions such as how enhancers contribute to gene expression variation (Hawkins et al. 2010). As with ChIP-chip, microarrays were also the first technology allowing genome-wide quantification of gene expression (Watson et al. 1998; Epstein and Butow 2000). High throughput sequencing has improved on microarrays with higher precision, dynamic range and genome coverage, and lower background (Wang et al. 2009; Ozsolak and Milos 2011).

Transcriptomes produced by RNA-Seq or microarrays can be integrated with genome-wide binding information from ChIP-chip or ChIP-Seq studies. Identifying the set of promoters or enhancers where factor binding correlates with gene expression increases the probability that the factor binding site is associated with adjacent gene expression (Ren et al. 2000; Simon et al.

2001; Wyrick and Young 2002). Additionally, this integration helps resolve potential ambiguities that occur in cases of intergenic ChIP binding in areas containing the promoters for two divergently transcribed genes (Wyrick and Young 2002). Furthermore, the sets of promoters / genes found within the intersection of ChIP-chip binding and expression datasets contains fewer false positives due to a reduction in noise (Wyrick and Young 2002).

Work on this thesis started out with trying to understand the precise mechanism of DIMM function. The most logical and technically feasible approach was to perform *in vivo* DIMM ChIP in adult fly heads. Experimental success was confirmed by testing for DIMM occupancy at its known enhancers in the first intron of PHM. Additionally, a set of genes whose expression was upregulated upon DIMM misexpression in embryonic neurons and downregulated in DIMM loss-of-function mutants was also tested for direct DIMM binding to putative enhancers. This gene list was integrated with a list of genes known to be enriched in normal adult DIMM-expressing neurons (Kula-Eversole et al. 2010). A portion of this work is presented in Chapter 2.

Upon demonstrating DIMM ChIP success, analysis was then extended to a genome-wide level by *in vivo* ChIP-chip. This revealed a set of 156 binding peaks that revealed a rich repertoire of 284 genes. ChIP-chip analysis provided new clues for how DIMM establishes and maintains the secretory capacity of a NE cell. A way to enhance the ChIP-chip data further was to obtain a gene expression profile of LEAP cells, which was not unknown. This was done by purifying normal DIMM+ adult neurons and profiling their transcriptome by deep sequencing. The two data sets were then intersected: genes enriched in DIMM+ cells compared to DIMM- cells were integrated with genes adjacent to or overlapping with DIMM binding peaks.

The resulting data set of 116 genes likely increased the quality of the data set by decreasing ambiguities behind intergenic binding and its assignment to peaks (Wyrick and Young 2002). Finally, many of the genes from the intersected list were tested in a functional

behavioral essay for DIMM-dependent sleep phenotypes, presented in Chapter 4. Due to the multidisciplinary and diverse nature of analysis, the integrated DIMM transcription network likely represents one of the first such networks revealed for a “master regulator” gene. This work is presented in Chapter 3.

REFERENCES

Abizaid, A., and Horvath, T.L. (2008). Brain circuits regulating energy homeostasis. *Regul. Pept.* **149**, 3-10.

Acampora, D., Postiglione, M.P., Avantaggiato, V., Di Bonito, M., Vaccarino, F.M., Michaud, J., and Simeone, A. (1999). Progressive impairment of developing neuroendocrine cell lineages in the hypothalamus of mice lacking the Orthopedia gene. *Genes Dev.* **13**, 2787-2800.

Allan, D.W., Park, D., St Pierre, S.E., Taghert, P.H., and Thor, S. (2005). Regulators acting in combinatorial codes also act independently in single differentiating neurons. *Neuron* **45**, 689-700.

Arvan, P., and Castle, J.D. (1987). Phasic release of newly synthesized secretory proteins in the unstimulated rat exocrine pancreas. *J. Cell. Biol.* **104**, 243-252.

Arvan, P., and Castle, D. (1992). Protein sorting and secretion granule formation in regulated secretory cells. *Trends Cell. Biol.* **2**, 327-331.

Arvan, P., and Castle, D. (1998). Sorting and storage during secretory granule biogenesis: looking backward and looking forward. *Biochem. J.* **332** (Pt 3), 593-610.
Bauerfeind, R., and Huttner, W.B. (1993). Biogenesis of constitutive secretory vesicles, secretory granules and synaptic vesicles. *Curr. Opin. Cell Biol.* **5**, 628-635.

Baumgardt, M., Karlsson, D., Terriente, J., Díaz-Benjumea, F.J., and Thor, S. (2009). Neuronal subtype specification within a lineage by opposing temporal feed-forward loops. *Cell* **139**, 969-982.

Benveniste, R.J., Thor, S., Thomas, J.B., and Taghert, P.H. (1998). Cell type-specific regulation of the *Drosophila* FMRF-NH2 neuropeptide gene by Apterous, a LIM homeodomain transcription factor. *Development* **125**, 4757-4765.

Beuret, N., Stettler, H., Renold, A., Rutishauser, J., and Spiess, M. (2004). Expression of regulated secretory proteins is sufficient to generate granule-like structures in constitutively secreting cells. *J. Biol. Chem.* **279**, 20242-20249.

Bier, E. (2005). *Drosophila*, the golden bug, emerges as a tool for human genetics. *Nat. Rev. Genet.* **6**, 9-23.

Biggin, M.D., and Tjian, R. (2001). Transcriptional regulation in *Drosophila*: the post-genome challenge. *Funct. Integr. Genomics* **1**, 223-234.

Biro, A.A., Holderith, N.B., and Nusser, Z. (2005). Quantal size is independent of the release probability at hippocampal excitatory synapses. *J. Neurosci.* **25**, 223-232.

Borgonovo, B., Racchetti, G., Malosio, M., Benfante, R., Podini, P., Rosa, P., and Meldolesi, J. (1998). Neurosecretion competence, an independently regulated trait of the neurosecretory cell phenotype. *J. Biol. Chem.* **273**, 34683-34686.

Borgonovo, B., Ouwendijk, J., and Solimena, M. (2006). Biogenesis of secretory granules. *Curr. Opin. Cell Biol.* **18**, 365-370.

Borycz, J.A., Borycz, J., Kubow, A., Kostyleva, R., and Meinertzhagen, I.A. (2005). Histamine compartments of the *Drosophila* brain with an estimate of the quantum content at the photoreceptor synapse. *J. Neurophysiol.* **93**, 1611-1619.

Burbach, J.P.H. (2011). What are neuropeptides? *Methods Mol. Biol.* **789**, 1-36.

Burgess, T.L., and Kelly, R.B. (1987). Constitutive and regulated secretion of proteins. *Annu. Rev. Cell. Biol.* **3**, 243-293.

Burgoyne, R.D., and Morgan, A. (2003). Secretory granule exocytosis. *Physiol. Rev.* **83**, 581-632.

Carrera, I., and Treisman, J.E. (2008). Message in a nucleus: signaling to the transcriptional machinery. *Curr. Opin. Genet. Dev.* **18**, 397-403.

Ceccarelli, B., Grohovaz, F., and Hurlbut, W.P. (1979). Freeze-fracture studies of frog neuromuscular-junctions during intense release of neurotransmitter .2. Effects of electrical-stimulation and high potassium. *J. Cell Biol.* **81**, 178-192.

Cheng, L.P., Arata, A., Mizuguchi, R., Qian, Y., Karunaratne, A., Gray, P.A., Arata, S., Shirasawa, S., Bouchard, M., Luo, P., et al. (2004). Tlx3 and Tlx1 are post-mitotic selector genes determining glutamatergic over GABAergic cell fates. *Nat. Neurosci.* **7**, 510-517.

Collas, P. (2009). The state-of-the-art of chromatin immunoprecipitation. *Methods Mol. Biol.* **567**, 1-25.

Craige, B., Salazar, G., and Faundez, V. (2004). Isolation of synaptic vesicles. In *Current Protocols in Cell Biology* / editorial board, Juan S Bonifacino [et al] Chapter 3, Unit 3.12.

Crivellato, E., Nico, B., and Ribatti, D. (2005). Ultrastructural evidence of piecemeal degranulation in large dense-core vesicles of brain neurons. *Anat. Embryol.* **210**, 25-34.

Crivellato, E., Nico, B., Bertelli, E., Nussdorfer, G.G., and Ribatti, D. (2006). Dense-core granules in neuroendocrine cells and neurons release their secretory constituents by piecemeal degranulation (review). *Int. J. Mol. Med.* **18**, 1037-1046.

Day, R., Benjannet, S., Matsuuchi, L., Kelly, R.B., Marcinkiewicz, M., Chretien, M., and Seidah, N.G. (1995). Maintained PC1 and PC2 expression in the Att-20 variant cell-line 6t3 lacking regulated secretion and POMC - restored POMC expression and regulated secretion after cAMP treatment. *DNA Cell Biol.* **14**, 175-188.

De Camilli, P., and Jahn, R. (1990). Pathways to regulated exocytosis in neurons. *Annu. Rev. Physiol.* **52**, 625-645.

de Kloet, A.D., and Woods, S.C. (2010). Molecular neuroendocrine targets for obesity therapy. *Curr. Opin. Endocrinol. Diabetes Obes.* **17**, 441.

De Matteis, M.A., and Luini, A. (2008). Exiting the Golgi complex. *Nat. Rev. Mol. Cell Biol.* **9**, 273-284.

Dikeakos, J.D., and Reudelhuber, T.L. (2007). Sending proteins to dense core secretory granules: still a lot to sort out. *J. Cell. Biol.* **177**, 191-196.

Ding, Y.Q., Marklund, U., Yuan, W.L., Yin, J., Wegman, L., Ericson, J., Deneris, E., Johnson, R.L., and Chen, Z.F. (2003). Lmx1b is essential for the development of serotonergic neurons. *Nat. Neurosci.* **6**, 933-938.

Docherty, K., and Steiner, D.F. (1982). Post-translational proteolysis in polypeptide hormone biosynthesis. *Annu. Rev. Physiol.* **44**, 625-638.

Edwards, R.H. (1998). Neurotransmitter release: variations on a theme. *Curr. Biol.* **8**, R883–5.

Eipper, B.A., Perkins, S.N., Husten, E.J., Johnson, R.C., Keutmann, H.T., and Mains, R.E. (1991). Peptidyl-alpha-hydroxyglycine alpha-amidating lyase - purification, characterization, and expression. *J. Biol. Chem.* **266**, 7827-7833.

Eipper, B.A., Stoffers, D.A., and Mains, R.E. (1992). The biosynthesis of neuropeptides: peptide alpha-amidation. *Annu. Rev. Neurosci.* **15**, 57-85.

Eipper, B.A., Bloomquist, B.T., Husten, E.J., Milgram, S.L., and Mains, R.E. (1993). Peptidylglycine alpha-amidating monooxygenase and other processing enzymes in the neurointermediate pituitary. *Ann. NY Acad. Sci.* **680**, 147-160.

Epstein, C.B., and Butow, R.A. (2000). Microarray technology - enhanced versatility, persistent challenge. *Curr. Opin. Biotechnol.* **11**, 36-41.

Ewer, J. (2005). Behavioral actions of neuropeptides in invertebrates: insights from *Drosophila*. *Horm. Behav.* **48**, 418-429.

Finck, B.N., and Kelly, D.P. (2007). Peroxisome proliferator-activated receptor gamma coactivator-1 (PGC-1) regulatory cascade in cardiac physiology and disease. *Circulation* **115**, 2540-2548.

Flames, N., and Hobert, O. (2009). Gene regulatory logic of dopamine neuron differentiation. *Nature* **458**, 885-889.

Freeman, M.R., Delrow, J., Kim, J., Johnson, E., and Doe, C.Q. (2003). Unwrapping glial biology: Gcm target genes regulating glial development, diversification, and function. *Neuron* **38**, 567-580.

Fricker, L.D., and Snyder, S.H. (1982). Enkephalin convertase - purification and characterization of a specific enkephalin-synthesizing carboxypeptidase localized to adrenal chromaffin granules. *Proc. Natl. Acad. Sci. USA* **79**, 3886-3890.

Fricker, L.D. (2005). Neuropeptide-processing enzymes: applications for drug discovery. *AAPS J.* **7**, E449-455.

Fuda, N.J., Ardehali, M.B., and Lis, J.T. (2009). Defining mechanisms that regulate RNA polymerase II transcription in vivo. *Nature* **461**, 186-192.

Fuhlendorff, J., Gether, U., Aakerlund, L., Langeland-Johansen, N., Thøgersen, H., Melberg, S.G., Olsen, U.B., Thastrup, O., and Schwartz, T.W. (1990). [Leu31, Pro34]neuropeptide Y: a specific Y1 receptor agonist. *Proc. Natl. Acad. Sci. USA* **87**, 182-186.

Fukuda, M., Kanno, E., and Yamamoto, A. (2004). Rabphilin and Noc2 are recruited to dense-core vesicles through specific interaction with Rab27A in PC12 cells. *J. Biol. Chem.* **279**, 13065-13075.

Fuller, S.D., Bravo, R., and Simons, K. (1985). An enzymatic assay reveals that proteins destined for the apical or basolateral domains of an epithelial-cell line share the same late golgi compartments. *EMBO J.* **4**, 297-307.

Geuze, H.J., Slot, J.W., Strous, J.A.M., Hasilik, A., and Vonfigura, K. (1983). Possible pathways for lysosomal-enzyme delivery. *J. Cell Biol.* **101**, 2253-2262.

Gilmour, D.S., and Lis, J.T. (1984). Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. *Proc. Natl. Acad. Sci. USA* **81**, 4275-4279.

Gilmour, D.S., and Lis, J.T. (1987). Protein-DNA cross-linking reveals dramatic variation in RNA polymerase II density on different histone repeats of *Drosophila* melanogaster. *Mol. Cell. Biol.* **7**, 3341-3344.

Glembotski, C.C. (1981). Subcellular fractionation studies on the post-translational processing of pro-adrenocorticotropic hormone-endorphin in rat intermediate pituitary. *J. Biol. Chem.* **256**, 7433-7439.

Goshu, E., Jin, H., Lovejoy, J., Marion, J.F., Michaud, J.L., and Fan, C.M. (2004). Sim2 contributes to neuroendocrine hormone gene expression in the anterior hypothalamus. *Mol. Endocrinol.* **18**, 1251-1262.

Griffiths, G., Warren, G., Stuhlfauth, I., and Jockusch, B.M. (1981). The Role Of Clathrin-Coated Vesicles In Acrosome Formation. *Eur. J. Cell Biol.* **26**, 52-60.

Griffiths, G., McDowall, A., Back, R., and Dubochet, J. (1984). On the preparation of cryosections for immunocytochemistry. *J. Ultrastruct. Res.* **89**, 65-78.

Gumbiner, B., and Kelly, R.B. (1982). Two distinct intracellular pathways transport secretory and membrane glycoproteins to the surface of pituitary tumor cells. *Cell* **28**, 51-59.

Hamanaka, Y., Park, D., Yin, P., Annangudi, S.P., Edwards, T.N., Sweedler, J., Meinertzhagen, I.A., and Taghert, P.H. (2010). Transcriptional orchestration of the regulated secretory pathway in neurons by the bHLH protein DIMM. *Curr. Biol.* **20**, 9-18.

Han, M., Park, D., Vanderzalm, P.J., Mains, R.E., Eipper, B.A., and Taghert, P.H. (2004). *Drosophila* uses two distinct neuropeptide amidating enzymes, dPAL1 and dPAL2. *J. Neurochem.* **90**, 129-141.

Hashimoto, S., Fumagalli, G., Zanini, A., and Meldolesi, J. (1987). Sorting of three secretory proteins to distinct secretory granules in acidophilic cells of cow anterior pituitary. *J. Cell. Biol.* **105**, 1579-1586.

Havas S, Aronne LJ, Woodworth KA (2009). The problem of obesity in the United States has become increasingly prominent and is now recognized as a critical target for public health intervention. Introduction. *Am. J. Med.* **122**(4 Suppl 1):S1-3.

Hawkins, R.D., Hon, G.C., and Ren, B. (2010). Next-generation genomics: an integrative approach. *Nat. Rev. Genet.* **11**, 476-486.

Hearn, S.A., Silver, M.M., and Sholdice, J.A. (1985). Immunoelectron microscopic labeling of immunoglobulin in plasma cells after osmium fixation and epoxy embedding. *J. Histochem. Cytochem.* **33**, 1212-1218.

Hendricks, T., Francis, N., Fyodorov, D., and Deneris, E.S. (1999). The ETS domain factor Pet-1 is an early and precise marker of central serotonin neurons and interacts with a conserved element in serotonergic genes. *J. Neurosci.* **19**, 10348-10356.

Helfrich-Förster, C., Täuber, M., Park, J.H., Mühlig-Versen, M., Schneuwly, S., and Hofbauer, A. (2000). Ectopic expression of the neuropeptide pigment-dispersing factor alters behavioral rhythms in *Drosophila* melanogaster. *J. Neurosci.* **20**, 3339-3353.

Hewes, R.S., Park, D., Gauthier, S.A., Schaefer, A.M., and Taghert, P.H. (2003). The bHLH protein Dimmed controls neuroendocrine cell differentiation in *Drosophila*. *Development* **130**, 1771-1781.

Hindmarch, C., Yao, S., Beighton, G., Paton, J., and Murphy, D. (2006). A comprehensive description of the transcriptome of the hypothalamoneuro-hypophyseal system in euhydrated and dehydrated rats. *Proc. Natl. Acad. Sci. USA* **103**, 1609-1614.

Hindmarch, C.C.T., and Murphy, D. (2010). The transcriptome and the hypothalamo-neurohypophyseal system. *Endocr. Dev.* **17**, 1-10.

Hökfelt, T., Holets, V.R., Staines, W., Meister, B., Melander, T., Schalling, M., Schultzberg, M., Freedman, J., Björklund, H., and Olson, L. (1986). Coexistence of neuronal messengers--an overview. *Prog. Brain. Res.* **68**, 33–70.

Horodyski, F.M., Riddiford, L.M., and Truman, J.W. (1989). Isolation and expression of the eclosion hormone gene from the tobacco hornworm, Manduca sexta. *Proc. Natl. Acad. Sci. USA* **86**, 8123–8127.

Horodyski, F.M., Ewer, J., Riddiford, L.M., and Truman, J.W. (1993). Isolation, characterization and expression of the Eclosion hormone gene of *Drosophila*-melanogaster. *Eur. J Biochem*. **215**, 221-228.

Hosoya, T., Oda, Y., Takahashi, S., Morita, M., Kawauchi, S., Ema, M., Yamamoto, M., and Fujii-Kuriyama, Y. (2001). Defective development of secretory neurones in the hypothalamus of Arnt2-knockout mice. *Genes Cells* **6**, 361-374.

Hwang, J.R., Siekhaus, D.E., Fuller, R.S., Taghert, P.H., and Lindberg, I. (2000). Interaction of *Drosophila* melanogaster prohormone convertase 2 and 7B2 - Insect cell-specific processing and secretion. *J. Biol. Chem.* **275**, 17886-17893.

Jiang, N., Kolhekar, A.S., Jacobs, P.S., Mains, R.E., Eipper, B.A., and Taghert, P.H. (2000). PHM is required for normal developmental transitions and for biosynthesis of secretory peptides in *Drosophila*. *Dev. Biol.* **226**, 118-136.

Johnson, C.L., Kowalik, A.S., Rajakumar, N., and Pin, C.L. (2004). Mist1 is necessary for the establishment of granule organization in serous exocrine cells of the gastrointestinal tract. *Mech. Dev.* **121**, 261-272.

Kalra, S.P., and Kalra, P.S. (2010). Neuroendocrine control of energy homeostasis: update on new insights. *Prog. Brain Res.* **181**, 17-33.

Kato, I., Yonekura, H., Tajima, M., Yanagi, M., Yamamoto, H., and Okamoto, H. (1990). 2 enzymes concerned in peptide-hormone alpha-amidation are synthesized from a single messenger-RNA. *Biochem. Biophys. Res. Commun.* **172**, 197-203.

Katopodis, A.G., Ping, D.S., Smith, C.E., and May, S.W. (1991). Functional and structural characterization of peptidylamidoglycolate lyase, the enzyme catalyzing the 2nd step in peptide amidation. *Biochemistry* **30**, 6189-6194.

Katz B (1969). *The Release of Neural Transmitter Substances*. (Liverpool: Liverpool University Press)

Kelly, R.B. (1985). Pathways of protein secretion in eukaryotes. *Science* **230**, 25-32.

Kim, T., Tao-Cheng, J.H., Eiden, L.E., and Loh, Y.P. (2001). Chromogranin A, an "on/off" switch controlling dense-core secretory granule biogenesis. *Cell* **106**, 499-509.

Knecht, S., Ellger, T., and Levine, J.A. (2008). Obesity in neurobiology. *Prog. Neurobiol.* **84**, 85-103.

Kohwi, M., and Doe, C.Q. (2009). Stem Cell Transcriptional Loops Generate Precise Temporal Identity. *Cell Stem Cell* **5**, 577-578.

Kolhekar, A.S., Roberts, M.S., Jiang, N., Johnson, R.C., Mains, R.E., Eipper, B.A., and Taghert, P.H. (1997). Neuropeptide amidation in *Drosophila*: separate genes encode the two enzymes catalyzing amidation. *J. Neurosci.* **17**, 1363-1376.

Kopeć, S. (1917). Experiments on metamorphosis of insects. *Bull. Acad. Sci. Cracovie Sci. Math. Nat. Ser. B*. **1917**: 57-60

Kopec, S. (1922). Studies on the necessity of the brain for the inception of insect metamorphosis. *Biol. Bull. Woods Hole Mass.* **42**:323—42

Kreil, G (1985) in *The Enzymology of Post-translational Modification of Proteins,* R.B. Freedman, H.G. Hawkins, ed. *(London: Academic Press), pp. 41–51.*

Kula-Eversole, E., Nagoshi, E., Shang, Y., Rodriguez, J., Allada, R., and Rosbash, M. (2010). Surprising gene expression patterns within and between PDF-containing circadian neurons in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **107**, 13497-13502.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protoc.* **1**, 729-748.

Lehman, J.J., Barger, P.M., Kovacs, A., Saffitz, J.E., Medeiros, D.M., and Kelly, D.P. (2000). Peroxisome proliferator-activated receptor gamma coactivator-1 promotes cardiac mitochondrial biogenesis. *J. Clin. Invest.* **106**, 847-856.

Leng, G., and Ludwig, M. (2008). Neurotransmitters and peptides: whispered secrets and public announcements. *J. Physiol.* **586**, 5625-5632.

Leoni, C., Menegon, A., Benfenati, F., Toniolo, D., Pennuto, M., and Valtorta, F. (1999). Neurite extension occurs in the absence of regulated exocytosis in PC12 subclones. *Mol. Biol. Cell* **10**, 2919-2931.

Levin, B. (2009). Synergy of nature and nurture in the development of childhood obesity. *Int. J. Obes. (Lond)* **33**, S53-S56.

Levine, M., and Tjian, R. (2003). Transcription regulation and animal diversity. *Nature* **424**, 147-151.

Li, X.-Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C.L., et al. (2008). Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* **6**, e27.

Littleton, J.T. (2000). A genomic analysis of membrane trafficking and neurotransmitter release in *Drosophila*. *J. Cell. Biol.* **150**, F77-82.

Loewi, O. (1921) "Über humorale Übertragbarkeit der Herznervenwirkung. I." *Pflügers Archiv*, **189**, pp. 239-242

Ludwig, M., and Leng, G. (2006). Dendritic peptide release and peptide-dependent behaviours. *Nat. Rev. Neurosci.* **7**, 126-136.

Macquarrie, K.L., Fong, A.P., Morse, R.H., and Tapscott, S.J. (2011). Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet.* **27**, 141-148.

Malosio, M.L., Benfante, R., Racchetti, G., Borgonovo, B., Rosa, P., and Meldolesi, J. (1999). Neurosecretory cells without neurosecretion: evidence of an independently regulated trait of the cell phenotype. *J. Physiol.* **520**, 43-52.

Martin, T.F.J. (2003). Tuning exocytosis for speed: fast and slow modes. *Biochim. Biophys. Acta* **1641**, 157-165.

Massari, M.E., and Murre, C. (2000). Helix-loop-helix proteins: Regulators of transcription in eucaryotic organisms. *Mol. Cell. Biol.* **20**, 429-440.

Matsui, K., and Jahr, C.E. (2006). Exocytosis unbound. *Curr. Opin. Neurobiol.* **16**, 305–311.

Matsuuchi, L., and Kelly, R.B. (1991). Constitutive and basal secretion from the endocrine cell-line, Att-20. *J. Cell Biol.* **112**, 843-852.

Meldolesi, J., Chieregatti, E., and Luisa Malosio, M. (2004). Requirements for the identification of dense-core granules. *Trends Cell. Biol.* **14**, 13-19.

Mellman, I., and Simons, K. (1992). The Golgi complex: in vitro veritas? *Cell* **68**, 829-840.

Mesce, K.A., and Fahrbach, S.E. (2002). Integration of endocrine signals that regulate insect ecdysis. *Front. Neuroendocrinol.* **23**, 179-199.

Michaud, J.L., Rosenquist, T., May, N.R., and Fan, C.M. (1998). Development of neuroendocrine lineages requires the bHLH-PAS transcription factor SIM1. *Genes Dev.* **12**, 3264-3275.

Miguel-Aliaga, I., Thor, S., and Gould, A.P. (2008). Postmitotic specification of *Drosophila* insulinergic neurons from pioneer neurons. *PLoS Biol.* **6**, 538-551.

Mills, J.C., and Taghert, P.H. (2011). Scaling factors: transcription factors regulating subcellular domains. *Bioessays* **34**, 10–16.

Moore, A.W., Barbel, S., Jan, L.Y., and Jan, Y.N. (2000). A genomewide survey of basic helix-loop-helix factors in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **97**, 10436-10441.

Moore, H.P., and Kelly, R.B. (1985). Secretory protein targeting in a pituitary cell line: differential transport of foreign secretory proteins to distinct secretory pathways. *J. Cell. Biol.* **101**, 1773-1781.

Morris, J.F., and Pow, D.V. (1991). Widespread release of peptides in the central-nervous-system - quantitation of tannic acid-captured exocytoses. *Anat. Rec. (Hoboken)* **231**, 437-445.

Morvan, J., and Tooze, S.A. (2008). Discovery and progress in our understanding of the regulated secretory pathway in neuroendocrine cells. *Histochem. Cell Biol.* **129**, 243-252.

Mukhopadhyay, A., Deplancke, B., Walhout, A.J.M., and Tissenbaum, H.A. (2008). Chromatin immunoprecipitation (ChIP) coupled to detection by quantitative real-time PCR to study transcription factor binding to DNA in Caenorhabditis elegans. *Nat. Protoc.* **3**, 698-709.

Nakai, S., Kawano, H., Yudate, T., Nishi, M., Kuno, J., Nagata, A., Jishage, K., Hamada, H., Fujii, H., Kawamura, K., et al. (1995). The POU domain transcription factor Brn-2 is required for the determination of specific neuronal lineages in the hypothalamus of the mouse. *Genes Dev.* **9**, 3109-3121.

Naser, K.A., Gruber, A., and Thomson, G.A. (2006). The emerging pandemic of obesity and diabetes: are we doing enough to prevent a disaster? *Int. J. Clin. Pract.* **60**, 1093-1097.

Nässel, D.R. (2002). Neuropeptides in the nervous system of *Drosophila* and other insects: multiple roles as neuromodulators and neurohormones. *Prog. Neurobiol.* **68**, 1-84.

Nègre, N., Lavrov, S., Hennetin, J., Bellis, M., and Cavalli, G. (2006). Mapping the distribution of chromatin proteins by ChIP on chip. *Methods Enzymol.* **410**, 316-341.

Nickel, W. (2010). Pathways of unconventional protein secretion. *Curr. Opin. Biotechnol.* **21**, 621-626.

Owald, D., and Sigrist, S.J. (2009). Assembling the presynaptic active zone. *Curr. Opin. Neurobiol.* **19**, 311-318.

Ozsolak, F., and Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87-98.

Pance, A., Livesey, F., and Jackson, A. (2006). A role for the transcriptional repressor REST in maintaining the phenotype of neurosecretory-deficient PC12 cells. *J. Neurochem.* **99**, 1435-1444.

Pance, A., Morgan, K., Guest, P.C., Bowers, K., Dean, G.E., Cutler, D.F., and Jackson, A.P. (1999). A PC12 variant lacking regulated secretory organelles: Aberrant protein targeting and evidence for a factor inhibiting neuroendocrine gene expression. *J. Neurochem.* **73**, 21-30.

Park, D., Hadzic, T., Yin, P., Rusch, J., Abruzzi, K., Rosbash, M., Skeath, J.B., Panda, S., Sweedler, J.V., and Taghert, P.H. (2011). Molecular organization of *Drosophila* neuroendocrine cells by dimmed. *Curr. Biol.* **21**(18):1515-24

Park, D., Han, M., Kim, Y.-C., Han, K.-A., and Taghert, P.H. (2004). Ap-let neurons--a peptidergic circuit potentially controlling ecdysial behavior in *Drosophila*. *Dev. Biol.* **269**, 95-108.

Park, D., Veenstra, J.A., Park, J.H., and Taghert, P.H. (2008a). Mapping peptidergic cells in *Drosophila*: where DIMM fits in. *PLoS ONE* **3**, e1896.

Park, D., Shafer, O.T., Shepherd, S.P., Suh, H., Trigg, J.S., and Taghert, P.H. (2008b). The *Drosophila* basic helix-loop-helix protein DIMMED directly activates PHM, a gene encoding a neuropeptide-amidating enzyme. *Mol. Cell. Biol.* **28**, 410-421.

Park, D., and Taghert, P.H. (2009). Peptidergic neurosecretory cells in insects: organization and control by the bHLH protein DIMMED. *Gen. Comp. Endocrinol.* **162**, 2-7.

Park, J.J., and Loh, Y.P. (2008). How peptide hormone vesicles are transported to the secretion site for exocytosis. *Mol. Endocrinol.* **22**, 2583-2595.

Park, P.J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10**, 669-680.

Pécot-Dechavassine, M., and Brouard, M.O. (1997). Large dense-core vesicles at the frog neuromuscular junction: characteristics and exocytosis. *J. Neurocytol.* **26**, 455-465.

Perello, M., and Nillni, E.A. (2007). The biosynthesis and processing of neuropeptides: lessons from prothyrotropin releasing hormone (proTRH). *Front. Biosci.* **12**, 3554-3565.

Peters, A., Palay, S.L., and Webster, H. (1991). *The Fine Structure of the Nervous System*, Third Edition (New York: Oxford University Press).

Pin, C.L., Bonvissuto, A.C., Konieczny, S.F. (2000). Mist1 expression is a common link among serous exocrine cells exhibiting regulated exocytosis. *Anat. Rec.* **259**, 157–167.

Pothos, E.N., Mosharov, E., Liu, K.-P., Setlik, W., Haburcak, M., Baldini, G., Gershon, M.D., Tamir, H., and Sulzer, D. (2002). Stimulation-dependent regulation of the pH, volume and quantal size of bovine and rodent secretory vesicles. *J. Physiol.* **542**, 453-476.

Powell, L.M., and Jarman, A.P. (2008). Context dependence of proneural bHLH proteins. *Curr. Opin. Genet. Dev.* **18**, 411-417.

Ramsey, V.G., Doherty, J.M., Chen, C.C., Stappenbeck, T.S., Konieczny, S.F., and Mills, J.C. (2007). The maturation of mucus-secreting gastric epithelial progenitors into digestive-enzyme secreting zymogenic cells requires Mist1. *Development* **134**, 211-222.

Rao, S., Lang, C., Levitan, E.S., and Deitcher, D.L. (2001). Visualization of neuropeptide expression, transport, and exocytosis in *Drosophila* melanogaster. *J Neurobiol* **49**, 159-172.

Rayburn, L.Y.M., Gooding, H.C., Choksi, S.P., Maloney, D., Kidd, A.R., Siekhaus, D.E., and Bender, M. (2003). amontillado, the *Drosophila* homolog of the prohormone processing protease PC2, is required during embryogenesis and early larval development. *Genetics* **163**, 227-237.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306.

Reynolds, S.E., Taghert, P.H., and Truman, J.W. (1979). Eclosion Hormone And Bursicon Titers And The Onset Of Hormonal Responsiveness During The Last Day Of Adult Development In Manduca-Sexta (L). *J Exp. Biol.* **78**, 77-86.

Rutter, G.A., and Tsuboi, T. (2004). Kiss and run exocytosis of dense core secretory vesicles. *Neuroreport* **15**, 79-81.

Salio, C., Lossi, L., Ferrini, F., and Merighi, A. (2006). Neuropeptides as synaptic transmitters. *Cell Tissue Res.* **326**, 583-598.

Salpeter, M.M., and Farquhar, M.G. (1981). High resolution analysis of the secretory pathway in mammotrophs of the rat anterior pituitary. *J. Cell. Biol.* **91**, 240-246.

Sandmann, T., Jakobsen, J., and Furlong, E. (2007). ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in *Drosophila* melanogaster embryos. *Nat Protoc* **1**, 2839-2855.

Sardiello, M., Palmieri, M., di Ronza, A., Medina, D.L., Valenza, M., Gennarino, V.A., Di Malta, C., Donaudy, F., Embrione, V., Polishchuk, R.S.*, et al.* (2009). A gene network regulating lysosomal biogenesis and function. *Science* **325**, 473-477.

Sargent, P.B., Saviane, C., Nielsen, T.A., DiGregorio, D.A., and Silver, R.A. (2005). Rapid vesicular release, quantal variability, and spillover contribute to the precision and reliability of transmission at a glomerular synapse. *J. Neurosci.* **25**, 8173-8187.

Scharrer, B., Scharrer, E. (1944). Neuro-secretion. VI. A comparison between the intercerebralis-cardiacum-allatum system of the insects and the hypothalamo-hypo-physeal system of the vertebrates. *Biol. Bull.* **87** : 243-5 1

Scharrer, B. (1987a). Neurosecretion: beginnings and new directions in neuropeptide research. *Annu. Rev. Neurosci.* **10**, 1-17.

Scharrer, B. (1987b). Insects as models in neuroendocrine research. *Annu. Rev. Entomol.* **32**, 1-16.

Scharrer, E. (1952). The general significance of the neurosecretory cell. *Scientia* **46**: 177-83

Scharrer, E., Brown, S. (1961). Neurosecretion. XII. The formation of neuro- secretory granules in the earthworm, Lumbricus terrestris *L. Z. Zell Forsch.* **54** : 530- 40

Scharrer, E., and Scharrer, B. (1937) Über Drüsen-Nervenzellen und neurosekretorische Organe bei Wirbellosen und Wirbeltieren. *Biol. Rev.* **12**, 185–216.

Schmidt, K., and Stephens, D.J. (2010). Cargo loading at the ER. *Mol. Membr. Biol.* **27**, 398-411.

Schnabel, E., Mains, R.E., and Farquhar, M.G. (1989). Proteolytic Processing Of Pro-Acth Endorphin Begins In The Golgi-Complex Of Pituitary Corticotropes And Att-20 Cells. *Mol. Endocrinol.* **3**, 1223-1235.

Schonemann, M.D., Ryan, A.K., McEvilly, R.J., Oconnell, S.M., Arias, C.A., Kalla, K.A., Li, P., Sawchenko, P.E., and Rosenfeld, M.G. (1995). Development and survival of the endocrine hypothalamus and posterior pituitary gld requires the neuronal POU domain factor Brn-2. *Genes Dev.* **9**, 3122-3135.

Seidah, N.G., and Chrétien, M. (1999). Proprotein and prohormone convertases: a family of subtilases generating diverse bioactive polypeptides. *Brain Res.* **848**(1-2): 45-62.

Seidah, N.G., Marcinkiewicz, M., Benjannet, S., Gaspar, L., Beaubien, G., Mattei, M.G., Lazure, C., Mbikay, M., and Chretien, M. (1991). Cloning And Primary Sequence Of A Mouse Candidate Prohormone Convertase PC1 Homologous To PC2, Furin, And Kex2 - Distinct Chromosomal Localization And Messenger-RNA Distribution In Brain and Pituitary Compared to PC2. *Mol. Endocrinol.* **5**, 111-122.

Seidah, N.G., and Prat, A. (2002). Precursor convertases in the secretory pathway, cytosol and extracellular milieu. In *Proteases in Biology and Medicine*, N.M. Hooper, ed. (Portland, Oregon: Portland Pr), pp. 79-94.

Sidyelyeva, G., Baker, N.E., and Fricker, L.D. (2006). Characterization of the molecular basis of the *Drosophila* mutations in carboxypeptidase D. Effect on enzyme activity and expression. *J. Biol. Chem.* **281**, 13844-13852.

Siekhaus, D.E., and Fuller, R.S. (1999). A role for amontillado, the *Drosophila* homolog of the neuropeptide precursor processing protease PC2, in triggering hatching behavior. *J. Neurosci.* **19**, 6942-6954.

Silver, R.A., Lubke, J., Sakmann, B., and Feldmeyer, D. (2003). High-probability uniquantal transmission at excitatory synapses in barrel cortex. *Science* **302**, 1981-1984.

Simon, I., Barnett, J., Hannett, N., Harbison, C.T., Rinaldi, N.J., Volkert, T.L., Wyrick, J.J., Zeitlinger, J., Gifford, D.K., Jaakkola, T.S., et al. (2001). Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* **106**, 697-708.

Smeekens, S.P., and Steiner, D.F. (1990). Identification of a Human Insulinoma cDNA-encoding a Novel Mammalian Protein Structurally Related To The Yeast Dibasic Processing Protease Kex2. *J. Biol. Chem.* **265**, 2997-3000.

Solomon, M.J., Larsen, P.L., and Varshavsky, A. (1988). Mapping protein-DNA interactions in vivo with formaldehyde: evidence that histone H4 is retained on a highly transcribed gene. *Cell* **53**, 937-947.

Southall, T.D., and Brand, A.H. (2007). Chromatin profiling in model organisms. *Brief. Funct. Genomic. Proteomic.* **6**, 133-140.

Speidel, C. C. (1919). Gland-cells of internal secretion in the spinal cord of the skates. *Carnegie Inst. Washington Publ.* No. **13**: 1-3 1

Stjarne L (2000) Do sympathetic nerves release noradrenaline in “quanta”? *J. Auton. Nerv. Syst.* **81**: 236–43

Südhof, T.C. (2004). The synaptic vesicle cycle. *Annu. Rev. Neurosci.* **27**, 509-547.

Südhof, T.C. (2007). Neurotransmitter Release. *Handb. Exp. Pharmacol.* **184**, 1-21.

Taghert, P.H., and Truman, J.W. (1982). Identification Of The Bursicon-Containing Neurons In Abdominal Ganglia Of The Tobacco Hornworm, Manduca-Sexta. *J. Exp. Biol.* **98**, 385.

Tazi, A., Dantzer, R., Lemoal, M., Rivier, J., Vale, W., and Koob, G.F. (1987). Corticotropin-Releasing Factor Antagonist Blocks Stress-Induced Fighting In Rats. *Regul. Pept.* **18**, 37-42.

Thiele, C., and Huttner, W.B. (1998). Protein and lipid sorting from the trans-Golgi network to secretory granules-recent developments. *Semin. Cell Dev. Biol.* **9**, 511-516.

Tooze, J., and Tooze, S.A. (1986). Clathrin-coated vesicular transport of secretory proteins during the formation of ACTH-containing secretory granules in AtT20 cells. *J. Cell. Biol.* **103**, 839-850.

Tooze, S.A., and Huttner, W.B. (1990). Cell-Free Protein Sorting To The Regulated And Constitutive Secretory Pathways. *Cell* **60**, 837-847.

Tooze, S.A., Martens, G.J., and Huttner, W.B. (2001). Secretory granule biogenesis: rafting to the SNARE. *Trends Cell Biol.* **11**, 116-122.

Tooze, S.A., and Stinchcombe, J.C. (1992). Biogenesis of secretory granules. *Semin. Cell Biol.* **3**, 357-366.

Toth, J., and Biggin, M.D. (2000). The specificity of protein-DNA crosslinking by formaldehyde: in vitro and in *Drosophila* embryos. *Nucleic Acids Res.* **28**, e4.

Truman, J.W. (2005). Hormonal control of insect ecdysis: endocrine cascades for coordinating behavior with physiology. *Vitam. Horm.* **73**, 1-30.

Truman, J.W., and Riddiford, L.M. (1970). Neuroendocrine control of ecdysis in silkmoths. *Science* **167**, 1624–1626.

Tsigos, C., and Chrousos, G.P. (2002). Hypothalamic-pituitary-adrenal axis, neuroendocrine factors and stress. *J. Psychosom. Res.* **53**, 865-871.

Vazquez-Martinez, R., Diaz-Ruiz, A., Almabouada, F., Rabanal-Ruiz, Y., Gracia-Navarro, F., and Malagón, M.M. (2011). Revisiting the regulated secretory pathway: From frogs to human. *Gen. Comp. Endocrinol.* **1**-9.

Von Euler, U. S., Gaddum, J. H. (1931). An unidentified depressor substance in certain tissue extracts. *J. Physiol.* **72**: 74-87

Wang, W.D., and Lufkin, T. (2000). The murine Otp homeobox gene plays an essential role in the specification of neuronal cell lineages in the developing hypothalamus. *Dev. Biol.* **227**, 432-449.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* **10**, 57-63.

Watson, A., Mazumder, A., Stewart, M., and Balasubramanian, S. (1998). Technology for microarray analysis of gene expression. *Curr. Opin. Biotechnol.* **9**, 609-614.

Winkler, H., and Fischer-Colbrie, R. (1998). Regulation of the biosynthesis of large dense-core vesicles in chromaffin cells and neurons. *Cell. Mol. Neurobiol.* **18**, 193-209.

Wyrick, J.J., and Young, R.A. (2002). Deciphering gene expression regulatory networks. *Curr. Opin. Genet. Dev.* **12**, 130-136.

Zanini, A., Giannattasio, G., Nussdorfer, G., Margolis, R., Margolis, R., and Meldolesi, J. (1980). Molecular organization of prolactin granules. II. Characterization of glycosaminoglycans and glycoproteins of the bovine prolactin matrix. *J. Cell. Biol.* **86**, 260-272.

Zenisek, D., Steyer, J.A., and Almers, W. (2000). Transport, capture and exocytosis of single synaptic vesicles at active zones. *Nature* **406**, 849-854.

Zhou, A., Webb, G., Zhu, X., and Steiner, D.F. (1999). Proteolytic processing in the secretory pathway. *J. Biol. Chem.* **274**, 20745-20748.

Zimmermann, R., Eyrisch, S., Ahmad, M., and Helms, V. (2011). Protein translocation across the ER membrane. *Biochim. Biophys. Acta* **1808**, 912-924.

Zimmet, P., Alberti, K.G., and Shaw, J. (2001). Global and societal implications of the diabetes epidemic. *Nature* **414**, 782-787.

Zupanc, G.K. (1996). Peptidergic transmission: from morphological correlates to functional implications. *Micron* **27**, 35-91.

Chapter 2.

Molecular organization of *Drosophila* neuroendocrine cells by *dimmed*

This chapter contains the following manuscript:
Park, D., Hadžić, T., Yin, P., Rusch, J., Abruzzi, K., Rosbash, M., Skeath, J.B., Panda, S., Sweedler, J.V., and Taghert, P.H. (2011). Molecular organization of Drosophila neuroendocrine cells by dimmed. *Curr. Biol.* 21(18):1515-24

I contributed the following to this manuscript: ChIP analysis (Figure 4 and Table S2): genetics, ChIP and analysis by qPCR), image in Figure 6F (genetics, immunohistochemistry and confocal microscopy), all bioinformatics analysis: comparison of DIMM-dependent embryonic transcripts and normal large LNv transcripts in the adult brain (Figure 7), analysis of embryo microarrays (Table S3, Table S4), analysis of adult microarrays (Table S7 and Table S8).

ABSTRACT

In *Drosophila* the bHLH protein DIMM coordinates the molecular and cellular properties of all major neuroendocrine cells, irrespective of the secretory peptides they produce. When expressed by non-neuroendocrine neurons, DIMM confers the major properties of the Regulated Secretory Pathway and converts such cells away from fast neurotransmission and towards a neuroendocrine state. We first identified 134 transcripts upregulated by DIMM in embryos, then evaluated them systematically using diverse assays (including embryo *in situ* hybridization, *in vivo* ChIP, and cell-based transactivation assays). We conclude that of 11 strong candidates, six are strongly and directly controlled by DIMM *in vivo*. The six targets include several large dense-core vesicle (LDCV) proteins, but also proteins in non-LDCV compartments such as the RNA-associated protein MAELSTROM. In addition, a functional *in vivo* assay, combining transgenic RNAi with MS-based peptidomics, revealed that three DIMM targets are especially critical for its action: These include two well-established LDCV proteins, the amidation enzyme PHM and the ascorbate-regenerating electron transporter Cytochrome-$b_{561\text{-}1}$. The third key DIMM target, CAT-4 (*CG13248)*, has not previously been associated with peptide neurosecretion – it encodes a putative cationic amino acid transporter, closely related to the SLIMFAST Arginine transporter. Finally, we compared transcripts upregulated by DIMM with those normally enriched in DIMM neurons of the adult brain and found an intersection of 18 DIMM-regulated genes, which included all six direct DIMM targets. The results provide a rigorous molecular framework with which to describe the fundamental regulatory organization of diverse neuroendocrine cells.

INTRODUCTION

Neuroendocrine (NE) cells embody highly dedicated secretory cell states. While different NE cells express unique arrays of neuropeptide/ neurohormone-encoding genes, all peptidergic

NE cells nevertheless share many critical cellular functions. These functions reflect a common need for enzymes that conduct the post-translational processing of numerous neuropeptide precursors, and a need for the structural and regulatory components that execute large dense-core vesicle (LDCV) biogenesis, packaging, maturation and trafficking [1-2]. They also exhibit a choreographed capacity to modify their cell properties in response to changing physiological needs [3]. Many proteins are normally enriched in neuroendocrine tissues, and these may be coordinately regulated under different physiological states [4-5]. Hence, NE programs of cell differentiation must reflect the operations of complex regulatory circuits, the details of which are only beginning to emerge. Examining the intracellular regulatory pathways that organize and modulate these specialized properties is an important question for cell biology, and specifically is also critical to understanding NE cell physiology.

Important clues to understanding peptidergic cell biology may come from parallel studies of other neuronal secretory systems. Developmental programs of neurotransmitter expression are governed by dedicated transcriptional organizers. For example, serotonergic neuron differentiation (but not fate specification or survival) is substantially controlled by an ETS domain transcription factor called PET-1 [6]. Likewise, dopaminergic (DA) neuron maturation is promoted by the ETS transcription factor, AST-1, although it is not necessary for DA neuron generation or survival [7]. To what extent therefore, do peptidergic secretory cells rely on similar transcriptional regulatory controls?

In the *Drosophila* model system, the transcription factor DIMM operates in peptidergic NE cells in many ways similar to PET-1 and AST-1 in aminergic cells. DIMM is specifically expressed in peptidergic NE cells [8-12] – termed LEAP cells (***L****arge,* ***E****pisodically-* ***R****eleasing,* ***A****midating* ***P****eptide producing* [13]. Notably, DIMM acts like a master cell regulator for professional secretory cell properties. It confers two cardinal features of the Regulated Secretory Pathway (RSP) [14] onto neurons that otherwise do not display such properties [15].

The first property is generation and accumulation of large dense core vesicles (LDCV), which can house neuropeptides if these are available. Second it activates the complete post-translational processing machinery, which is sufficient to produce biologically active peptides from neuropeptide precursors. These cell biological observations raise important questions as to how DIMM, a single transcription factor, can efficiently organize a complete and functional sub-cellular domain like the RSP. A large part of the answer must lie in the nature of the target genes that DIMM activates, and this has motivated our current efforts.

We reasoned that direct targets provide the most compelling basis to describe DIMM action mechanisms. Here we describe a genome-wide screen to define candidate targets for DIMM regulation and found 11 genes rapidly up-regulated following DIMM over-expression in the embryo. We then extended the analysis with several independent measures and concluded that at least six of the 11 genes are authentic direct DIMM targets. Further, we show that three of these six are essential to the promotion of neuroendocrine cell properties by DIMM. The novel targets and the molecular details of DIMM transcriptional regulation provide a molecular foundation to help define the development and physiology of NE cells.

RESULTS

Identification of candidate genes as putative direct DIMM targets. To amass candidate DIMM targets in addition to *Phm* [12], we used genome-wide microarray profiling by over-expressing DIMM throughout the embryonic nervous system. We compared profiles from experimental (*elav>dimm*) and control (*elav*-GAL4) embryos at 22-26 hr and 28-32 hr after egg laying (AEL). The design was intended to identify transcripts consistently up-regulated shortly after the induction of DIMM, reasoning that the targets most likely to be directly affected by DIMM would be the earliest changed. We focused on genes that were up-regulated at least 1.5 fold following DIMM over-expression, at both time points: this process identified 134 candidate DIMM targets (Table S3). Many genes were also down-regulated (Table S4), but we did not

study these further.

We used quantitative real time PCR to verify the authenticity of genes affected in the embryo by DIMM-over-expression: by this method, 22 genes were confirmed to be up-regulated at least 1.5-fold by DIMM overexpression and of these, 18 genes up-regulated at least 2-fold (Table S5). Many of the 134 candidates were not tested based on low expression levels, or were tested but not detected by qPCR at this stage, or had primer sets that failed quality control tests. Finally, we focused on the 18 most highly-elevated transcripts and reasoned that if these 18 genes were true DIMM targets, they would be expressed in the central nervous system (CNS), and their RNA levels would decline in a *dimm* loss-of-function mutant. Using conventional PCR, we found that transcripts of all 18 genes were in fact expressed in CNS (Table S5). In addition, in 1st instar larvae of severe *dimm* hypomorphs (*Rev4/Rev8*: [8]), transcript levels for 11 of the 18 genes were significantly decreased. We therefore focused on these 11 genes: their identities, and GO terms, are listed in Table S6.

Consistent with previous findings, *Phm* is included in the 11-gene list, providing critical internal validation to the profiling method just described. In addition, *CG1275* is notable because it encodes a cytochrome (Cyt) trans-membrane electron transporter most related to the $b_{561\text{-}1}$ protein [16]. The Cyt-$b_{561\text{-}1}$ is a specific to dense-core secretory vesicles that contain catecholamines or amidated peptides, as are found in chromaffin cells and elsewhere. Cyt-$b_{561\text{-}1}$ regenerates ascorbate which is an essential co-factor to support the biosynthetic functions of peptidylglycine-a-hydroxylating monooxygenase (PHM) [17-18] from within the vesicles. The remaining nine candidates were not as readily associated with participation in the peptidergic secretory pathway; however, we have also evaluated them experimentally and discuss that work below.

Evaluating the 11 candidate DIMM targets. We performed three independent experiments to evaluate the degree to which the 11 candidate targets are strongly regulated by

DIMM, and devised a functional screen to determine the potential contributions of the candidates to DIMM action mechanisms. Figure 1 presents an overview of the work plan. As summarized in Table 1 and argued in the Discussion, these experiments suggest the 11 targets are divisible into a "strongly-regulated" set of six genes (direct regulation), two additional ones that may be directly-targeted, and another set of three likely indirect targets.

Experiment I. Embryonic RNA in situ hybridization. We first analyzed transcript distributions for each of the 11 candidates in embryos of diverse stages, and of two different genotypes (control and *elav>dimm* – see Methods for precise genotypes). Representative specimens are shown in Figure 2; control specimens are on the left and *dimm* over-expression specimens are on the right of each column. By Stage 16-17 in control embryos, we observed a clear cell-specific pattern of *dimm* transcripts in the CNS ([8]; Figure 2A). Up-regulation of *dimm* by *elav*GAL4 produced the expected pan-neuronal activation of *dimm* transcripts (Figure 2A'). Transcripts for the only previously known direct DIMM target, *Phm*, display a moderate level of expression in control embryonic tissues (Figure 2B) and clear up-regulation with DIMM over-expression (Figure 2B'), although the up-regulation was not as pronounced as the one shown by *dimm* transcripts.

Seven other candidates (*CG1275, CG13248, mael (CG11254), CG17293, CG7785*, *CG32850 and CG6522*) displayed weak-to-moderate expression levels in control embryos (Figure 2C-H). In the case of *CG6522,* transcripts were moderately to broadly expressed in normal tissues, and were modestly increased with *dimm* over-expression (data not shown). None of these candidate targets displayed a normal expression pattern exactly matching that of *dimm*. *CG13248* (Figure 2D) and *CG32850* (Figure 2H) were broadly expressed with very prominent accumulation in paired mid-line cells of the ventral nerve cord. In the cases of *Phm* and *CG13248,* those patterns were present early (~Stage 13) but resolved to stronger staining in lateral CNS regions later in embryogenesis that resembled that of *dimm* (Figure S1). *mael*

(*CG112545*) transcripts were not clearly detected in normal CNS (Figure 2E) although we clearly observed *mael* in normal germ cells (cf. [19]). *CG17293* was weakly expressed throughout the CNS (Figure 2F), and *CG7785* moderately expressed (Figure 2G).

DIMM over-expression resulted in heightened target RNA accumulation for seven of the 11 candidate targets. The *dimm*-stimulated transcript patterns varied, but did show some similarities: most prominently, many patterns included heavy accumulation in the paired midline cells (Figures 2B', C', D', E', G' and H'). Such mid-line cells do not normally express high levels of *dimm* (Figure 2A). Transcripts for the three other candidate targets (*CG14621, CG31346, and pastrel (CG8588*) were not detectable in control CNS (although some were evident in non-neuronal tissues). Likewise, we could not detect them following *dimm* over-expression (data not shown).

In summary, seven of the 11 candidate genes are regulated by DIMM according to embryonic stage *in situ* hybridization experiments – *Phm*, *Cyt-b*$_{561\text{-}1}$, *CG13248*, *mael*, *CG17293*, *CG7785*, and *CG32850*.

Experiment II. In vitro trans-activation assay: Our second test for strength of regulation measured DIMM's ability to trans-activate regulatory fragments of the candidate gene targets using Luciferase levels as a readout. We generated test constructs containing the *luciferase* reporter downstream of a mini-SV promoter and the putative DIMM-binding locus of the gene. Putative DIMM binding sites were selected by the presence and locations of specific E-box sequences (CATATG or CAGCTG) based on our previous analysis [12]: therefore, we focused especially on E-boxes within the first intron. Most candidate genes contained either CATATG or CAGCTG E-box sequences throughout the gene locus (Figure S2). In addition, we also combined the data from ChIP analysis (see below). Although our earlier work on DIMM regulation of *Phm* utilized mammalian *hEK-293* cells [12], we wished to approximate a more homologous cellular context, and so tested a *Drosophila* neuronal cell line, BG3-c2 [20]. First,

we confirmed that the positive control (*Phm*-sv-Luc) displayed the expected responsiveness (~10 fold induction) when *dimm* is co-transfected, but not when a mutant *dimm* isoform was co-transfected, all such results consistent with our previous observations [12]. Next, we measured *dimm*-responsiveness for each of the candidate genes. We found that in addition to *Phm*, five candidate DIMM targets (*CG1275*, *CG13248*, *mael*, *CG17293*, and *CG6522*) all displayed significant transactivation responses to DIMM (Figure 3).

To evaluate DIMM transactivation in more detail, we constructed a series of sequence variants of the *CG13248* responsive fragment – the ~1 kB fragment that comprises the 1st intron and which contain three E boxes. We mutated three of the six consensus bases of the E boxes: we tested each mutated E box in single, double and triple format: the positions of the E boxes and the results obtained are shown in Figure 3B. Once again, the wild type *CG13248* sequence produced an activity level comparable to that of *Phm*. Of the three single E box mutations, only E3 proved necessary for full DIMM transactivation. Furthermore, all mutant combinations containing the E3 mutation were not at all transactivated. Finally, the E1/E2 double mutant showed a diminished level of transactivation. In summary, the evidence points to the involvement of multiple E boxes in the transactivation of *CG13248*, with the E3 sequence playing the largest role.

Experiment III. In vivo Chromatin immunopreciptation (ChIP). To determine the potential occupancy by DIMM at candidate target gene binding sites *in vivo*, we used an epitope-tagged *dimm* transgene and ChIP methods. The transgene - UAS-*dimm*-MYC (II) - was selectively expressed in *dimm*-containing neurons by using the *c929*-GAL4 driver. Furthermore, we restricted GAL4: UAS activities to adult stages using temperature-sensitive GAL^{80} (*tub*-gal80^{ts}) to avoid the lethality that results from *dimm* over-expression at earlier stages (T Hadzic, unpublished). After raising the flies at the restrictive temperature (18^{o}C), 1-3 day old adults were transferred to 29^{o}C for a three-day period to permit expression of DIMM-MYC protein. We

confirmed DIMM-MYC expression under these conditions by Western blot analysis of fly head extracts, and an absence of such without transfer to 29^{o}C (data not shown).

In a previous report, we demonstrated that DIMM activates *Phm* directly via three palindromic E-boxes located in the *Phm* 1st intron (having sequences – CATATG and CAGCTG) [12]. Therefore, as a positive control, we asked whether DIMM is resident at this *Phm* intron and took a comparative approach by testing each of two sites in the *Phm* gene locus for DIMM occupancy – the E-box-containing test site within the 1st intron, and a control site located about 6 kb upstream (See Figure S2; control sites listed in Table S2). Across two biological replicates, DIMM-MYC CHIP'ed samples for the *Phm* gene showed a strong (~32 fold) average enrichment over samples from the negative control genotype (*c929*-GAL4; *tub*-gal80^{ts}). This difference was significantly different ($p < 0.05$) from the average enrichment found at its negative control site. We then asked whether DIMM is resident at the other target genes *in vivo* and selected test sites based on inclusion of potential DIMM-binding E-boxes in the immediate 5' upstream or 1st intronic regions. Two of these candidates displayed a statistically higher level of enrichment compared to controls – *CG1275* and *CG6522*. For seven other genes - *CG13248, CG11254 (mael), CG17293, CG7785, CG32850, CG14621* and *CG31436* - we observed a trend for DIMM enrichment at the E-box containing site (Figure 4; Table 1). However, these seven other examples did not achieve statistical significance.

In summary, ChIP analysis suggests that DIMM protein is normally resident in adult head DIMM cells *in vivo* in the regulatory DNA of at least three (*Phm*, *CG1275*, *CG6522*) and likely as many as ten, of the 11 candidate targets. Furthermore, through all three independent tests of validation, ten of 11 candidate target genes subsequently scored positive in at least one independent assay for direct DIMM regulation (Summarized in Table 1). Five genes were responsive in all three (*Phm*, *CG1275, CG13248*, *mael*, and *CG17239*) and two others*, CG7785 and CG6522,* were responsive in two of the assays. *CG32850, CG14621, CG31436* all showed

DIMM residency *in vivo* by ChIP, but of these three candidates, only *CG32850* also displayed mRNA induction by DIMM in embryos. Only the eleventh candidate, *pst,* failed to show significant responsiveness in any test.

Experiment IV. Adult in vivo RNAi and LC-MS. To determine if DIMM target genes contribute to its action mechanisms, we devised a functional screen to extend our analyses of gene regulation. We previously demonstrated that ectopic expression of DIMM in photoreceptor neurons confers peptidergic neuroendocrine properties onto them – photoreceptor neurons do not normally display such properties of the Regulated Secretory Pathway. When such cells are also forced to express a heterologous neuropeptide precursor (ppMII), the MII peptide is fully processed [15]. We therefore used the DIMM-dependent accumulation of processed MII peptide as an end-point assay to measure potential contributions of single candidate targets to the DIMM-generated secretory pathway. We used transgenic RNAi methods to knock down individual DIMM targets exclusively within photoreceptor cells and employed quantitative mass spectrometry to analyze the processing of MII. In the control condition (*GMR*> UAS-*dimm*, UAS-*ppMII*), fully-processed MII peptide is detected at 1710.69 m/z ratio [15]. To normalize the results of MII peptide accumulation across conditions, we also measured an endogenous *Drosophila* brain peptide, not found in photoreceptors, as an internal standard. For this, we chose the Drm-MT2 peptide derived from the HUGIN neuropeptide precursor (SVPFKPRLamid e, m/z 942.59: [21], because *hugin*-expressing neurons do not express the RNAi transgene in this experimental design. We predicted that levels of MT2 peptide should not change as a function of condition. An additional negative control in this design tested the effect of an RNAi transgene for a neuropeptide precursor never found in photoreceptors – *ecdysis triggering hormone* (*eth*); *eth* is normally expressed in the endocrine Inka glands, which are cells that normally express DIMM [8].

We used both labeling and label-free quantitative mass spectrometry approaches to

analyze MII and MT2 peptide levels in *Drosophila* head extracts using CapLC-MALDI-TOF/TOF MS. The MS-based labeling method is well validated [22-23] but involves multiple, sample-handling steps and is less effective for samples with low concentrations and small volumes. While we were able to quantify MT2 with the labeling approach (Figure S3), MII was not observed after labeling. We therefore turned to label-free quantitation because isolated peak heights can be directly compared across conditions for the following reasons: the instrument parameters are kept constant, the chemical environment of the peptide of interest is similar, and the total ion acquisition is equivalent for each measurement.

Using the label-free quantitation, MT2 levels showed equal intensities in the control (*GMR>dimm;>ppMII*) and experimental (*GMR>dimm; >ppMII*; *Phm-RNAi*) samples (Figure 5A), consistent with labeling results (Figure S3) and supporting its use as an internal standard. MII levels in the experimental sample (GMR>*dimm, >ppMII*; > *Phm-RNAi*) were significantly decreased from those in the control sample (GMR>*dimm*; >*ppMII*) (Figure 5B). After normalization to MT2, relative MII levels were quantified and we then determined if its production was significantly affected by the RNAi mediated knockdown of ten of the 11 candidate targets. As shown in Figure 5C, MII levels were significantly lower following RNA interference knockdown of *Phm*, *CG1275* and *CG13248* (Figure 5C). *CG17293* RNAi displayed less fully-processed peptide and *CG11254* (*mael*) RNAi showed a trend of up-regulation, but their final values did not exceed statistical significance. MII levels in the other six RNAi tests were very similar to control values.

In summary, three of ten candidate DIMM targets (*Phm*, *CG1275* and *CG13248*) make critical contributions to DIMM mechanisms, as tested *in vivo* under conditions in which the peptide secretory pathway is completely organized by DIMM. Two candidates (*Phm* and *CG1275*) are clearly involved in processing, while *CG13248* may as well be involved in overall organization of the secretory pathway (see Discussion). The functional assay we employed was

sensitive to disturbances of processing, but that is not the only interpretation for a reduction in the level of the secretory peptide. A reduced level could also result from an inability to accumulate, properly traffic or retain secretory peptides. We did not detect a build up of intermediates with *CG13248* RNAi and this likely indicates a defect at a stage different from processing.

The CG13248 protein is enriched in DIMM-neurons and dependent on dimm.

CG13248 encodes a putative cationic amino acid transporter, orthologous to mammalian cationic amino acid transporter 4 (CAT-4; SLC7a4), and closely related to *Drosophila slimfast* (CAT-1). CG13248 protein has not previously been associated with NE cells and so to confirm this prediction, we analyzed its normal expression (anti-CAT-4). We detected immunoreactivity (IR) for CAT-4 in both the adult brain in the 3rd instar larval brain and evaluated CAT-4-like IR in the context of DIMM expression.

First, we performed double immunostaining with *dimm* neurons labeled by anti-GFP (in *c929*-GAL4>*GFP* brains). In both developmental stages, CAT-4-like IR is heavily enriched in DIMM-positive neurons; Figure 6A shows an example of an adult brain. We also noted minorities of DIMM-only and CAT-4-only stained neurons. CAT-4-like IR was found mainly in cell bodies and terminals (not shown), and very little within axonal tracts. CAT-4-like IR was specific in that it was greatly reduced by the action of a *CG13248* RNAi transgenic construct: when driven by *386y*-GAL4 CAT-4-like IR was lost, while DIMM-like-IR was unaffected (Figure 6C and C'). *386y*-GAL4 is an insertion in the prohomone convertase gene *dPC2* and its drives expression in most or all of *dimm* neurons, as well as in other cells [24].

We next asked whether CAT-4 expression is dependent on DIMM in loss-of-function and gain-of-function *dimm* states. We generated UAS-*dimm* RNAi flies, confirmed the ability of these transgenes to produce a large-scale reduction of DIMM-like IR (Figure 6D and D') and observed a concomitant reduction of CAT-4-like IR (Figure 6D-E). This result indicates that high level

CAT-4 expression in *dimm*-positive neurons depends on DIMM. To test the effects of DIMM over-expression, we turned to the four-cell Tv cluster of the larval CNS. CAT-4-like IR is normally found in the two of the four Tv cluster neurons – the peptidergic Tv and Tvb, but not in the Tva or Tvc neurons (Figure 6E, cf., [25]). We used an *ap*-GAL4 driver to misexpress DIMM in all four Tv cells and observed ectopic CAT-4-like expression within the Tva and Tvc cells as well (Figure 6F). These results confirm that *in vivo* CAT-4 is specifically enriched in DIMM-positive cells and that it is regulated by *dimm*. The distribution of CAT-4-like immunoreactivity within DIMM neurons was studied in various identified peptidergic neurons of the adult brain (Figure S4), including in diverse neurons of the Pars Intercerebralis, HUG-positive neurons of the sub-esophageal neuromeres, and PDF-positive large LNv. In these neurons, CAT-4-like IR was strongly expressed by many DIMM neurons and weakly by others. It appeared principally cytoplasmic, and displayed heterogeneous accumulations.

A computational analysis afforded by prior transcript profiling of mature DIMM neurons. A recent microarray analysis of identified neurons from the adult *Drosophila* brain provided a fortuitous means to independently assess the authenticity of the original 134-gene list [26]. Importantly it was conducted on wild type brains containing normal DIMM levels, and so it serves as a useful counterpoint to our study of DIMM over-expression. Kula-Eversole *et al.* [26] profiled three types of neurons – the large lateral neuron ventral (l-LNv) and the small lateral neuron ventral (s-LNv). These two identified neuron groups are similar in that both are circadian pacemakers and both are neuropeptide PDF-expressing cells [27]. However they are different in that only the large LNv are DIMM-positive, while the small-LNv are not ([24]; Figure 7, and see [13]). Kula-Eversole *et al.* [26] compared l-LNv and s-LNv with a generic (ELAV-positive) brain neuron type for ~19,000 transcripts. Using that primary data set, we identified 579 genes enriched in large over small LNv. By comparing our 134-gene list derived from DIMM over-expression in embryonic stages to those ~579 transcripts normally enriched in DIMM-

positive cells from adult stages, we find an intersection of 18 putative DIMM direct target genes specifically enriched in adult l-LNv versus s-LNv (Figure 7; Table S3). Significantly, the six DIMM targets we identified by experimental analysis are all included in this intersection – *Phm*, *CAT-4*, *mael*, *CG17293*, *CG7785*, and *Cyt-b*$_{561\text{-}1}$.

DISCUSSION

The experiments reported here address the mechanisms underlying DIMM's regulatory functions within peptidergic neuroendocrine cells in *Drosophila*. The results from a genome-wide screening revealed a diverse array of potential DIMM targets and illustrated that the scope of DIMM actions is likely broad. The actions of its direct targets appear to extend from the nucleus (*CG17293*) to regulation of mRNAs (*CG11254*) to the ER and Golgi (*CG13248*) to peptide-containing LDCVs (*Phm* and *CG1254*). We found no neuropeptide–encoding genes on any of our lists, even the larger 134-gene list of transcripts exhibiting up-regulation with DIMM over-expression. We showed previously that DIMM is very inefficient by itself at driving ectopic neuropeptide gene expression [8-10]. Together these findings are consistent with our previous speculation that in *Drosophila*, specific neuropeptide expression is controlled by differing sets of transcription factors working within complex combinatorial codes [9, 13]. In contrast, DIMM provides parallel instructions for the cell biological machinery within which neuropeptides can be made, stored and trafficked [9, 13, 15].

Because we used an over-expression screen to generate a primary list of candidate targets, it was important to authenticate those results by reference to genes enriched in "normal" DIMM cells (i.e., cells in which DIMM levels were not artificially manipulated). We were fortunate to have access such information from the recently published Gene array study of Kula-Eversole *et al.* [26], from which we found that 13% of the 134 gene candidates were in fact highly-enriched in DIMM-positive neurons (versus DIMM-negative peptidergic neurons). While several candidates performed well in many of these tests and exhibit properties of direct DIMM targets,

most did not score positive in all tests employed (only *Phm, CG1275, CG13248*, *CG11254* and *CG17293* did). The results emphasize the importance of employing multiple tests to fully evaluate and properly interpret lists of regulated transcripts. Of the 11 genes passing the first test, we then used diverse experimental criteria to divide them into sets of six direct targets, two likely-direct targets, and three indirect targets (Table 1). We emphasize that our categorization of direct targets is based on highly stringent criteria and here discuss the significance of the findings for neuroendocrine cell biology.

Phm, Cyt-b$_{561-1}$ and CG13248 are key and direct DIMM mediators. The inclusion of *Phm* and *Cyt-b$_{561-1}$* genes in the original list of 11 candidates increased our confidence in the list's authenticity because both play well-established roles in LDCVs [18]. Furthermore, we had previously demonstrated that *Phm* is a true transcriptional DIMM target both in heterologous cells and *in vivo* [12]. Likewise the subsequent strong performance of *Phm* and *Cyt-b$_{561-1}$* in all four downstream assays provided further support for the validity of the experimental design to identify authentic DIMM targets.

CG13248 is a direct DIMM target and encodes a putative arginine transporter, CAT-4. In addition to *Phm* and *Cyt-b$_{561-1}$*, these studies show that a third *bona fide* DIMM target gene, *CG13248* is critical to normal regulation of neuroendocrine cell properties. Notably, in results described by Kula-Eversole *et al.* [26], *Phm*, *Cyt-b$_{561-1}$* and *CG13248* all ranked near the top for absolute transcript abundance in DIMM-positive neurons. The identification of *CG13248* as an integral component of neuroendocrine physiology is a significant new finding, but its specific contribution is a mystery because its precise molecular functions are not known. It is the clear sequence orthologue to mammalian cationic amino acid transporter 4 (CAT-4) and is therefore a candidate member of the system y+ (Na+ and pH-independent) cationic amino acid-preferring transport activities [28].

CAT proteins form a branch of the solute carrier family 7 (SLC7) [29]. Murine CAT-1, -2 and -3 all display arginine transporter activity when heterologously expressed, but to date, CAT-4 does not [30]. Notably, the *Drosophila* orthologue of the CAT-1 protein is the transporter SLIMFAST, which mediates arginine transport in fat body, couples to the mTOR pathway and helps that tissue function as a nutrient sensor that restricts global growth through a humoral mechanism [31]. In murine pancreatic acinar cells (which are regulated by the DIMM orthologue MIST1, [32], CAT-4 is a membrane-associated protein of secretory granules [33]. Future pursuit of the exact mechanisms and pathways in which CAT-4 operates in DIMM-expressing neurons will help illuminate fundamental neuroendocrine cell physiology.

Additional direct DIMM targets. Regarding the other direct DIMM targets, we make special mention of a few for potential novel insights into mechanisms of neuroendocrine cell regulation. *CG11254* (*mael*): By transcript profiling, Kula-Eversole *et al.* [26] report that *mael* is highly enriched in the DIMM-positive l-LNv's. It was originally characterized in nuclei and perinuclear particles of germ cells. MAEL helps localize components of the microRNA pathway and contributes to cellular polarization [19]. *CG17293* encodes a protein highly related to mammalian WDR82, and *CG7785* encodes a protein highly related to CCLD6 – both of which suggest a connection of DIMM mechanisms to chromatin-modifying properties [34].

There were two genes we concluded likely to be directly targeted (Table 1: "Maybe Direct") - *CG6522* encodes a member of the Testin/ Prickle family of proteins. Notably, the Prickle-like protein RILP interacts with REST and acts as a nuclear translocation factor [35]. The significance of potential Prickle-REST interactions is that REST displays a suppressive effect on neurosecretory properties of PC12 cells [e.g., 36]. In addition, *CG32850* encodes a protein orthologous to Ring Finger protein 11, which is a membrane-associated E3 ligase that is expressed widely in brain [37]. Finally, in the larger list of 18 genes representing the intersection of the embryonic and adult DIMM-regulated transcripts (Figure 1), there was sizable

representation of genes encoding proteins previously implicated in regulated neuropeptide secretion (*Rph*) and probable elements of the secretory pathway (*PPADC1, RCN2* and *Rabx4*).

DIMM directs a core program for neuroendocrine cells. These results define principal elements of what we anticipate will be a core program for neuroendocrine cell organization. Among mammalian bHLH proteins, DIMM is most similar to MIST1 [32]. Mills and colleagues have identified several candidate MIST1 targets, including RAB3D [38]. We note that five of the six DIMM targets that responded to DIMM in the trans-activation assay contained E boxes in their first exon-intron regions. The importance of 1st intron E boxes was already established for the case of *Phm* [12] and is also true for MIST1 target genes so far identified [38]. Furthermore, studies of *Phm* and *CG13248* suggest they can define a consensus DIMM binding profile: they both contain three boxes within the first intron, two of which have the sequence CATATG, all of which contribute synergistically, and one of which appears to have the strongest contribution to DIMM transactivation. We predict many other DIMM targets will display a similar E box profile. Furthermore, how individual target gene products contribute to the DIMM program, and how many more genes are involved, are now pertinent questions that will require additional studies. We anticipate that further analysis of this core DIMM program will help explain the regulatory organization of neuroendocrine cells and their evolution in different phyla. Because DIMM protein persists for the life of neuroendocrine cells in *Drosophila,* this work may also inform studies of neuroendocrine cell physiology and plasticity.

Developmental Generation of Peptidergic Phenotypes. In the case of neurons that utilize fast conventional neurotransmitters, transcriptional regulatory systems typically exert direct control over genes that encode biosynthetic enzymes, as well as ones for key transporter proteins that retrieve and recycle transmitters back into the lumen of synaptic vesicles [39]. For example, PET-1 supports serotonergic differentiation and directly targets genes that encode the critical biosynthetic enzyme TBH-1 and the serotonin transporter SERT1 [6]. It is striking

therefore that our limited but highly validated list of DIMM targets similarly includes genes essential for neuropeptide biosynthesis (*Phm* and *Cyt-b*$_{561\text{-}1}$) as well as a transporter that is specifically expressed by neuroendocrine cells (CAT-4). We propose that there may exist an unexpected but essential parallelism in the developmental regulation of secretory systems for small transmitters and for small amidated peptides. This hypothesis can help design experiments to further illuminate the mechanisms that underlie the developmental generation of peptidergic phenotypes.

EXPERIMENTAL PROCEDURES

Fly stocks. Details of the fly lines used are provided in Supplemental Information.

Microarray analysis and quantitative real-time PCR. We combined UAS-*dimm-myc* (II) with *elav*-GAL4; for microarray analysis, these were compared to *elav*-GAL4. Crosses were maintained 18^{o}C to minimize lethality. Details of the RNA preparation [40] and qPCR analysis are provided in Supplemental Information. All primer sets are listed in Table S1. All microarray data are publically available (GEO accession # GSE31113).

Whole mount *in situ* hybridizations. We followed general methods previously described [41]; details are provided in Supplemental Information.

Transactivation. Assays followed the general methods previously described [12]; details are provided in Supplemental Information.

Immunocytochemistry and Imaging. We generated antibodies and used immunostaining methods as previously described [8]; details are provided in Supplemental Information.

Chromatin Immunoprecipitation (ChIP). ChIP was performed as described by Menet *et al.* [42]. w^{1118}; *c929*-GAL4;*tub*-GAL80^{ts} females were crossed to $y^{1}w^{1}$; UAS-*dimm-myc* (II) males at 18^{o}C. Further details are available in Supplemental Information.

Real-time quantitative PCR (qPCR) following ChIP. The PCR mixture contained Platinum Taq polymerase (Invitrogen) and optimized concentrations of Sybr-Green (Invitrogen). The sequences of primers used are listed in Table S2; further details are available in Supplemental Information.

Quantitative mass spectrometry. Fly heads of the control (*GMR*>UAS-*ppMII*; UAS-*dimm*) and experimental (*GMR*>UAS-*ppMII*; UAS-*dimm*; UAS-RNAi) transgenic lines were collected in frozen state. MS-based quantitation was based on prior quantitative measurement approaches [22-23, 43]. Further technical details are provided in Supplemental Information.

ACKNOWLEDGMENTS

We thank our laboratory members for helpful discussions and Jason Mills for his comments on an earlier draft of this manuscript. We thank Jennifer Trigg and Weihua Li for excellent technical assistance. We also thank the Bloomington Stock Center, the DGRC and the VDRC for flies, and the *Drosophila* Genome Center for providing information. This work was supported by NIH P30-DA018310 and a NIDA grant NS031609 (to JVS), by a grant from the Dana Foundation and by NIE grant EY016807 (to SP), by NINDS grant NS036570 and NSF IOS-0744261 (to JBS), by NIH P01 NS044232 and P30-NS045713 and the Howard Hughes Medical Institute (to MR), and NINDS grant NS21749 (to PHT), and by a P30-NS057105 to Washington University. Imaging was performed at the Washington University Bakewell Center. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

REFERENCES

1. Dannies, P.S. (1999). Protein hormone storage in secretory granules: mechanisms for concentration and sorting. Endocr. Rev. *20*, 3-21.
2. Kim T., Gondré-Lewis M.C., Arnaoutova I. and Loh Y.P. (2006) Dense-core secretory granule biogenesis. Physiol. *21*, 124-133.
3. Burbach, J.P., Luckman, S.M., Murphy, D. and Gainer, H. (2001). Gene regulation in the magnocellular hypothalamo-neurohypophysial system. Physiol. Rev. *81*, 1197-1267.
4. Yue, C., Mutsuga, N., Verbalis, J. and Gainer, H. (2006). Microarray analysis of gene expression in the supraoptic nucleus of normoosmotic and hypoosmotic rats. Cell. Mol. Neurobiol. *26*, 959-978.
5. Hindmarch, C., Yao, S., Beighton, G., Paton, J. and Murphy, D. (2006). A comprehensive description of the transcriptome of the hypothalamoneurohypophyseal system in euhydrated and dehydrated rats. Proc. Nat'l. Acad. Sci. U S A *103*, 1609-1614.
6. Hendricks, T., Francis, N., Fyodorov, D. and Deneris, E.S. (1999). The ETS domain factor Pet-1 is an early and precise marker of central serotonin neurons and interacts with a conserved element in serotonergic genes. J. Neurosci. *19*, 10348-10356.
7. Flames, N. and Hobert, O. (2009). Gene regulatory logic of dopamine neuron differentiation. Nature *458*, 885-889.
8. Hewes, R.S., Park, D., Gauthier, S.A., Schaefer, A.M. and Taghert, P.H. (2003). The bHLH protein Dimmed controls neuroendocrine cell differentiation in *Drosophila*. Develop. *130*, 1771-1781.
9. Allan, D.W., Park, D., St. Pierre, S.E., Taghert, P.H., and Thor, S. (2005). Regulators acting in combinatorial codes also act independently in single differentiating neurons. Neuron *45*, 689-700.
10. Hewes, R.S., Gu, T., Brewster, J.A., Qu, C. and Zhao, T. (2006) Regulation of secretory protein expression in mature cells by DIMM, a basic helix-loop-helix neuroendocrine differentiation factor. J. Neurosci. *26*, 7860-7869.
11. Park, D., Veenstra, J.A., Park, J.H. and Taghert, P.H. (2008a). Mapping peptidergic cells in *Drosophila*: where DIMM fits in. PLoS ONE. *3*, e1896.
12. Park D., Shafer O.T., Shepherd S.P., Suh H., Trigg J.S. and Taghert P.H. (2008b). The *Drosophila* bHLH protein Dimmed directly activates *Phm*, a gene encoding a neuropeptide amidating enzyme. Mol. Cell. Biol. *28*, 410-421.
13. Park, D. and Taghert, P.H. (2009). Peptidergic neurosecretory cells in insects: Organization and control by the bHLH protein DIMMED. Gen. Comp. Endocrinol. *162*, 2-7.
14. Gumbiner, B. and Kelly, R.B. (1982). Two distinct intracellular pathways transport secretory and membrane glycoproteins to the surface of pituitary tumor cells. Cell *28*, 51-59.
15. Hamanaka, Y., Park, D., Yin, P., Annangudi, S.P., Edwards, T.N., Sweedler, J., Meinertzhagen, I.A., and Taghert, P.H. (2010). Transcriptional orchestration of the regulated secretory pathway in neurons by the bHLH protein DIMM. Curr. Biol. *20*, 9-18.
16. Verelst, W. and Asard, H. (2003) A phylogenetic study of cytochrome b561 proteins. Genome Biol. *4*, R38.
17. Eipper, B.A., Stoffers, D.A. and Mains, R.E. (1992). The biosynthesis of neuropeptides: peptide alpha-amidation. Ann. Rev. Neurosci. *15*, 57–85.
18. Perin M.S., Fried V.A., Slaughter C.A. and Südhof T.C. (1988). The structure of cytochrome b561, a secretory vesicle-specific electron transport protein. *EMBO J. 7*, 2697-2703.
19. Findley, S.D., Tamanaha, M., Clegg, N.J. and Ruohola-Baker, H. (2003). *Maelstrom*, a *Drosophila* spindle-class gene, encodes a protein that colocalizes with Vasa and RDE1/AGO1 homolog, Aubergine, in nuage. Develop. *130*, 859-871.

20. Ui, K., Nishihara, S., Sakuma, M., Togashi, S., Ueda, R., Miyata, Y. and Miyake, T. (1994) Newly established cell lines from *Drosophila* larval CNS express neural specific characteristics. In Vitro Cell Dev. Biol. Anim. *30*, 209–221.
21. Baggerman, G., Cerstiaens, A., Loof, A.D. and Schoofs, L. (2002). Peptidomics of the Larval *Drosophila melanogaster* Central Nervous System. J. Bio. Chem. *277*, 40368–40374.
22. Brockmann, A., Annangudi, S.P., Richmond, T.A., Ament , S.A., Xie, F., Southey, B.R., Rodriguez-Zas, S.R., Robinson, G.E. and Sweedler, J.V. (2009) Quantitative Peptidomics Reveal Brain Peptide Signatures of Behavior. Proc. Natl. Acad. Sci., U.S.A. *106*, 2383–2388.
23. Ramos-Ortolaza, D.L., Bushlin, I., Abul-Husn, N., Annangudi, S.P., Sweedler, J. and Devi, L.A. (2010). Quantitative neuroproteomics of the synapse. Methods Mol. Biol. *615*, 227-46.
24. Taghert, P.H., Hewes, R.S., Park, J.H., O'Brien, M.A., Han, M. and Peck, M.E. (2001). Multiple amidated neuropeptides are required for normal circadian locomotor rhythms in *Drosophila*. J. Neurosci. *21*, 6673-6686.
25. Park, D., Han, M., Kim, Y.C., Han, K.A. and Taghert, P.H. (2004). Ap-let neurons—a peptidergic circuit potentially controlling ecdysial behavior in *Drosophila*. Dev. Biol. *269*, 95-108.
26. Kula-Eversole E., Nagoshi E., Shang Y., Rodriguez J., Allada R. and Rosbash M. (2010) Surprising gene expression patterns within and between PDF-containing circadian neurons in *Drosophila*. Proc. Natl. Acad. Sci. U S A. *107*, 13497-13502.
27. Helfrich-Förster, C. (1995). The *period* clock gene is expressed in central nervous system neurons which also produce a neuropeptide that reveals the projections of circadian pacemaker cells within the brain of *Drosophila melanogaster*. Proc. Nat'l. Acad. Sci. U S A. *92*, 612-616.
28. White, M.F. and Christensen, H.N. (1982). The two-way flux of cationic amino acids across the plasma membrane of mammalian cells is largely explained by a single transport system. J. Biol. Chem. *257*, 10069–10080.
29. Closs, E.I., Boissel, J.P., Habermeier, A. and Rotmann, A. (2006) Structure and function of cationic amino acid transporters (CATs). J. Membr. Biol. *213*, 67-77.
30. Wolf, S., Janzen, A., Vékony, N., Martiné, U., Strand, D. and Closs, E.I. (2002). Expression of solute carrier 7A4 (SLC7A4) in the plasma membrane is not sufficient to mediate amino acid transport activity. Biochem. J. *364*, 767-775.
31. Colombani, J., Raisin, S., Pantalacci, S., Radimerski, T., Montagne, J. and Léopold, P. (2003). A nutrient sensor mechanism controls *Drosophila* growth. Cell *114*, 739-749.
32. Pin C.L., Bonvissuto A.C. and Konieczny S.F. (2000). Mist1 expression is a common link among serous exocrine cells exhibiting regulated exocytosis. Anat. Rec. *259*, 157-167.
33. Rindler, M.J., Xu, C.F., Gumper, I., Smith, N.N. and Neubert T.A. (2007). Proteomic analysis of pancreatic zymogen granules: identification of new granule proteins. J. Proteome Res. *6*, 2978-92.
34. Lee J.H. and Skalnik D.G. (2008) Wdr82 is a C-terminal domain-binding protein that recruits the Setd1A Histone H3-Lys4 methyltransferase complex to transcription start sites of transcribed human genes. Mol. Cell Biol. *28*, 609-618.
35. Shimojo, M. and Hersh, L.B. (2003). REST/NRSF-interacting LIM domain protein, a putative nuclear translocation receptor. Mol. Cell Biol. *23*, 9025-9031.
36. D'Alessandro, R., Klajn, A., Stucchi, L., Podini, P., Malosio, M.L. and Meldolesi, J. (2008). Expression of the neurosecretory process in pc12 cells is governed by rest. J. Neurochem., *105*, 1369–1383.

37. Anderson, L.R., Betarbet, R., Gearing, M., Gulcher, J., Hicks, A.A., Stefánsson, K., Lah, J.J., and Levey, A.I. (2007). PARK10 candidate RNF11 is expressed by vulnerable neurons and localizes to Lewy bodies in Parkinson disease brain. J. Neuropathol. Exp. Neurol. *66*, 955-964
38. Tian, X., Jin, R.U., Bredemeyer, A.J., Oates, E.J., Błazewska, K.M., McKenna, C.E. and Mills, J.C. (2010). RAB26 and RAB3D are direct transcriptional targets of MIST1 that regulate exocrine granule maturation. Mol. Cell Biol. *30*, 1269-1284.
39. Amara, S.J. and Kuhar, M.J. (1993). Neurotransmitter transporters: recent progress. Annu. Rev. Neurosci *16*, 73–93.
40. Lin Y., Han M., Shimada B., Wang L., Gibler T.M., Amarakone A., Awad T.A., Stormo G.D., Van Gelder R.N. and Taghert P.H. (2002). Influence of the *period*-dependent circadian clock on diurnal, circadian, and aperiodic gene expression in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U S A. *99*, 9562-9567.
41. Kosman D., Ip Y.T., Levine M. and Arora K. (1991). Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. Science *254*, 118-122.
42. Menet J.S., Abruzzi K.C., Desrochers J., Rodriguez J. and Rosbash M. (2010) Dynamic PER repression mechanisms in the *Drosophila* circadian clock: from on-DNA to off-DNA. Genes and Dev*. 24*, 358-367.
43. Hattan, S.J. and Parker, K.C. (2006). Methodology utilizing MS signal intensity and LC retention time for quantitative analysis and precursor ion selection in proteomic LC-MALDI analyses. Anal. Chem. *78*, 7986-7996.

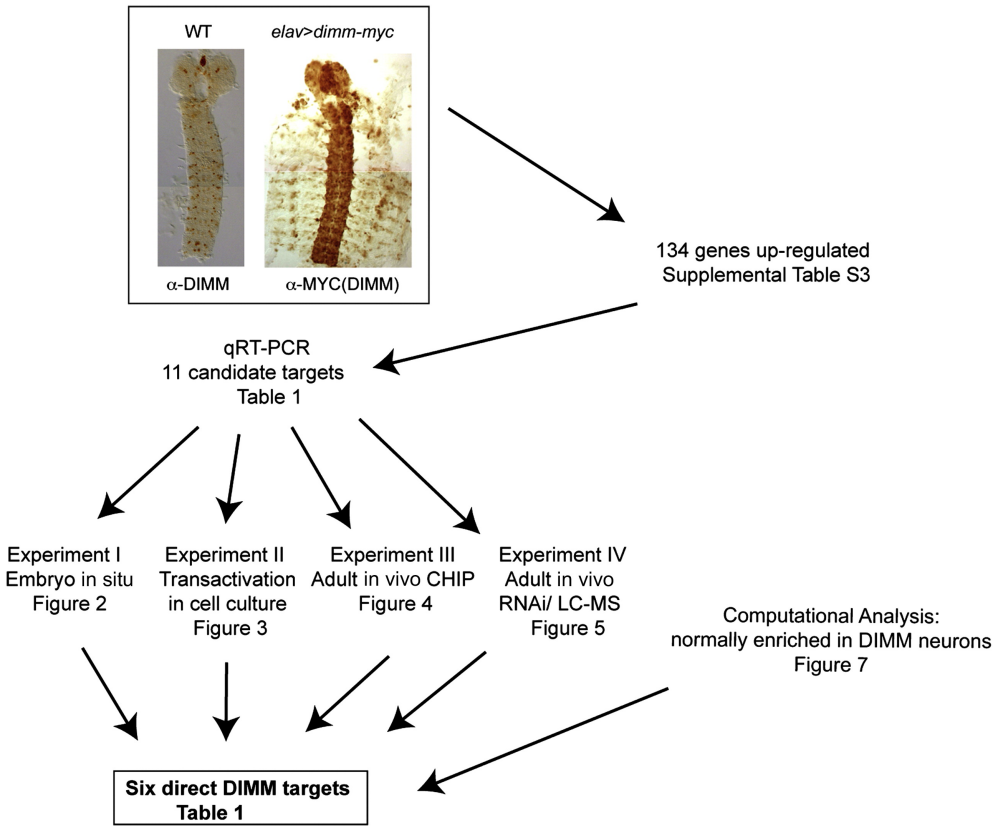

Figure 1. A schematic overview of the workplan for this study. A genome wide search for *dimm* targets performed by Genechip, then evaluated by three downstream analyses of DIMM regulation, and further evaluated by a functional *in vivo* assay.

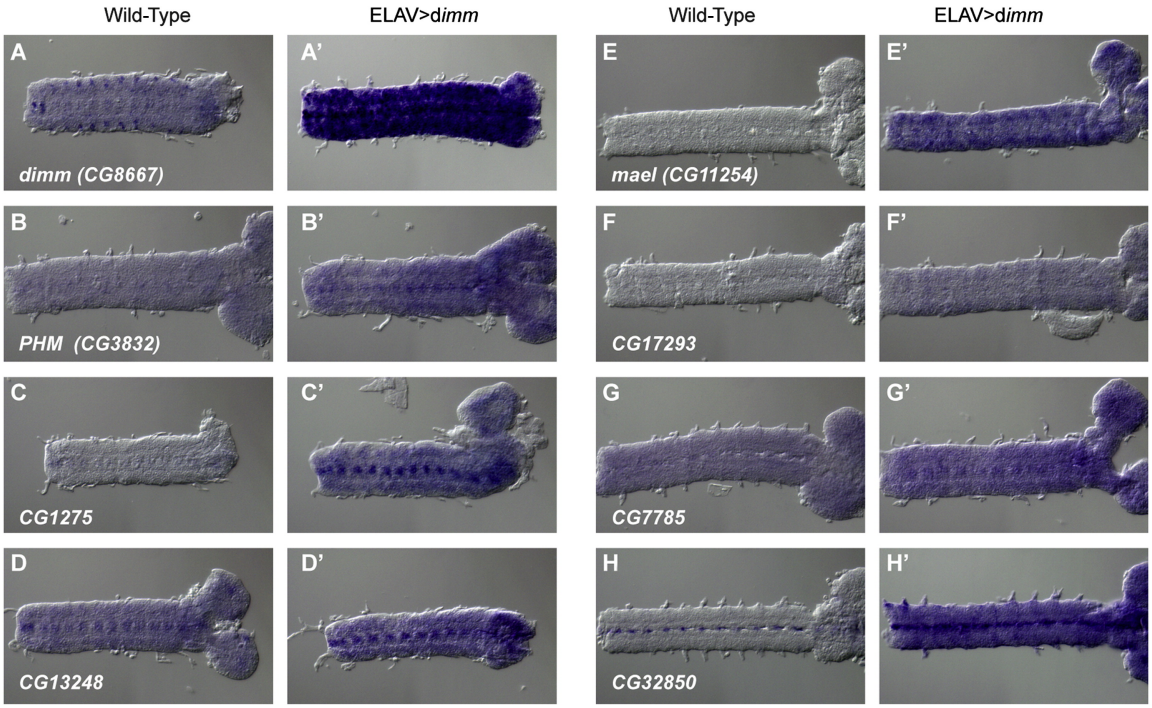

Figure 2. RNA *in situ* hybridization in control embryos and embryos that over-express *dimm*. Left - control (w^{1118}); Right - *elav> dimm*; Probes: A, B) *dimm*; C,D) *Phm*; E,F) *CG13248*; G,H) *CG11254* (*mael*); I,J) *CG17293*; K,L) *CG6522*. See also Figure S1.

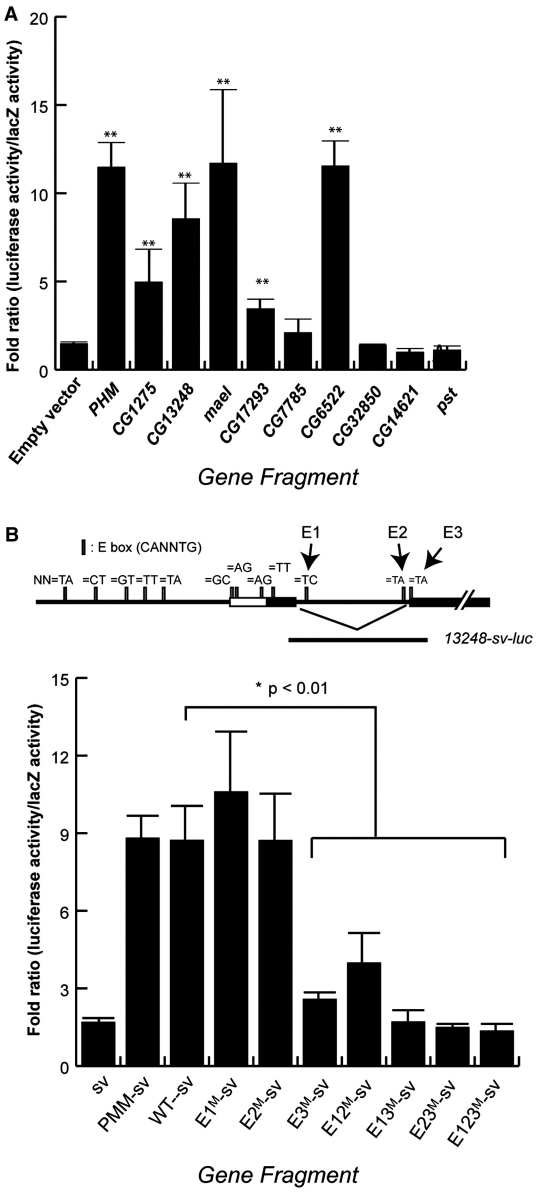

Figure 3. Trans-activation by DIMM of genomic fragments of candidate genes in *Drosophila BG3-c2* neuronal cell lines. Fold ratios represent Luciferase levels with *dimm* co-transfection divided by those without. Histograms represents means and SEMs of at least three independent replicate assays. (A). Results of testing ten DIMM targets. (B) Results from analysis of E box sequence requirements within the CG13248 regulatory region. E1, E2 and E3 indicate three separate E boxes, which were mutated singly, doubly or in triple-format. In (A), * $p<0.05$; ** $p< 0.01$ vs empty vector, by student's T-test. In (B), * $p< 0.01$ vs *CG13248* WT

sequence, by student's T-test. See also Figure S3 which illustrates E box positions in and around these candidate targets.

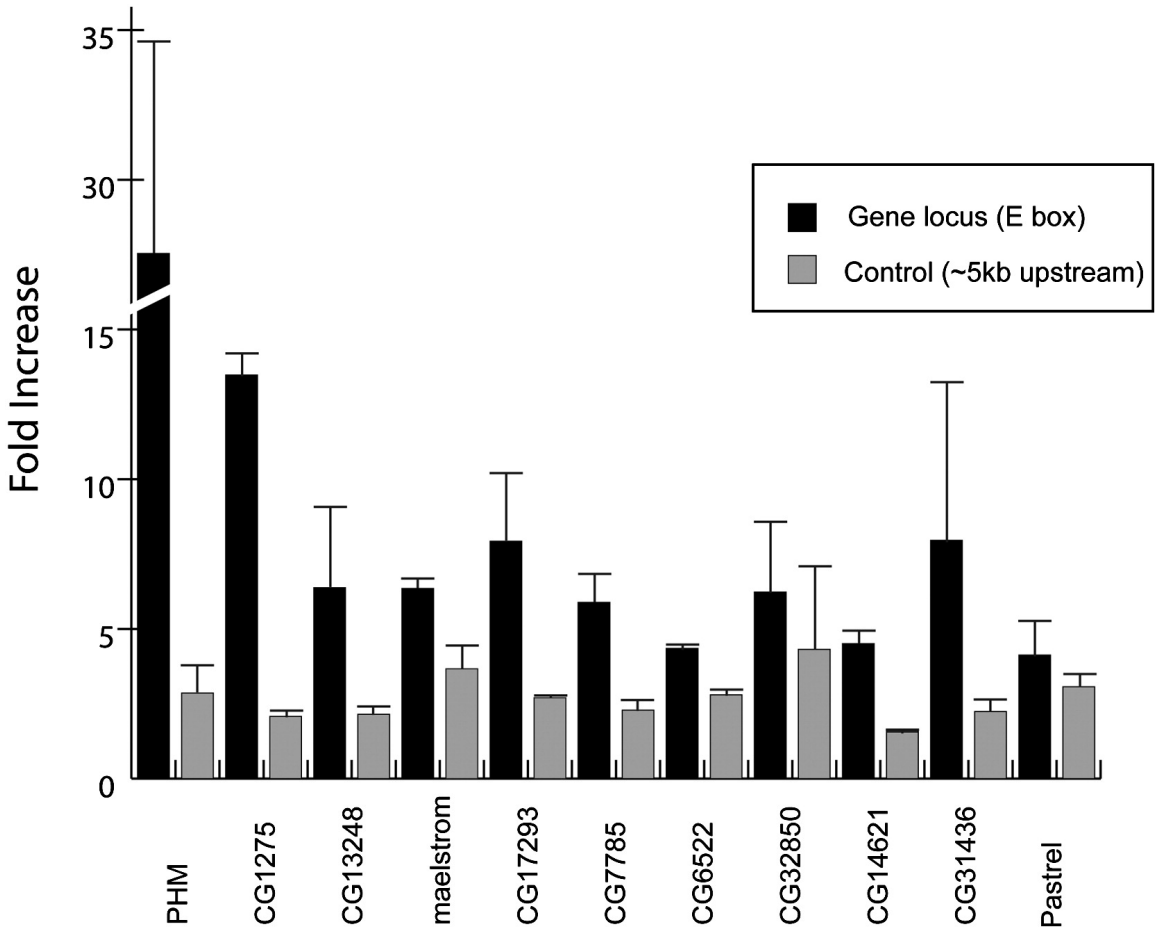

Figure 4. ChIP analysis *in vivo* of putative DIMM binding at E-boxes within *dimm*-dependent candidates genes. Darker histograms show the level of enrichment of the putative DIMM binding sites defined by the value in experimental versus control genotypes (see Methods). Lighter histograms show the level of enrichment of arbitrarily chosen sites ~6 kB upstream of the putative DIMM-binding sites, in the same experimental versus control genotypes. Histograms represents average and SEMs at least two independent assays (two biological replicates).

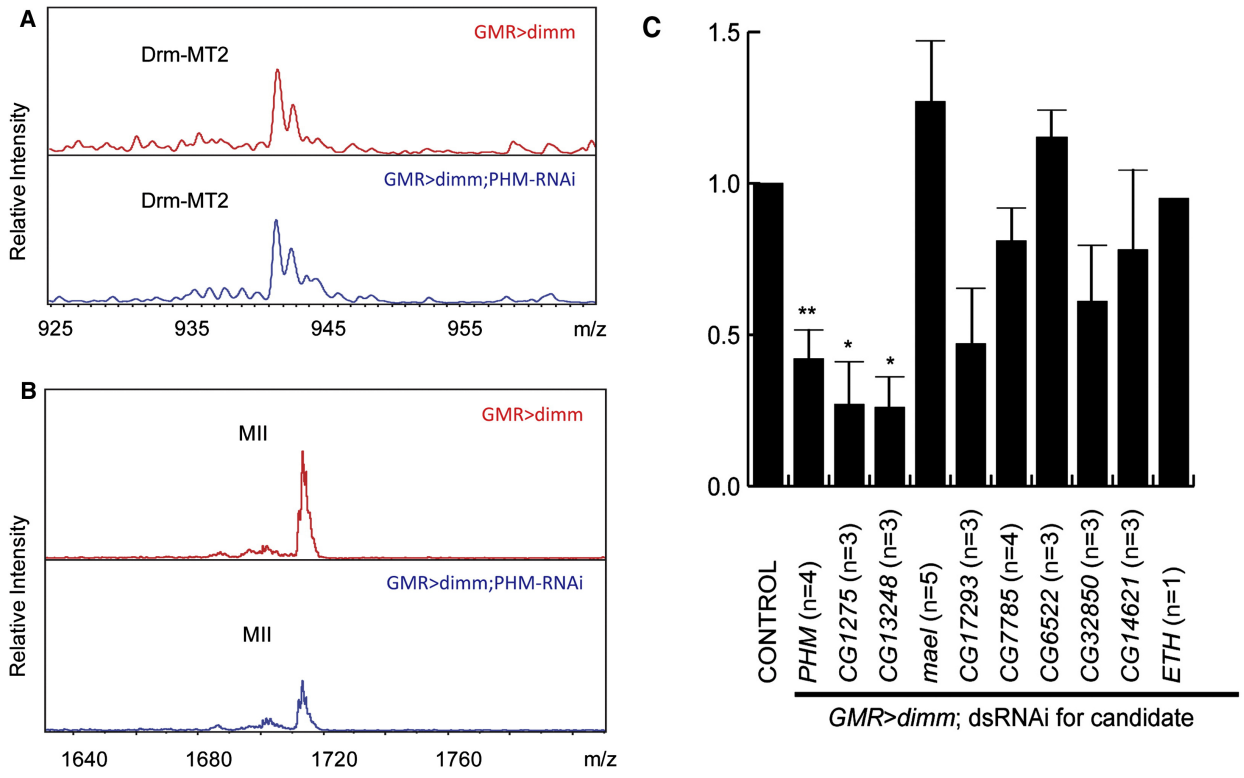

Figure 5. MS-based label-free quantitative analysis of alterations in MII peptide accumulation in photoreceptors following RNA interference of candidate DIMM targets.

(A) Mass spectra of MII in the control (*GMR*>UAS-*ppMII;* UAS-*dimm*, red) and experimental (*GMR*>UAS-*ppMII;* UAS-*dimm*; UAS-*PHM-RNAi*, blue) samples. The intensity ratio of MII in experimental versus control samples in these spectra is 0.54. B) Mass spectra of MT2 in the control and experimental samples. The intensity ratio of MT2 in experimental to control in these spectra is 1.02. C) Exogenous MII level in the experimental samples with RNAi compared to that in the control samples. Histograms represents means and SEMs, * $p<0.05$; vs control (*GMR*>UAS-*dimm*), by student's T-test. N, biological replicates. See also Figure S3 for analysis of MS-based quantitation of the endogenous peptide Drm-MT2.

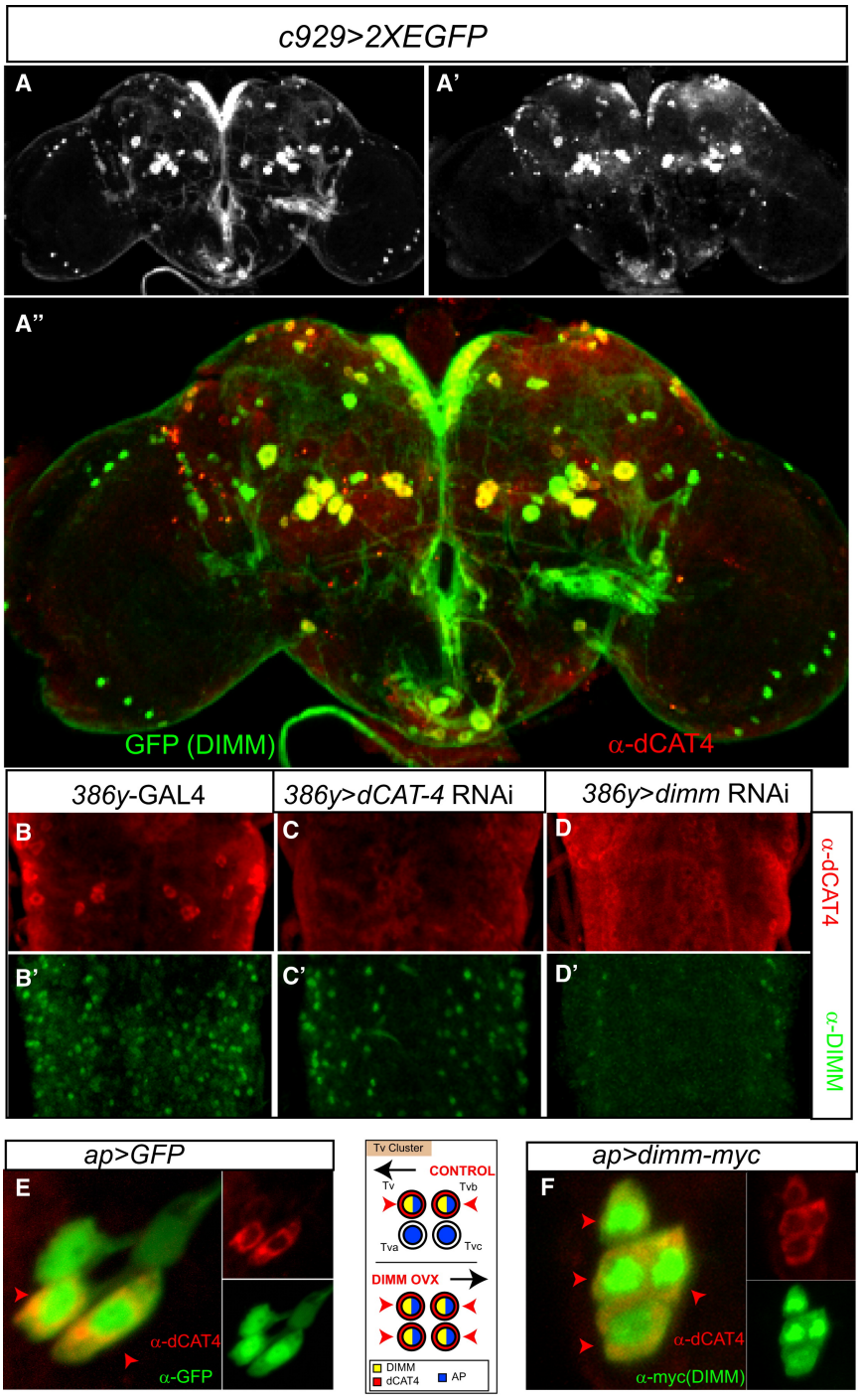

Figure 6. Enrichment of CG13248 (CAT-4) in DIMM neurons and its regulation by *dimm in vivo*.

A) DIMM-like and CAT-4-like IR are extensively co-localized in the adult brain. *c929*>UAS-*GFP* (anti-GFP, green), anti-CAT-4 (red); bottom: merged image. B-D) CAT-4-like IR following RNAi knock-down of *dimm*. B, B') parental control (*386Y*-GAL4); C, C') *386Y>UAS-DCR2/UAS-CAT-4-RNAi; D, D') 386Y>DCR2/ dimm-RNAi*; (B-D) anti-DIMM (green), (B'-D') anti-CAT-4 (red). (E-F) CAT-4-like IR is normally present in the two DIMM-positive neurons of the four-cell Tv-cluster; following DIMM mis-expression throughout the cluster, CAT-4 appears in all four cells. E) *ap*> *GFP*, anti-GFP (green), anti-CAT-4 (red), F) *ap*> *dimm*, anti-MYC (=DIMM) (green); anti-CAT-4 (red). See also Figure S4 for high power images of cDAT-4-like immunoreactivity in different adult brain DIMM neurons.

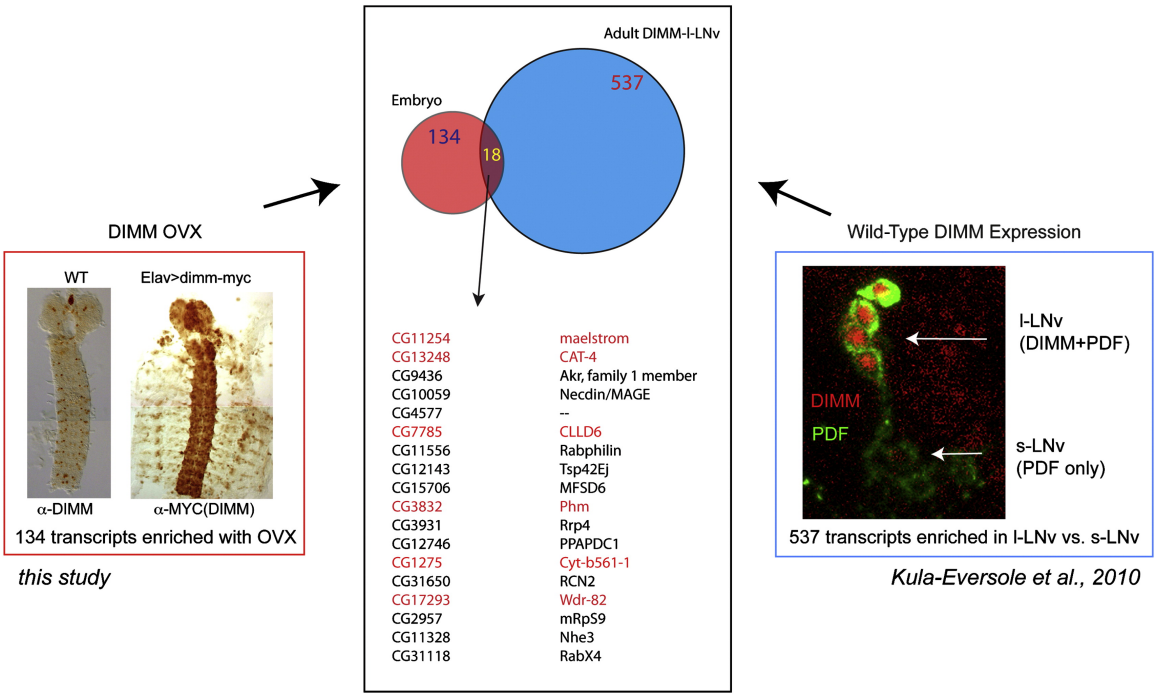

Figure 7. Comparison of transcripts upregulated by DIMM in embryos with transcripts enriched in DIMM-positive peptidergic neurons of the adult brain. Top: A Venn diagram illustrating the identification of 18 genes (13% of 134) of those up-regulated by DIMM over-expression among the 537 normally enriched in DIMM-positive large LNv [29]. Bottom lists the 18-gene intersection, asterisks mark those genes that were shown to be direct DIMM targets by current experiments.

Gene	Embryo In Situ[a]		Transactivation in Cell Culture	Adult In Vivo ChIP[b]	Adult In Vivo RNAi/MS	Enriched in Adult DIMM Neurons[c]	Conclusion
	Control	*Dimm*-OVX					
CG3832 (*Phm*)	+	++	++	++	+	+	direct
CG1275 (*Cyt* b_{561-1})	−	++	++	++	+	+	direct
CG13248 (*CAT-4*)	+	++	++	+	+	+	direct
CG11254 (*mael*)	−	++	++	+	−	+	direct
CG17293 (*wdr-82*)	+	++	++	+	−	+	direct
CG7785 (*CLLD6*)	+	++	−	+	−	+	direct
CG6522	+	+[d]	++	++	−	−	maybe direct
CG32850 (*RNF-11*)	+	++	−	+	−	−	maybe direct
CG14621 (*Slc35e1*)	−	−	−	+	−	−	indirect
CG31346	−	−	ND	+	ND	−	indirect
CG8588 (*pastrel*)	−	−	−	−	−	−	indirect

[a] −, no expression; +, moderate expression detected; ++, strong expression detected; ND, not determined.
[b] ++$p < 0.05$; +, positive but not surpassing statistical significance.
[c] Greatly enriched in DIMM-positive l-LNvs versus s-LNvs [26].
[d] No obvious change.

Table 1. Summary of Results to Determine the Validity of Eleven Candidate Direct DIMM Genes

SUPPLEMENTARY INFORMATION

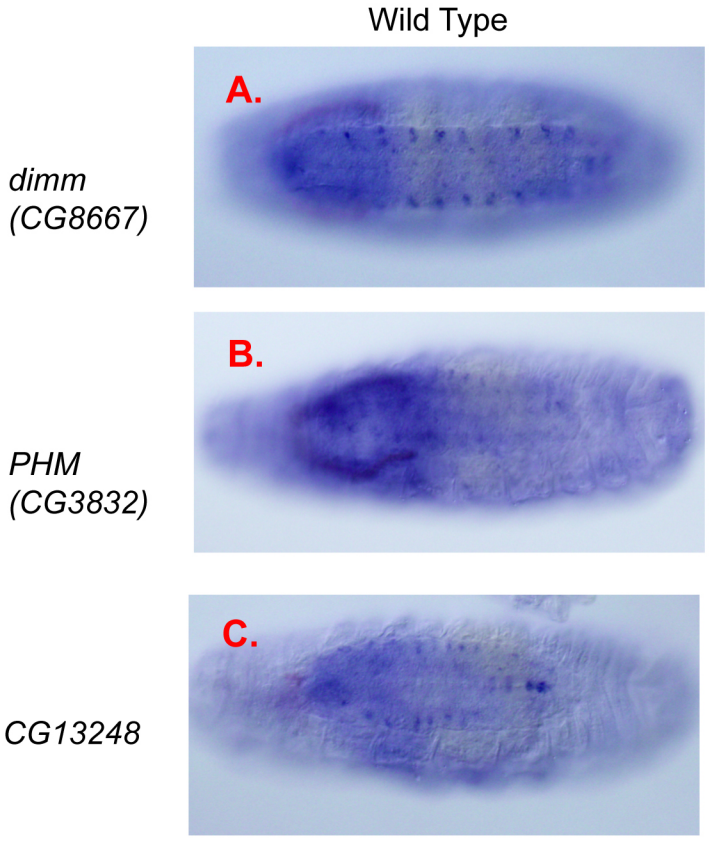

Figure S1. Similar spatial distribution of RNAs for *dimm*, *Phm* and *CG13248* in late stage control embryos. These data support the spatial analysis of DIMM target RNAs illustrated in Figure 2.

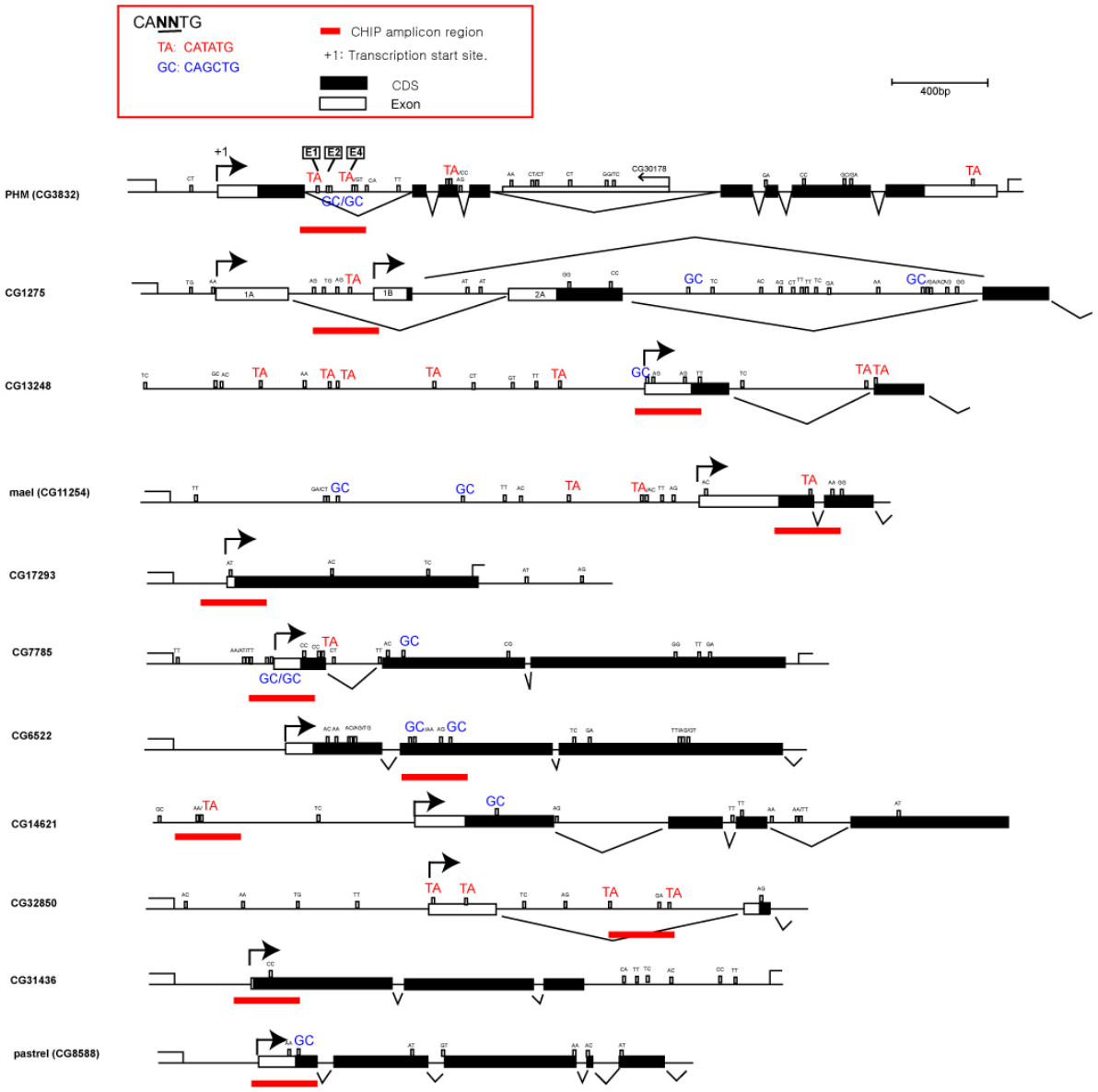

Figure S2. The distribution of CATATG and CAGCTG E-boxes within and around the candidate gene loci.

The schematic cartoon shows the putative DIMM binding sites within eleven candidate genes. Each box represents an E-box, CANNTG; Two nucleotides on the top of box indicate the sequence of NN. This information supports the study of DIMM trans-activation illustrated in Figure 3.

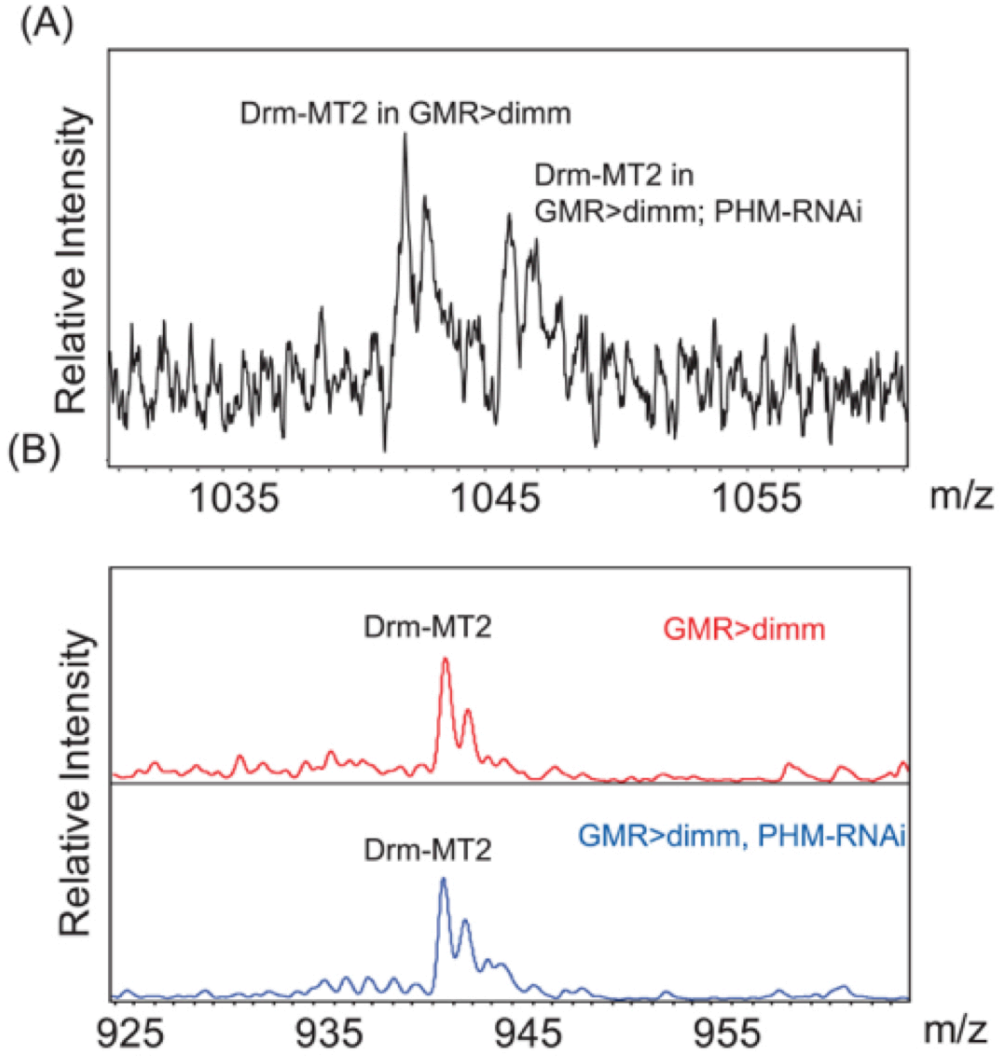

Figure S3. MS-based quantitation results from endogenous peptide Drm-MT2.

MALDI-TOF/TOF mass spectrum of labeled MT2 with light form of succinic anhydride in the control (***GMR*>UAS-*ppMII;* UAS-*dimm***) and with heavy form of succinic anhydride in the experimental (***GMR*>UAS-*ppMII*; UAS-*dimm*; UAS-*Phm-RNAi***) samples. The relative intensity

ratio of MII in experimental versus control genotypes is 0.95. This data supports the analysis illustrated in Figure 5.

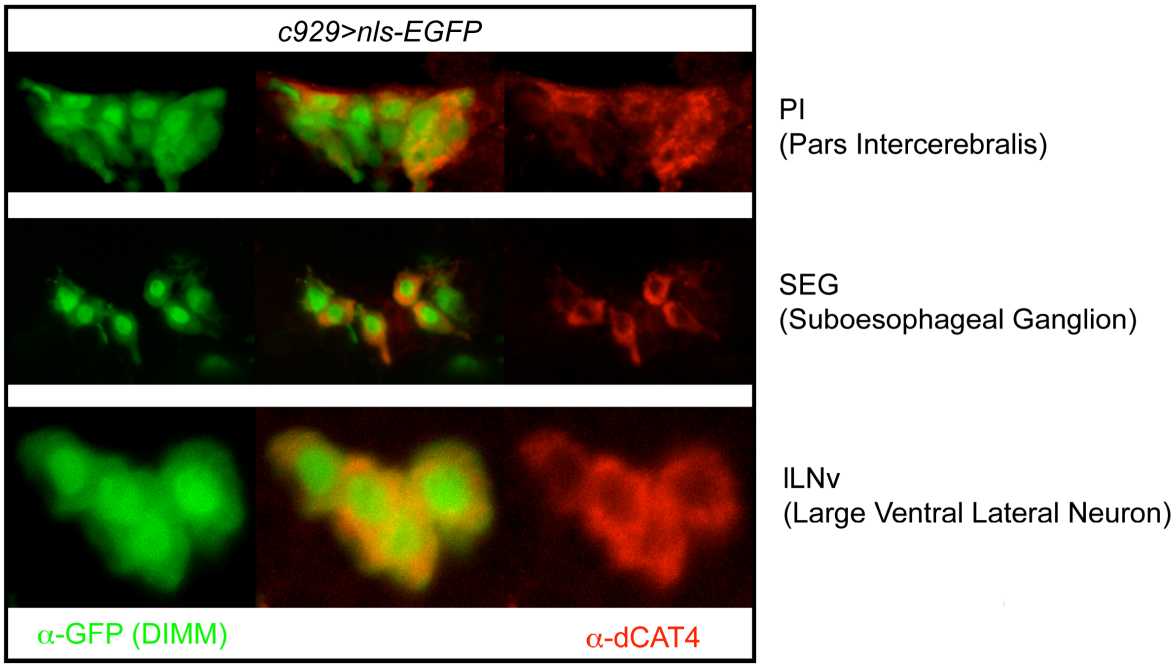

Figure S4. DIMM-positive adult brain neurons double stained for *c929*-GAL4 (green) and dCAT-4-like immunoreactivity (red). Note that the latter is primarily cytoplasmic and within individual cell bodies, appears heterogeneous (spotty). These images support the data presented in Figure 6.

Table S1. Information on primers used for analysis of cDNAs identified by genome-wide microarray screen of embryos following DIMM-over-expression. Please see attached / uploaded Microsoft Excel spreadsheet file.

Table S2. Information on primers used for ChIP analysis. Please see attached / uploaded Microsoft Excel spreadsheet file.

Table S3. Probes for 134 genes enriched > 1.5 fold change in DIMM overexpressing samples (ED34-E34) & (ED12-E12) over controls. Ordered from most to least enriched (average fold change, both developmental time points). The meaning of limma - calculated parameters is explained in the spreadsheet. Please see attached / uploaded Microsoft Excel spreadsheet file. Excel file also available from: http://www.sciencedirect.com/science/MiamiMultiMediaURL/1-s2.0-S0960982211008906/1-s2.0-S0960982211008906-mmc2.xls/272099/html/S0960982211008906/7502adb188d50fc94acf32597de56d44/mmc2.xls

Table S4. Probes for 23 genes down-regulated < 0.5 fold in DIMM over expressing embryos (ED34-E34) & (ED12-E12) over controls. Ordered from most to least down-regulated (average fold change, both developmental time points). The meaning of limma calculated parameters is explained in the spreadsheet. Please see attached / uploaded Microsoft Excel spreadsheet file. Excel file also available from: http://www.sciencedirect.com/science/MiamiMultiMediaURL/1-s2.0-S0960982211008906/1-s2.0-S0960982211008906-mmc2.xls/272099/html/S0960982211008906/7502adb188d50fc94acf32597de56d44/mmc2.xls

Table S5. qRT-PCR analysis of the 135 genes up-regulated by *dimm* over-expression in the embryonic CNS. 53 genes were tested and 18 showed greater than 2-fold levels following DIMM over-expression. The 11 genes selected for further study are marked in yellow. NT – not tested; ND – not detected. Please see attached / uploaded Microsoft Excel spreadsheet file.

Table S6. The eleven candidate DIMM-regulated genes – identification, mammalian orthologues and GO annotations. Please see attached / uploaded Microsoft Excel spreadsheet file.

Table S7. Probes for 535 genes enriched > 1.5 fold change and satisfying a 5% False Discovery Rate in large LNvs versus small LNvs. Ordered from most to least enriched (average fold change). The meaning of limma-calculated parameters is explained in the spreadsheet. Please see attached / uploaded Microsoft Excel spreadsheet file. Excel file also available from: http://www.sciencedirect.com/science/MiamiMultiMediaURL/1-s2.0-S0960982211008906/1-s2.0-S0960982211008906-mmc2.xls/272099/html/S0960982211008906/7502adb188d50fc94acf32597de56d44/mmc2.xls

Table S8. Probes for 537 genes enriched > 1.5 fold change and satisfying a 5% False Discovery Rate in small LNvs versus large LNvs. Ordered from most to least enriched (average fold change). The meaning of limma-calculated parameters is explained in the spreadsheet. Please see attached / uploaded Microsoft Excel spreadsheet file. Excel file also available from: http://www.sciencedirect.com/science/MiamiMultiMediaURL/1-s2.0-S0960982211008906/1-s2.0-S0960982211008906-mmc2.xls/272099/html/S0960982211008906/7502adb188d50fc94acf32597de56d44/mmc2.xls

SUPPLEMENTARY EXPERIMENTAL PROCEDURES

Fly stocks. The following fly lines were used: *dimm* deletion mutant alleles (*Rev4* and *Rev8*); UAS-*dimm*-myc (II or III) [1]; UAS-*ppMII* [2], UAS-*2X EGFP* and the pan-neuronal driver, *elav*-GAL4 (III), *GMR*-GAL4 (II). For quantitative mass spectrometry, we used UAS-RNAi transgenic lines obtained from the Vienna RNAi stock center (VDRC) to knockdown target genes (See Supplemental Information), combined with UAS-*DCR2* (obtained from Stefan Thor, Linkoping Univ., Sweden). The following strains were used for quantitative mass spectrometry:

w, UAS-*DCR2;pGMR*-GAL4 *(II);* UAS-*ppMII (III)*

w; UAS-*dimm-myc (II or III)*;UAS- *dsRNAi* lines (II or III).

The following strains were used for ChIP analysis**:**

w^{1118}*; c929*-GAL4/UAS-*dimm-myc (II);;tub*-GAL80^{ts}

w^{1118}*; c929*-GAL4*;tub*-GAL80^{ts} (as a negative control strain)

Microarrays. We combined UAS-*dimm-myc* (II) with *elav*-GAL4; for microarray analysis, these were compared to *elav*-GAL4. Crosses were maintained 18^{o}C to minimize lethality. RNA was prepared from embryos collected at 22-26 hr (~ embryo Stage 14) and at 28-32 hr (~ embryo Stage 16). A pair of GeneChip® *Drosophila* Genome 2.0 arrays (Affymetrix Co., Santa Clara, CA) were tested with each RNA sample (two experimental and two control, for each of two time points). Eight microarrays were used to hybridize cDNAs from experimental *dimm::myc*-expressing and control flies, sampled at two different embryonic stages. Four microarrays were hybridized with cDNA from the *dimm::myc*-expressing embryos of genotype w;UAS-*dimm-myc* (II)/+;*elav*-GAL4/+; (prefix "ED"). Another four microarrays were hybridized with cDNAs from the parental control strain: w;;*elav*-GAL4 (prefix “E”). Samples ED3 and ED4, E3 and E4 were hybridized with cDNA from stage 14 embryos (24-26hr development). Samples ED1 and ED2, E1 and E2 were hybridized with cDNA from stage 16 embryos (28-32 hr old embryos). A pair of GeneChip® Drosophila Genome 2.0 arrays (Affymetrix Co., Santa Clara,

CA) were tested with each RNA sample (two experimental and two control, for each of two time points). All microarray data are publically available (GEO accession #GSE31113).

Affymetrix CEL files raw probe intensity data was loaded into the statistical computing language R v. 2.12.1 for Snow Leopard (32-bit build) with the affy / simpleaffy libraries (ReadAffy). Microarrays were quality checked with the affyPLM and affycoretools libraries. Probe intensities were then $\log_2$ normalized with the gcrma algorithm version 2.22.0 (default settings). All eight arrays were gcrma-normalized together as one data set. Gcrma-normalized data sets were then processed with the limma package (Linear Models for Microarray Data) v. 3.6.9 and default limma settings [3]. Contrast matrices were set up to look for probes that were differentially expressed in the ED samples compared to E samples for each age: (ED3,ED4)-(E3,E4) and (ED1,ED2)-(E1,E2).

Limma comparison results are displayed in terms of linear scale fold changes when ED probes are compared to E probes. The ED34-E34 comparison yielded 360 probes that were at least 1.5 fold up-regulated. There were 48 probes that were below 0.5 fold in the ED34-E34 comparison, indicating down-regulation. The ED12-E12 comparison yielded 342 ED probes that were at least 1.5 fold above E probes and 97 ED probes below 0.5 fold of E probes.

In order to correct for multiple comparisons, Benjamini-Hochberg False Discovery Rate (FDR) corrections implemented in limma were used. There were 880 and 912 differentially expressed genes with conventional p-values < 0.05 in the ED34-E34 and ED12-E12 comparisons, respectively. Of these, 21 / 880 and 44 / 912 satisfied a 5% FDR cutoff. Failure to obtain a higher number of significant hits after Benjamini-Hochberg correction at 5% FDR was likely related to the relatively small number of replicates (two) in the experiment. Furthermore, sampling whole embryos as opposed to only *elav*-positive cells yielded a low signal to noise ratio.

We therefore decided to use the 1.5 fold enrichment cutoff in identifying differentially expressed genes. We first sought to identify overlapping probes that were above 1.5 fold enriched in both the early and late embryos. To obtain the intersection of these probe sets, queries were run in Galaxy, an open platform for genomic research [4-5]. Galaxy queries were setup to find which of the 360 ED34-E34 enriched probes overlapped with the 342 ED12-E12 enriched probes. This intersection yielded a list of 136 probes (135 genes, Supplementary Table 3). We also identified overlap between the 48 probes below 0.5-fold in ED34-E34 and the 98 probes in ED12-E12. This intersection included 22 probes that were downregulated in both early and late embryos at 0.5-fold level or less (Supplementary Table 4).

We compared the list of DIMM overexpression-enriched genes generated in this study with the data generated by Kula-Eversole *et al.* [6] obtained from purified “normal” DIMM neurons (adult LNv). In order to directly compare our study with theirs, we reanalyzed the raw data from [6], with the same methods used for the embryo data. Additionally, Kula-Eversole *et al*. compared large LNv and small LNv probe data to *elav* cells sat different circadian time points. We wanted to compare large LNv neurons directly to small LNv neurons regardless of circadian time. For these reasons, data from Kula-Eversole *et al.* were reanalyzed as follows: all CEL files from the study were obtained from the Gene Expression Omnibus depository under the accession number GSE22308. The 24 CELs were gcrma-normalized together followed by limma analysis of probes enriched in Large-Small LNv neurons.

We analyzed all Large LNv (10 samples) and small LNv (4 samples) files regardless of the circadian time when RNAs from purified LNv neurons were harvested. For this analysis, we assumed that DIMM is not involved directly in the circadian system and that therefore those genes that DIMM activates directly should be enriched in Large LNv neurons across circadian time points. Limma was used to identify differentially expressed probes in the Large LNv-small LNv comparison at a Benjamini-Hochberg FDR of 5% and a 1.5 fold minimum fold enrichment.

535 probes were 1.5 fold enriched in Large LNvs over small LNvs with a 5% FDR (Supplementary Table 7). 537 Large LNv probes were below 0.5 fold of small LNv with a 5% FDR (Supplementary Table 8).

Next, the goal was to identify overlapping probe sets that are enriched upon embryonic DIMM overexpression and in normal adult DIMM-positive Large LNv neurons when compared to DIMM-negative small LNvs. Galaxy queries were run to find the intersection of 136 enriched probes from the embryo and the 535 Large LNv-enriched probes from the adult. The intersection of these data sets yielded a set of 18 genes (Figure 7; Supplementary Table 3). There were no overlapping hits when 22 probes that were <0.5 fold in both embryonic samples (Supplementary Table 4) were compared to probes that are <0.5 fold and 5% FDR in the Large-small comparison (Supplementary Table 8). The term DIMM-negative is used as a descriptor and is not meant to suggest that the small LNv have no DIMM expression.

qPCR analyses. To evaluate candidate genes, quantitative real time-PCR (qPCR) was performed using RNA derived from 24-32hr embryos collected at 18^{o}C and mis-expressing either DIMM (UAS-*dimm*) or CD8-EGFP (UAS-*CD8-EGFP*) driven by *elav*-GAL4. To analyze *dimm* mutants, first instar larvae trans-heterozygous for two *dimm* alleles (*rev4/rev8* – [1]), or first instar larvae from the control w^{1118}, were collected at RT. To analyze neural RNAs, one hundred third instar larval CNS of w^{1118} were manually dissected. Total RNA was isolated with Trizol reagent (Sigma, St Louis, MO), digested with RNase-free DNase I, and purified with RNAeasy columns (Qiagen, Madison, WI). Reverse transcription reactions were performed following the manufacturer's protocols (NEB). We measured transcript quantities using SYBR-green incorporation on an ABI 7000 machine and made genotypic comparisons with the double delta Ct method [7]. Levels of *RP49* RNA were used for normalization control. The primers used are listed in Supplementary Table 1.

***In situ* hybridizations and antibody staining of embryos.** We followed general methods previously described [8]. Clones containing the cDNA of interest were obtained from the PCR of the cDNA template. For probe synthesis, plasmids were linearized with appropriate restriction enzymes at the 5' end of the coding sequence. Transcription reactions used the digU RNA labeling mix (Roche, final concentrations 1mM each ATP, CTP and GTP, 0.65 mM UTP, 0.35 mM DIG-11-UTP) and the appropriate RNA polymerase (NEB, Roche, or Promega) for 2 hrs at 37^{o}C. Probes were then hydrolyzed in carbonate buffer (final concentration 60 mM Na_2CO_3, 40mM $NaHCO_3$, pH 10.2) for 40 minutes at 65^{o}C. The reactions were neutralized with stop solution (final concentration 0.1 M sodium acetate pH 6.0), and the RNA is precipitated with LiCl and re-suspended in 150 ml hybridization solution (see below). Typically, between 2 and 4 ul of probe was used in a 100 ml hybridization reaction.

Embryos were collected and aged on yeasted grape plates. Due to embryonic lethality at room temperature, for the *dimm* overexpression studies all collections and aging was done at 18^{o}C. Embryos were dechorionated in 50% bleach and fixed in 50:50 heptane: embryo fix buffer (1xPBS, 50 mM EGTA pH 8.0, 10% formaldehyde) for 25 minutes with vigorous agitation, followed by devitellinization with MeOH. The embryos were cleared using Xylenes, postfixed in 5% formaldehyde in PBT (1xPBS + 0.1% Tween 20) and then treated with proteinase K (Roche, 4 ug/ml in PBT) for 10 minutes. This was followed by another postfix and prehybridization for 1 hr in hybridization solution (50% de-ionized formamide (American Bioanalytical), 5x SSC, 100 ug/ml sonicated, boiled salmon sperm DNA (Sigma), 50 ug/ml heparin (Sigma), 0.1% Tween 20 (Fisher)) at 55^{o}C. The embryos were then hybridized with antisense RNA probes in hybridization solution for 18 hrs at 55^{o}C. After washing with hybridization solution at 55^{o}C and PBT at room temperature, the embryos were incubated with alkaline phosphatase labeled anti-digoxygenin antibody (Fab fragments, Roche, pre-absorbed, diluted at 1:2,000) overnight at 4^{o}C. After washing with PBT, the staining was developed with 4-nitro blue tetrazolium (NBT, Roche,

0.675mg/ml) and X-phosphate (Roche, 0.35mg/ml) in AP staining buffer (100mM NaCl, 50mM MgCl2, 100mM Tris pH 9.5, 0.1% Tween 20). Embryos were mounted in permount (Fisher) or 70% glycerol in PBS and viewed using a Zeiss Axioplan2 microscope. Images were obtained with an Olympus DP71 camera and manufacturer's software.

Transactivation. Assays followed the general methods previously described [9]; Genomic fragments from the loci encoding candidate genes were subcloned into the *pGL3* vector (Promega, Madison WI). *Drosophila BG3-c2* neuronal cells (1x 10^6 cells per well) were transiently transfected with a total 1.5mg of DNA by mixing with 5 ul of FuGENE^{HD} (Roche) and incubating for two days at RT. Transactivation was measured by a Luciferase assay system (Promega, Madison, WI). For each experiment, a vector containing *pActin-LacZ* was co-transfected to normalize the transfection efficiency, and each transfection was performed at least three times independently. The significance was tested by the student's T-test (two tailed).

Immunocytochemistry and Imaging. We generated antibodies and used immunostaining methods as previously described [1, 10, 11]. Affinity purified guinea pig anti-DIMM (1:200; [10]), rabbit anti-FMRFa (1:1000; [11]), mouse monoclonal anti-GFP 3E6 (1:800, Molecular Probes, Carlsbad CA) and rabbit anti-GFP (1:500; rabbit polyclonal, #AB3080 Chemicon, Temecula, CA) were used as primary antibodies. Cy3-conjugated- (Jackson Immunoresearch, West Grove, PA) or Alex-488-conjugated- (Molecular Probes, Carlsbad CA) secondary antibodies were used for immunocytochemistry. To generate an antibody against the CG13248 protein product, we constructed a GST-fusion containing the predicted C-terminal (cytoplasmic) region of the protein (596-669 A.A.; 75 amino acids) and produced GST-CAT-4 in *E. coli*. Guinea pig anti-CAT-4 protein was used for immunostaining, using methods previously described [1, 10, 11]. Confocal images were acquired on an Olympus FV500 laser scanning confocal microscope and manipulated by ImageJ and Adobe Photoshop software to adjust contrast.

Chromatin Immunoprecipitation (ChIP) analysis. ChIP was performed as described by Menet *et al.* [12]. w^{1118}; *c929*-GAL4;*tub*-GAL80^{ts} females were crossed to y^1w^1; UAS-*dimm-myc* (II) males at 18^{o}C. To control for non-specific antibody binding, non-specific DNA enrichment (through binding to sepharose beads, reaction surfaces etc. [13]), ChIPs were also performed on the w^{1118}*; c929*-GAL4; *tub*-GAL80^{ts} parental strain that lacks the *dimm::myc* transgene. Progeny developed at 18^{o}C through adult eclosion. Two to five day old adults of the following two genotypes (y^1 w^1/ +*; c929*-GAL4/UAS-*dimm-myc* (II); *tub*-GAL80^{ts}/+ and w^{1118}*; c929*-GAL4*; tub*-GAL80^{ts}) were shifted to 30^{o}C for 3 days. Flies were collected on dry ice, and heads were sieved the following day through 25 and 40 micron sieves. One milliliter of fly heads was homogenized in 7 ml Wheaton-Dounce homogenizers with the type B (tight) pestle in 3 mL of NEB buffer (10 mM HEPES-Na at pH 8.0, 10 mM NaCl, 0.1 mM EGTA-Na at pH 8.0, 0.5 mM EDTA-Na at pH 8.0, 1 mM DTT, 0.5% Tergitol NP-10, 0.5 mM Spermidine, 0.15 mM Spermine plus protease inhibitor cocktail [Roche mini]) for a total of 30 min (2-min homogenization 10 times, 1 min on ice 10 times). Homogenate was dumped into a 70-mm cell strainer placed in a 50-mL falcon tube and centrifuged at 300g for 2 min. Filtered homogenate was split into three fractions, transferred to microfuge tubes and centrifuged at 6,000 x g for 10 min. The nuclei-containing pellets were resuspended in 1 mL of NEB and centrifuged at 20,000 x g for 20 min on sucrose gradient (0.65 mL of 1.6 M sucrose in NEB, 0.35 mL of 0.8 M sucrose in NEB). The pellet was resuspended in 1 mL of NEB and 11% formaldehyde (diluted in Schneider's media; Sigma) was added to a final concentration of 1%. Nuclei were cross-linked for 10 min at room temperature on a rotator before cross-linking was quenched by adding 1/10 vol of 1.375 M glycine for 5 min. The nuclei were collected by centrifugation at 6,000 x g for 5 min. All centrifugations steps were carried out at 4^{o}C. Nuclei were washed twice in 1 mL of NEB and resuspended in 450 mL of Sonication buffer (10 mM HEPES-Na at pH 7.5, 2 mM EDTA at pH 8.0, 1% SDS, 0.2% Trition X-100, 0.5 mM Spermidine, 0.15 mM Spermine plus Roche mini

protease inhibitor cocktail). Two biological replicates of nuclei were sonicated using a Fisherbrand Sonic Dismembranator at setting 2 (57 W) five times for 15 sec on and 2 min off (set on ice). Three other biological replicates were sonicated on a Misonix 3000 sonicator fitted with a microtip at setting P=5.0 (12-15W) 6 - 12 times for 15 sec on and 1 minute off (set in ice-ethanol bath). Sonicated nuclei were centrifuged at 15,000 x g for 10 min and frozen at -80°C in 150 µl aliquots. The majority of the sonicated chromatin had a unimodal sequence length distribution with a peak around 500 base pairs. Twenty-five microliters of sonicated chromatin were removed for the input sample. The remaining 125µl of chromatin was diluted in 1.1 mL of IP buffer (50 mM HEPES/KOH at pH 7.6, 2 mM EDTA, 1% Triton, 0.1% NaDeoxycholate in PBS). Samples were rotated overnight at 4°C after adding antibodies: 15 µl of anti-c-MYC tag goat antibody (Abcam ab9132). Protein G-Sepharose beads (Zymed) or protein G-sepharose beads (Sigma, Saint Louis, MO) were blocked overnight in 0.1 mg/mL yeast tRNA and 1 mg/mL BSA in IP buffer. After overnight incubation, the beads were washed once in IP buffer, added to the chromatin/antibody mixture, and then incubated for an additional 2 h at 4°C.

Beads were spun down at 10,000 rpm for 20 sec and were washed once in 1.5 mL of ChIP Wash buffer (50 mM HEPES-KOH at pH 7.6, 1 mM EDTA, 1% Triton, 0.1% NaDeoxycholate, 0.1% Sarkosyl, 0.1% BSA, 0.5 M KCl in PBS). Beads were resuspended in 1 mL of ChIP Wash buffer and rotated for 30 min at 4°C. Beads were then washed once in Li Wash Buffer (10 mM Tris-Cl at pH 8.0, 0.25 M LiCl, 0.5% NP40, 0.5% NaDeooxychoalte, 1 mM EDTA) and once in cold TE (pH 8.0) before being eluted with 150 mL of ChIP Elution buffer (50 mM Tris-HCl at pH 8.0, 10 mM EDTA, 1% SDS, 1 mM DTT, 0.1 mg/mL Proteinase K). ChIP Elution buffer (150 mL) was also added to the input sample. Both IP and input samples were incubated for 2 h at 37°C. The sepharose beads were removed from the IP samples and then all samples were decrosslinked overnight at 65°C. DNA was isolated from the samples using PCR purification kit (Qiagen).

Real-time quantitative PCR (qPCR) following ChIP. The PCR mixture contained Platinum Taq polymerase (Invitrogen) and optimized concentrations of Sybr-Green (Invitrogen). The sequences of primers used are listed in Table S2qPCR was performed in triplicate on at least two, in most cases three independent biological replicates of chromatin obtained from different sets of fly heads on different days. For each primer set, 1:10 dilutions of ChIP and input DNA from both strains were prepared from the same master mix. Additionally, for each primer set, one of the inputs was further diluted to make 1:100 and 1:1,000 dilutions. This was necessary to create calibration curves to ensure reaction linearity. Fluorescence intensities were plotted versus the number of cycles by using an algorithm provided by Corbett Research (Qiagen). For each gene, at least two primer sets were designed. Only reactions whose R^2 values were > 0.99 were considered valid. Real-time qPCR was performed a Corbett Research Rotor-Gene 3000 real-time cycler or a RotorGene Q real-time PCR machine with cycling parameters: 3 min at 95°C, followed by 40 cycles of 30 sec at 95°C, 45 sec at 55°C, and 45 sec at 72°C. One set amplified regions around or immediately adjacent to CANNTG canonical E-boxes in genes' first introns or first exons / 5' UTR. A second set amplified genomic regions approximately 6 kilobases upstream of E-box primer sets. Each ChIP was normalized to its input by delta-Ct value. Delta-delta-Ct value was then calculated by subtracting delta-Ct of negative control from delta-Ct of the tagged strain. The fold-difference between experimental and control samples was obtained by 2^(-delta-deltaCt). The properties of the primers used are listed in Supplementary Table 2.

Quantitative Mass Spectrometry. Fly heads of the control (*GMR*>UAS-*ppMII*; UAS-*dimm*) and experimental (*GMR*>UAS-*ppMII*; UAS-*dimm*; UAS-RNAi) transgenic lines were collected in frozen state. MS-based quantitation was based on prior quantitative measurement approaches [14-16]. Frozen heads were homogenized in acidified acetone (40:6:1, acetone:water:concentrated HCl, v/v/v; 10 ul/head). The homogenate was centrifuged at 12,000

rpm for 5 min: supernatants were concentrated by drying in a SpeedVac (Thermo Electron Co.), and then reconstituted with 2% acetonitrile aqueous solution (containing 0.1% FA and 0.01% TFA) for subsequent analysis. The quantitative analysis on MII fully processed from ppMII was performed using capillary liquid chromatography coupled to matrix-assisted laser/desorption ionization time-of-flight mass spectrometry (CapLC-MALDI-TOF/TOF MS). We performed relative quantitation using two approaches. For the labeling approach, the samples were labeled with succinic anhydride as previous described [14-15]. Briefly, 2 μl of 2 M light and heavy forms of succinic anhydride (Sigma Aldrich) in DMSO were used to label the control and experimental samples, and the sample solutions were adjusted to be basic. After the reaction, the labeled control and experimental samples were combined before being desalted using spin columns (Pierce, Rockford, IL). The combined and desalted samples were then analyzed with CapLC-MALDI-TOF/TOF MS (Bruker Daltonics, Billerica, MA). 5 μl samples were separated with a reverse phase column (Alltech Associates Inc., Alltima HP C18, 150 mm × 300 um, 3 μm particle diameter, 100 Å pore size) at a flow rate of 2 μl/min. Solvent A contains 95% water, 5% acetonitrile, 0.1% FA, and 0.01% TFA, and Solvent B contains 5% water, 95% acetonitrile, 0.1% FA, and 0.01% TFA. The 45-min gradient started from 2% B to 20% B over 10 min, to 40% B in another 15 min, continued to 80% B in 7 min, and stayed at 80% B for 3 min before ramping back to 2% B. The fraction collection started at 20 min, and a total of 24 one-minute fractions were collected on a MALDI target for each sample. 1 μl of *a*-cyano-4-hydroxycinnamic acid (Sigma Aldrich) in 70/30 (v/v) acetonitrile/ water solution was used as matrix, and MALDI-TOF/TOF mass spectra were acquired in the mass range of 600-5000 Da. The relative peak intensities in the peak pair were used to indicate the relative amount of peptides between control and experimental samples.

In the label-free quantitation approach, *Drosophila* head extracts were analyzed with CapLC-MALDI-TOF/TOF MS described earlier just after they were reconstituted. Endogenous

peptide Drm-MT2 (hugin (*CG6371*) (m/z 942.59) and exogenous peptide MII (m/z 1710.69) have been well studied in *Drosophila* previously [2, 17], and their presence was determined by mass matching between the theoretical masses and the experimental masses at signal-to-noise ratio above 3 within 20 ppm. MT2 and MII were respectively isolated with MALDI-TOF/TOF tandem MS, and the intensity of the isolated parent ion was used to reflect the amount of the peptide after normalization. Student's t-test was used for statistics with at least three biological replicates.

RNAi transgenic lines used in this study.

CG#	**GENE**	**RNAi line**
CG13248	*CAT-4*	GD5384 (VDRC)
CG1275	*Cyt-b*$_{561\text{-}1}$	GD4103 (VDRC)
CG11254	*mael*	GD18198 (VDRC)
CG7785	*cddl6*	GD36650 (VDRC)
CG32850	*rnf11*	JF01121 (Harvard)
CG6522		GD22500(VDRC)
CG14321		GD29814 (VDRC)
CG17293	*wdr82*	GD25246 (VDRC)
CG31436		GD21341 (VDRC)
CG14621		GD8661 (VDRC)

SUPPLEMENTARY REFERENCES

1. Hewes, R.S., Park, D., Gauthier, S.A., Schaefer, A.M. and Taghert, P.H. (2003). The bHLH protein Dimmed controls neuroendocrine cell differentiation in *Drosophila*. Develop. *130*, 1771-1781.
2. Hamanaka, Y., Park, D., Yin, P., Annangudi, S.P., Edwards, T.N., Sweedler, J., Meinertzhagen, I.A., and Taghert, P.H. (2010). Transcriptional orchestration of the regulated secretory pathway in neurons by the bHLH protein DIMM. Curr. Biol. *20*, 9-18.
3. Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mo.l Biol. *3*, Article3.
4. Goecks, J., Nekrutenko, A., Taylor, J. and The Galaxy Team. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.
5. Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A. and Taylor, J. (2010). "Galaxy: a web-based genome analysis tool for experimentalists". *Current Protocols in Molecular Biology*. Chapter 19: Unit 19.10.1-21.
6. Kula-Eversole E., Nagoshi E., Shang Y., Rodriguez J., Allada R. and Rosbash M. (2010) Surprising gene expression patterns within and between PDF-containing circadian neurons in *Drosophila*. Proc. Natl. Acad. Sci. U S A. *107*, 13497-13502.
7. Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 29, e45.
8. Kosman D., Ip Y.T., Levine M. and Arora K. (1991). Establishment of the mesoderm-neuroectoderm boundary in the *Drosophila* embryo. Science *254*, 118-122.
9. Park D., Shafer O.T., Shepherd S.P., Suh H., Trigg J.S. and Taghert P.H. (2008b). The *Drosophila* bHLH protein Dimmed directly activates *Phm*, a gene encoding a neuropeptide amidating enzyme. Mol. Cell. Biol. *28*, 410-421.
10. Allan, D.W., Park, D., St. Pierre, S.E., Taghert, P.H., and Thor, S. (2005). Regulators acting in combinatorial codes also act independently in single differentiating neurons. Neuron *45*, 689-700.
11. Copenhaver, P.F. and Taghert, P.H. (1989). Development of the enteric nervous system in the moth. I. Diversity of cell types and the embryonic expression of FMRFamide-related neuropeptides. Dev. Biol. *131*, 70-84.
12. Menet J.S., Abruzzi K.C., Desrochers J., Rodriguez J. and Rosbash M. (2010) Dynamic PER repression mechanisms in the *Drosophila* circadian clock: from on-DNA to off-DNA. Genes and Dev. *24*, 358-367.
13. Sandmann, T., Jakobsen, J.S. and Furlong, E.E.M. (2007). ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in *Drosophila melanogaster* embryos. *Nat. Protoc. 1*: 2839-2855
14. Brockmann, A., Annangudi, S.P., Richmond, T.A., Ament, S.A., Xie, F., Southey, B.R., Rodriguez-Zas, S.R., Robinson, G.E. and Sweedler, J.V. (2009) Quantitative Peptidomics Reveal Brain Peptide Signatures of Behavior. Proc. Natl. Acad. Sci., U.S.A. *106*, 2383–2388.
15. Ramos-Ortolaza, D.L., Bushlin, I., Abul-Husn, N., Annangudi, S.P., Sweedler, J. and Devi, L.A. (2010). Quantitative neuroproteomics of the synapse. Methods Mol. Biol. *615*, 227-246.
16. Hattan, S.J. and Parker, K.C. (2006). Methodology utilizing MS signal intensity and LC retention time for quantitative analysis and precursor ion selection in proteomic LC-MALDI analyses. Anal. Chem. *78*, 7986-7996.
17. Baggerman, G., Cerstiaens, A., Loof, A.D. and Schoofs, L. (2002). Peptidomics of the Larval *Drosophila melanogaster* Central Nervous System. J. Bio. Chem. *277*, 40368–40374.

Chapter 3.

Integrative analysis of the gene regulatory interactome and transcriptome of the neuroendocrine scaling factor DIMMED

This chapter will be adapted for the manuscript:
T. Hadžić, K.C. Abruzzi, D. Park, J.S. Trigg, M. Rosbash and P. Taghert: Integrative analysis of the gene regulatory interactome and transcriptome of the neuroendocrine scaling factor DIMMED. In preparation.

I performed the following experiments presented in this chapter: ChIP-chip (genetic crosses, collections and temperature shift, Western blots, ChIP, tiling array hybridization (with help from KCA for the first biological replicate), bioinformatic analysis of ChIP-chip data: MAT and Galaxy analysis, intersection, conservation, genomic annotation, selection of enhancer fragments for luciferase testing), RNA-Seq (genetic crosses, dissection (with help from PHT, DP, Laura Duvall and Seol Hee Im), cell dissociation for FACS, RNA purification, bioinformatics analysis of RNA-Seq data: enrichment quantification, heat map generation and intersection of ChIP-chip and RNA-Seq data.)

INTRODUCTION

Neuroendocrine (NE) cells possess features characteristic of neurons and glandular tissue (Scharrer B 1987). They have evolved specialized adaptations to synthesize, process and store large amounts of neuropeptides until they receive an appropriate signal for their release (Dannies 1999; Kim et al. 2006). To accomplish this, NE cells possess a repertoire of neuropeptide processing enzymes, as well as a special subcellular compartment, the Large Dense Core Vesicles (LDCVs, Dannies 1999; Kim et al. 2006). LDCVs house most of the processing enzymes and the neuropeptide cargo (Dannies 1999; Kim et al. 2006). LDCVs are also the key intracellular sites for neuropeptide processing, for long-term storage, for transport and ultimately for the secretion of the biologically active peptides (Burbach et al. 2001). LDCVs inside NE cells are made, stored, trafficked and exocytosed out of the cell thanks to a large complement of structural and regulatory components that ensure precisely timed release in response to stimuli (Crivellato et al. 2010). With the exception of a few proteomic studies, LDCV components remain largely undefined by high throughput studies (Gauthier et al. 2008; Wegrzyn et al. 2010). By utilizing this special subcellular compartment, NE cells have developed a unique mode of protein secretion, the Regulated Secretory Pathway (RSP; Dannies 1999; Kim et al. 2006). The primary function of this pathway is the secretion of substances (mainly neuropeptides) that produce modulatory changes in target neurons and other cells (De Camilli and Jahn 1990). NE cells, through the RSP, engage in episodic release of cargo that has the potential to affect cells located very far away (Burbach et al. 2001). Although fast neurotransmission is the dominant form of neural signaling, NE cells have evolved in parallel with neurons (Südhof 2007). NE cells integrate inputs and adapt their responses to various stimuli in order to maintain long-term homeostasis in living organisms (Burbach et al. 2001).

Thus far, little is known about how LDCVs and the RSP are established and developed in NE precursor cells during their maturation into professional secretory cells. Only a few groups

have used microarrays to obtain gene expression profiles of several normal and stimulated NE tissues in mammals and *Drosophila* (Yue et al. 2006; Hindmarch et al. 2006; Kula-Eversole et al. 2010; Nagoshi et al. 2010). Whereas Yue et al. (2006) used laser capture microdissection to collect individual cells from the hypothalamic supraoptic nucleus, Hindmarch et al. (2006) used more heterogeneous tissues from whole hypothalamic nuclei. These two studies suggest the existence of coordinately regulated NE-specific gene batteries whose function is to ensure proper secretion from NE tissues (Park and Taghert 2009).

In *Drosophila*, two groups have examined gene expression profiles of purified *Drosophila* circadian neurons, some of which are neuroendocrine (Kula-Eversole et al. 2010; Nagoshi et al. 2010). In both papers, manual collection of freshly dissociated cells was used to source the RNA material (Kula-Eversole et al. 2010; Nagoshi et al. 2010). Both studies focused on the portion of the transcriptome that cycles with circadian time in clock cells. The advantage that these studies had over others is that the circadian system in *Drosophila* is one of the most well understood systems implicated in the control of behavior (Nitabach and Taghert 2008). On the other hand, what remains unanswered is how cells such as the hypothalamic thirst-controlling NE neurons and the *Drosophila* circadian NE neurons establish those properties required of professional secretory cells: a robust RSP with an appreciable LDCV subcellular compartment and an appropriately regulated transcriptome and epigenome.

Generally speaking, the transcriptome of NE cells likely consists of those genes that are required in all cells, and others that allow NE cells to perform their unique roles. Although the “house keeping” portion of the transcriptome is likely not identical in all cells, it probably consists of a core set of genes with some quantitative variations in their expression (Doyle et al. 2008). On the other hand, each cell type has a unique portion of its transcriptome that orchestrates its functions. The definition of “unique” changes depending on the reference cell type to which the cell population of interest is being compared to (Dougherty et al. 2010). Thus, the degree of

transcriptome uniqueness changes depending on whether a cell type is being compared against a related cell type (e.g. one type of neuron against another) or against a distantly related cell type (e.g. neurons against germ line cells). Regardless of the characteristics, there likely exist specialized intracellular circuits and pathways that set up, amplify and maintain a particular distinguishing cell feature, such as the RSP in NE cells, synaptic vesicles in neurons or mitochondria in cardiomyocytes (Mills and Taghert 2011). Furthermore, such master regulators, transcriptional organizers and scaling factors are likely not static, but must be able to adapt to the ever changing environment (Mills and Taghert 2011.

Two known transcriptional organizers of the developmental programs of neuronal cell subtypes are factors PET-1 and AST-1 (Park and Taghert 2009; Hendricks et al. 1999; Flames and Hobert 2009). AST-1 is an ETS transcription factor that controls expression of all dopamine pathway genes in all dopaminergic cell types in *Caenorhabditis elegans* (Flames and Hobert 2009). PET-1 is another ETS domain transcription factor that appears to control expression of serotonin pathway genes in mammalian serotonergic neurons (Hendricks et al. 1999). Both of these factors control major elements of the biosynthetic machinery known to distinguish their respective neuronal subtypes from neurons expressing other biogenic amines or classic neurotransmitters. Neither factor, however, is necessary for the initial fate specification or survival of the respective neuron type that they control (Park and Taghert 2009). There is also no evidence as of yet that either of these factors controls a whole subcellular compartment, such as LDCVs, endosomes or synaptic vesicles.

One known factor that controls a whole subcellular compartment is Transcription Factor EB (TFEB, Sardiello et al. 2009). TFEB exerts coordinated transcriptional control over most mammalian lysosomal genes (Sardiello et al. 2009). *In vitro*, TFEB induces lysosomal biogenesis and increases degradation of lysosomal complex molecules (Sardiello et al. 2009). Another factor, Peroxisome proliferator–activated receptor Gamma Coactivator-1 (PGC-1), has

been shown to control the number of mitochondria inside heart cells, as well as their function in response to energy demands (Lehman et al. 2000; Finck and Kelly 2007). Though it has been convincingly shown that PGC-1 is a master regulator of myocardial energy metabolism, this protein is a not a sequence-specific transcription factor, but rather a co-activator for such factors (Finck and Kelly 2007).

The above examples demonstrate master regulators / scaling factors that control the biogenesis, maintenance and homeostasis of lysosomes and mitochondria (Sardiello et al. 2009; Lehman et al. 2000). Another subcellular compartment of great consequences for human health is the LDCV compartment. Peptide hormone insulin is stored in LDCVs of pancreatic beta cells and any pathologic changes in LDCVs can lead to beta cell disease and diabetes (Liu et al. 2009). In general, LDCVs are found in NE, endocrine, as well as other secretory cells such as exocrine and hematopoetic mast cells (Crivellato et al. 2010; Kim et al. 2006). A known NE-specific master regulator and scaling factor that is likely responsible for the establishment and dynamic maintenance of the RSP in NE cells is the *Drosophila* basic helix loop transcription (bHLH) factor *dimmed* (DIMM) (Park and Taghert 2009; Hewes et al. 2003; Park et al. 2008b; Hamanaka et al. 2010; Park et al. 2011). DIMM coordinates the molecular and cellular properties of all major NE cells, irrespective of the secretory peptides they produce (Park et al. 2011). DIMM is a well conserved protein, with a mammalian orthologue called Mist1 (Moore et al. 2000). Mist1 is the transcriptional organizer of serous exocrine cell-specific secretory features in mammals (Pin et al. 2000; Ramsey et al. 2007; Johnson et al. 2004).

The goal of this thesis was to identify the precise mechanism of DIMM action *in vivo*. Since DIMM is a transcription factor, it functions by binding to and activating expression of a set of genes. Therefore, I took the specific aim of defining all DIMM genomic targets. Chromatin immunoprecipitation (ChIP) has become the technique of choice to identify direct genomic targets of a transcription factor (Collas 2009; Lee et al. 2006; Sandmann et al. 2007). DNA

sequences from DNA-protein adducts isolated by ChIP can be globally analyzed by microarray hybridization (ChIP-chip, Lee et al. 2006; Southall and Brand 2007). ChIP-chip allows for unbiased detection of protein–DNA interactions, as it requires no prior knowledge of candidate binding sites for any particular transcription factor (Sandmann et al. 2007). In a previously published study, (Chapter 2), I demonstrated the application of *in vivo* ChIP to analyze candidate DIMM binding sites in the enhancers of a few genes (Park et al. 2011). In Chapter 2, a broad multidisciplinary approach was undertaken to validate a small number of candidate DIMM-dependent genes as *bona fide* targets with functional consequences for NE cell physiology (Park et al. 2011). Nevertheless, the study was limited to evaluating a selection of 134 genes that were derived from a DIMM-over-expression paradigm in the embryonic nervous system (Park et al. 2011).

In this chapter, I extend the ChIP analysis to the whole genome by ChIP-chip. This *in vivo* approach allows for identification of the full complement of potential DIMM regulated sites in the *Drosophila* genome in an unbiased, systematic fashion. As I will describe, the ChIP-chip data yielded a stringent DIMM gene regulatory interactome of at least 156 high fidelity cis-regulatory modules, which is highly interesting by itself. Although Sardiello et al. (2009) used in vitro expression profiling to identify many lysosomal genes as TFEB targets, they did not conduct a genome-wide survey of TFEB binding. There have so far been no reports of scaling factor regulatory interactomes being revealed by ChIP-chip or ChIP coupled to deep sequencing (ChIP-Seq).

Once I successfully identified the gene regulatory interactome of DIMM, it became evident that DIMM's interactome could be greatly enhanced by integrating it with the transcriptome of DIMM-expressing LEAP cells. Therefore, I also describe herein how I purified DIMM+ and DIMM- cells by Fluorescence Activated Cell Sorting (FACS) in order to profile their gene expression by deep sequencing (RNA-Seq). Once the LEAP transcriptome was

successfully obtained, I intersected the DIMM gene regulatory interactome with the newly derived transcriptome of DIMM+ LEAP cells. This yielded a set of genes that are high fidelity direct DIMM targets.

MATERIALS AND METHODS

ChIP-chip

ChIP was conducted as described previously (Chapter 2, Park et al. 2011; Menet et al. 2010). Briefly, tagged ChIP was carried out with a DIMM::MYC-tag fusion transgene that can rescue diminished neuropeptide levels in DIMM null animals (Hewes et al. 2003). This transgene was expressed in a spatiotemporally controlled manner by using a GAL4 driver that overlaps to a great extent with DIMM (c929-GAL4; Hewes et al. 2003; Park et al. 2008a). Furthermore, I restricted GAL4 activity to DIMM-expressing neurons in adult animals by using a temperature sensitive allele of a ubiquitously expressed GAL80 protein (tub-GAL80^{ts}) – TARGET (McGuire et al. 2003; Brand and Perrimon 1993). In this system, GAL4 activation of transcription from the UAS element is inhibited by GAL80 as long as the animal is kept at the restrictive temperature (18°C, McGuire et al. 2003). Shifting the animals to 30°C (permissive temperature) results in a derepression of GAL4 activity and transcription from the UAS element. Two sets of animals were used for this experiment: i) c929-GAL4/UAS-DIMM::MYC; tub-GAL80^{ts}/+ [experimental group] and ii) c929-GAL4; tub-GAL80^{ts} maternal strain [negative control]. Both sets of animals were raised at 18°C to allow for normal development, as well as to prevent the lethality that results from DIMM overexpression in LEAP cells (unpublished results). Freshly eclosed adult flies were then shifted to 30°C to induce expression of the DIMM::MYC transgene only in adult DIMM+ cells. Specific induction of the DIMM::MYC protein within 72 hours was confirmed by Western blotting (Figure 1). I performed tagged ChIP 72 hours later, using methods described in Chapter 2 (Park et al. 2011).

One limitation of ChIP is that it requires a high quality antibody that can recognize fixed antigens in solution (Orlando 2000; Lee et al. 2006; Sandmann et al. 2007). A second limitation is that ChIP requires at least one million cells for success (Collas 2009). c929-GAL4 is expressed in about 300 NE cells in the adult brain, which is less than 1% of all CNS cells (D Park and P Taghert, unpublished observations). As expected from these technical limitations, ChIP with a native DIMM antibody failed to produce successful results. Therefore, I employed a "tagged ChIP strategy based on the availability of the MYC-tagged DIMM transgene. Affinity tags such as MYC are increasingly used in ChIP assays to detect *in vivo* binding of transcription factors to their target genes in chromatin (e.g., Kolodziej et al. 2009).

Following crosslink reversal, DNA was processed according to the Affymetrix® Chromatin Immunoprecipitation Assay Protocol (P/N 702238 Rev. 3). For each sample, 10 microliters of undiluted ChIP DNA or 1:10 diluted input DNA was amplified by linear PCR-based amplification. Quantitative PCR was used to show that DIMM occupancy at its known binding site in PHM was preserved after amplification. Identical amounts (6.0 micrograms) of each amplified sample were then fragmented and biotin-labeled by Terminal Deoxynucleotidyl Transferase. Labeled samples were then hybridized to Affymetrix GeneChip® *Drosophila* 2.0 Tiling Arrays and scanned according to Affymetrix protocols at the National Center for Behavioral Genomics (Brandeis University).

Bioinformatic analysis of ChIP-chip data

To detect statistically significant DIMM binding peaks throughout the genome, an analysis algorithm titled Model-based Analysis of Tiling-arrays (MAT) was used (Johnson et al. 2006). MAT considers the 600-basepair window surrounding each tiling array probe and computes a trimmed mean of all of the t values in the window (Johnson et al. 2006). Additionally, MAT models baseline probe behavior by considering the 25-mer probe sequence and copy number of all probes on a single tiling array (Johnson et al. 2006). MAT detects ChIP regions using an

enrichment measure known as MAT score, which is calculated for each 600 bp window and assigned to the probe at the center of the window. Essentially, MAT score, which is calculated from normalized probe intensities, represents a statistical likelihood that a particular genomic region is enriched in the immunoprecipitated relative to the control sample (Abruzzi et al. 2011). MAT analysis was carried out with MAT source build 2.11272006 on a 32-bit Linux machine. Genome visualization .bar files produced by MAT are incompatible with the University of California in Santa Cruz (UCSC) Genome Browser. Therefore, UCSC Genome Browser-compatible files were obtained by running an implementation of MAT (rMAT) in the statistical computing language R (Droit et al. 2010). This analysis produces a graphic representation of MAT scores in sliding windows on a log2 scale. Visual output from all three MAT analyses performed (DIMM ChIP/Input, NEG ChIP/Input and DIMM ChIP / NEG ChIP) can be directly compared to each other as long as the window minima and maxima are identical. This is true because an identical number of arrays was used for experimental and control samples, with the same number of biological replicates and DNA amounts hybridized to tiling arrays.

Genomic annotation of DIMM binding sites

Annotation of DIMM binding sites with respect to introns (Figure 3A) was performed in the Galaxy, an open platform for genomic research (Goecks et al. 2010). A data table describing all *Drosophila* introns from the UCSC dm3 *Drosophila* genome build (UCSC Main on *D. melanogaster*: flyBaseGene: Introns) was downloaded from the UCSC Genome Browser database. Galaxy queries were then run to identify the percent of overlap between DIMM-bound regions and all known introns from the dm3 genome build (Figure 3A). To annotate DIMM binding sites against exons, introns and promoters, PinkThing was used (Figure 3B). PinkThing is an open-source software for quick annotation of a set of genomic regions against transcription start sites and gene features (TSSs, PinkThing.cmbi.ru.nl, F. Nielsen, M. Kooyman and M. Huynen; Kramer et al. 2011). PinkThing was used under default parameters to annotate DIMM

binding sites against genomic features defined as: gene body (introns versus exons), as well as upstream and downstream regions. A DIMM-bound region was classified as: Near Upstream or Downstream = 5 kb upstream or downstream of a TSS; Far Upstream = 5 – 25 kb from the closest TSS and Distant Upstream = >25 kb from the closest TSS). Furthermore, precise distance of ChIP-chip binding sites from transcription start sites was calculated and plotted in the statistical programming language R v. 2.14.0 with the BioConductor package ChIPpeakAnno (Zhu et al. 2010). ChIPpeakAnno allows for visual representation of peak distance from annotated TSSs (Zhu et al. 2010).

Binding motif identification and conservation analysis

Binding motif analysis was carried out in the Cistrome project with the SeqPos motif finding tool (He et al. 2010; Liu et al. 2011). DIMM binding peaks with 1% - 100% of intron overlap were analyzed with the following settings: region width = 600 bp; p-value cutoff = 0.05. For conservation analysis, Galaxy queries were used to extract Multiple Alignment File (MAF) blocks for the set of 156 DIMM-bound genomic intervals and a 15-way multiZ (dm3) alignment with the genomes of twelve other *Drosophila* species, as well as the bee (*Apis mellifera*) and mosquito (*Anopheles gambiae*) genomes (Blankenberg et al. 2011). A query to calculate MAF Coverage Stats was then run in Galaxy and the resulting coverage reported on a scale from 0 (no conservation) to 1 (100% conservation).

In vitro luciferase DIMM transactivation assays

Luciferase assays were performed as in Chapter 2 (Park et al. 2011. DIMM-bound genomic fragments that were tested in these assays were selected based on their MAT scores, and in some cases based on their intronic location. Assays were performed with at least n=3 biological replicates with the exception of the crc-F fragment.

Preparation of cells for FACS sorting

Brains of homozygous w^{1118},UAS-dcr2; c929-GAL4, UAS-CD8::EGFP; flies were used to harvest cells for FACS sorting and subsequent RNA-Seq. UAS-dcr2 was included in the genetic background because of plans to carry out the same type of analysis in a *dimm* RNAi loss-of-function background. This strain (w^{1118},UAS-dcr2; c929-GAL4, UAS-CD8::EGFP;) is essentially wild-type with respect to DIMM physiology, as evidenced by normal neuropeptide staining (data not shown). Young flies (5-12 days old) of both sexes were grown at 25°C on standard *Drosophila* cornmeal media. They were anesthetized with carbon dioxide and collected on ice. RNA was isolated from brains dissected on two separate occasions, in order to get as much material as possible, to minimize RNA amplification. Brains of 140 (first experiment) and 180 flies (second experiment) were dissected within a two to three hour time period. Adult heads were detached from ventral nerve cords and dissected in cold saline as in Nagoshi et al. (2010) with some modifications. Sterile, ice-cold Modified Dissecting Saline (MDS) was used for dissection (9.9 mM HEPES-KOH buffer, 137 mM NaCl, 5.4 mM KCl, 0.17 mM NaH_2PO4, 0.22 mM KH_2PO4, 33 mM glucose, 43.8 mM sucrose, pH 7.4; Jiang et al. 2005). After dissection, brains were immediately transferred into sterile, chilled modified SM^{active} medium and pooled (SM^{active} medium containing 5 mM Bis-Tris; Küppers-Munther et al. 2004).

At the end of dissection, brains were pooled in a 2 ml Eppendorf DNA LoBind nuclease-free tube and washed with 1 ml of chilled MDS. Fly brains were then centrifuged at 1,000g for 30 seconds in a microcentrifuge. Frozen L-cysteine-activated papain (50 units/ml in dissecting saline; Worthington) was thawed at the beginning of dissection and activated at 37°C for 20 minutes immediately prior to the end of dissection. Four hundred and fifty microliters of activated papain were then added to the brains. Fly brains were gently resuspended and incubated with papain for 20 minutes at room temperature. Papain digestion was quenched by adding 1.5 ml of Schneider's medium supplemented with 1% heat inactivated Fetal Calf Serum. The sample was centrifuged at 500g for 30 seconds at room temperature. Brains were washed two more times

and were finally resuspended in 600 microliters of Schneider's medium supplemented with 1% FCS.

Papain-digested brains were then triturated 30 times with a flame-rounded P1000 filter tip (medium tip opening), 30 times with a flame-rounded P1000 filter tip (small opening), 30 times with a flame-rounded P200 filter tip (large opening), 40 times with a flame-rounded P200 filter tip (medium tip opening) and 20 times with a flame-rounded P200 filter tip (small opening). Pipette volumes used for trituration were 500 microliters and 150 microliters for the P1000 and P200 pipettes, respectively. Trituration efficiency was such that very small pieces of brain tissue were still visible under a dissection microscope.

Brain homogenate was then strained through a sterile 70 micron cell strainer (Fishersci) mounted on a 50 ml Falcon, then centrifuged briefly at 400 rpm at 4˚C. Flow through containing filtered dissociated cells was then strained through a 35 micron sterile cell strainer (BD Falcon) mounted on a 50 ml Falcon, then centrifuged briefly at 200 rpm. Four hundred and fifty microliters of dissociated single cells were collected in a 6 ml BD Falcon tube. Cells were then transported on ice to the Washington University Siteman Flow Cytometry Core for FACS sorting.

FACS sorting

Cell sorting was performed on a DAKO-Cytomation MoFlo High Speed Sorter at the Siteman Flow Cytometry Core, Washington University in St Louis. The Green Fluorescent Protein (GFP+) gate was set based on established criteria and experience of the Core's personnel with the sorter (Figure 5). Sorting GFP+ and GFP- cells was completed within on hour of the start of sorting. Roughly, 1% - 3% of all cells in the samples were GFP+, which is consistent with the number of c929+ cells in the adult fly brain (Park and Taghert 2009). In order to establish the viability of cells after dissociation, a pilot sorting experiment was performed in which cells were sorted for GFP, as well as their ability to exclude the vitality dye 7-Amino-Actinomycin D. This

nucleic acid dye allows detection of dying cells by entering such cells thanks to fragmentation of their membranes. (Schmid et al. 1992). The pilot experiment showed that the majority of gated GFP+ and GFP- cells were excluding 7-AAD, and were therefore alive at the time of sorting. The fact that sorted DIMM+ cells were viable shortly before RNA capture increased confidence that any subsequent expression profiling would faithfully reproduce the transcriptome of live DIMM+ cells. Because double sorting against GFP and 7-Amino-Actinomycin D doubled the amount of time needed for sorting, sorting experiments for RNA collection were conducted without 7-Amino-Actinomycin D. The combined number of cells sorted from the two experiments was 3.0 x 10^5 GFP+ cells and 1.35 x 10^6 GFP- cells.

RNA isolation, processing and Illumina HiSeq 2000 sequencing

At the end of sorting, cells were transferred directly into Qiagen's RLT buffer supplemented with beta-mercaptoethanol (Qiagen RNA MinElute kit). Most investigators harvesting nano scale RNA samples for microarray gene expression profiling use the Arcturus PicoPure kit (Borghese et al. 2006; Nagoshi et al. 2010; Ramsey et al. 2007; Tian et al. 2010). This kit contains a poly(dI:dC)-based proprietary nucleic carrier embedded in the column. This carrier does not interfere with microarray applications, but could interfere with deep sequencing of RNAs isolated this way. Therefore, we decided to use the Qiagen kit for small samples, which lacks carriers. After transferring cells into RLT buffer, they were vortexed for 6 seconds, and lysed by passing the suspension five times through a 21.5 gage needle mounted on a 3 ml syringe. RNA was isolated as recommended by Qiagen's MinElute protocol, with the exception of using 65°C-heated water for enhanced RNA recovery. Next, DNA in the sample was digested with the DNaseI Turbo DNA-free kit (Ambion). After DNA removal, RNAs isolated from GFP+ and GFP- cells from each of the two experiments were pooled.

Samples with pooled GFP+ and GFP- RNAs were then submitted to the Genome Technology Access Center (Washington University Genetics Department). RNA quality was

then checked by Qubit (Invitrogen), NanoDrop (GE Health) and Bioanalyzer (Agilent) quality control assays. Neither sample showed signs of degradation or DNA contamination, and the samples were moved into the deep sequencing pipeline at the same facility. After taking sample aliquots for quality control, there were ~35 nanograms of RNA from GFP+ cells left and ~100 nanograms of GFP- RNA. Since the two samples would be directly compared to each other, the GFP- sample was diluted in order to have equal quantities of both samples. The samples were sequenced on Illumina's HiSeq 2000 platform (Han et al. 2011).

Because Illumina's deep sequencing protocols require 10 micrograms of RNA, the samples had to be amplified. Although various oligo dT-based RNA amplification protocols have been used, a new product emerged recently, which foregoes this type of amplification. The NuGen Ovation RNA-Seq system is a single primer-based RNA amplification product that uses isothermal amplification (Kurn et al. 2005; Dafforn et al. 2004). In this system, the RNA is reverse transcribed, then partially degraded in order to synthesize a second DNA strand from the first-strand cDNA template (Kurn et al. 2005; Dafforn et al. 2004). Double-stranded DNA is purified and amplified using single primer isothermal amplification. This amplification process starts with RNase H cleaving RNA in DNA/RNA heteroduplexes located on the ends of double-stranded DNA. Next, the proprietary primer binds to cDNA and polymerase starts replication at the end of the primer by displacing the existing forward strand. Finally, random hexamers initiate linear amplification of the second-strand cDNA (Kurn et al. 2005; Dafforn et al. 2004). After amplification with the NuGen system, samples were prepared for deep sequencing according to standard Illumina procedures. This included a poly-A selection step with beads, as well as linker ligation with unique multiplexed ends. Multiplexing allowed each sequenced read to be distinguished by a 6 basepair (bp) tag at the end of the sequencing adaptor. Sequencing was performed in a single lane of an Illumina HiSeq 2000 machine. Base calls were made by

Illumina's software (Eland). Next, reads were de-multiplexed to allow the GFP+ sample's reads to be distinguished from the GFP- sample.

An algorithm called Tophat was then used to align raw sequence reads to the reference genome and to map splice junctions (Ensembl *Drosophila* BDGP5.25.62 release; Trapnell et al. 2009). Tophat output was then used as input for the algorithm Cufflinks v. 1.0.3, which assembled whole transcripts and estimated their abundance (Roberts et al. 2011). Cuffdiff was then used to test for differential expression in RNA-Seq samples. Cuffdiff identified 4,846 out of 22,872 transcripts that were enriched at 1.5-fold or higher in the GFP+/GFP- sample. Tophat-generated alignments of RNA-Seq data in BAM format were indexed and sorted for visualization in the Integrated Genomics Viewer (Robinson et al. 2011). Indexed RNA-Seq reads were visualized as coverage tracks. In order to be able to compare samples directly, coverage tracks were always graphed with identical y-axis coordinates. Heat maps were generated in MultiExperiment Viewer (MeV; Saeed et al. 2006).

ChIP-chip and RNA-Seq data integration

ChIP-chip and RNA-Seq gene lists were integrated using Galaxy (Goecks et al. 2010). In order to establish whether or not the statistical overlap was due solely to chance, a cumulative hypergeometric probability was calculated as follows: population size = 22,872 total transcripts; number of successes in population = 4,846 enriched transcripts; sample size = 597 ChIP-chip associated transcripts; number of successes in the sample = 170 ChIP-chip transcripts that are enriched at 1.5-fold or higher in the population. Instead of a simple hypergeometric probability, the more stringent cumulative probability of obtaining 170 or more hits due to chance alone was calculated: $P(X \geq 170)$ was equal to 1.21×10^{-5}.

RESULTS

Genome-wide identification of direct DIMM targets by *in vivo* ChIP-chip

detection of genome-wide DIMM binding

To identify *in vivo* DIMM occupancy at binding sites throughout the genome, a tagged ChIP-chip strategy was employed (Ezhkova and Tansey 2006; see Materials and Methods). Success of the tagged ChIP approach was measured by examining DIMM occupancy at its known binding site in the first intron of PHM and lack of occupancy at a nearby (control) site (Park et al. 2011). Once I could confirm the specificity of DIMM binding to PHM intron 1, I initiated ChIP-chip analysis.

I performed anti-MYC ChIPs from both experimental and control *Drosophila* head extracts. DIMM-bound DNA fragments were identified using *Drosophila* Tiling Arrays 2.0 (Affymetrix). For each genotype, arrays were hybridized with ChIPed DNA, as well as pre-ChIP genomic input DNA (n=2 biological replicates for each genotype and sample type). Thus, I used a total of eight arrays: two ChIP arrays, two input arrays per genotype. I identified DIMM binding peaks using the MAT algorithm (Johnson et al. 2006).

In order to identify true DIMM binding in a rigorous way, I employed three different types of MAT comparative analyses: i) DIMM::MYC ChIP sample was compared against its input (DIMM ChIP/Input), ii) the control (maternal strain) ChIP sample was compared against its own input (NEG ChIP/Input), and iii) DIMM ChIP sample was compared against the control ChIP sample (DIMM ChIP / NEG ChIP). The last comparison (DIMM ChIP / NEG ChIP) was scientifically valid because an identical amount of DNA was hybridized to each tiling array, regardless of sample type or genotype. Each of these MAT analyses, including the second analysis (NEG ChIP/Input) yielded enriched regions at a commonly accepted p-value of 1×10^{-5}: i) DIMM ChIP/Input - 5,644 bound regions ii) NEG ChIP/Input - 3,177 regions and iii) DIMM ChIP / NEG ChIP - 258 regions.

Although ChIP-chip from the NEG ChIP/Input (control) sample was theoretically supposed to yield no specifically-bound regions by MAT, many regions was detected as significant (3,177 regions). This likely happened due to the MYC-tag antibody binding non-specifically to DNA, or due to specific interactions between genomic DNA and agarose beads, and other reaction vessels (Sandmann et al. 2007). Nevertheless, increasing the statistical stringency of MAT analysis of both samples by progressively decreasing the p-value eventually reduces the number of bound regions in the NEG ChIP/Input to a few, while leaving many more bound regions in the DIMM ChIP/Input sample intact. Based on the need to fairly compare all three analyses against each other, I left the p-value cutoff unchanged. Instead, after consultations with the author of MAT (X. Liu, personal communication), I decided that the most prudent way to analyze the data was to subtract NEG ChIP/Input bound regions from the DIMM ChIP/Input bound regions, and then intersect the resulting regions with the 258 bound peaks in the DIMM ChIP / NEG ChIP comparison.

The resulting intersection of the DIMM ChIP/Input bound regions lacking any NEG ChIP/Input overlap with DIMM ChIP / NEG ChIP bound regions yielded a total of 156 bound peaks (Table 1). As with ChIP-qPCR in Chapter 2, I again used DIMM occupancy in the first intron of PHM as an internal positive control. Indeed, by DIMM ChIP / NEG ChIP analysis, the DIMM binding peak in the first intron of PHM was the fifth highest scoring peak (MAT score 139), and it was the 25th most highly scoring peak in the DIMM ChIP/Input comparison (MAT score 280.8). Visualization of this previously known *bona fide* DIMM enhancer in the genome browser is shown in Figure 2. Lack of any significant binding in the NEG ChIP/Input comparison is striking. I therefore submit that the subtraction and intersection of the three MAT analyses yields a high quality data set of DIMM-occupied regions.

Genomic annotation of DIMM-bound peaks

Having identified DIMM-occupied genomic regions, I then annotated them with respect to various genomic features such as transcription start sites (TSSs), exons and introns. It has been well documented that bHLH transcription factors bind not only to promoters, but also to introns, where they can recruit the basal transcriptional machinery (Cao et al. 2010; Tian et al. 2010; Menet et al. 2010). DIMM binds to specific palindromic E-boxes (CATATG and CAGCTG) that are located in introns of target genes in addition to PHM (Park et al. 2008; Park et al. 2011). Similarly, its mammalian orthologue Mist1 binds to CATATG E-boxes located in the first intron of 6 of its targets (Tian et al. 2010).

Of the 156 high stringency DIMM-bound peaks, 51 (33%) do not overlap with introns, although some might be in their vicinity. Another 43 peaks (28%) have a degree of intron overlap ranging between 1% and 50% of their length. 26 more peaks have 50% - 97% overlap with an intron. Finally, 36 peaks (23%) have 100% overlap with an intron (Figure 3A). Binding sites were also annotated with respect to introns, exons and promoters by using a gene structure annotation algorithm (PinkThing, Figure 3B). This analysis showed that the majority of the peaks overlap with introns (39%) and promoters (23%).

Next, a search for statistically overrepresented sequence binding motifs was conducted and one E-box motif was isolated. The E-box motif with the consensus sequence CATATGKTTS was isolated with a z-score -2.6297 and a p-value of 0.004273. Statistical overrepresentation of a motif containing the CATATG E-box is entirely consistent with the E-boxed that DIMM binds to in the first intron of PHM (Park et al. 2008b). When conservation status of all 156 DIMM-bound peaks was examined by employing pairwise alignment comparisons between DIMM-bound regions and homologous regions in 13 insect species, a high degree of conservation was observed (Figure 3D). Nucleotide scores ranging from 0.956

for droSec1 genome to 0.513 for the droGri2 *Drosophila* species, and lower conservation with bees and mosquitoes (0.11 - 0.15; Figure 3D).

The 156 DIMM-bound peaks that passed the high stringency criteria were annotated by using tools from the open source Galaxy platform for genomic analysis (Goecks et al. 2010). Peaks were annotated by fetching the closest non-overlapping genomic features from: all *Drosophila* exons, introns, 5 kilobase upstream regions and 5' untranslated regions (UTRs). Using this annotation strategy, 156 peaks as input produced a total of 284 DIMM ChIP-chip binding-associated genes (Table 2). Additionally, the 284 genes that are associated with DIMM-bound regions produce a total of 598 transcripts.

DIMM transactivation of enhancers identified by *in vivo* ChIP-chip

Based on previous evidence, DIMM and Mist1 likely function as transcriptional activators (Park et al. 2008b; Park et al. 2011; Tian et al. 2010). To support that supposition, we tested DIMM's ability to transactivate ChIP-chip-identified-regulatory fragments of the candidate gene targets, using luciferase levels as a readout. This assay (carried out by Ms. Jennifer Trigg) used the BG3 *Drosophila* neuronal cell line as described in Chapter 2 (Park et al. 2011). 39 of 156 DIMM-bound regions were tested in luciferase transactivation assays (Figure 4). DIMM transactivated 19 out of 39 fragments with reporter induction ranging from 5.9 to 93-fold (p-value < 0.05). Interestingly, many of the tested enhancers produced inductions that were many fold higher than the strongest inductions recorded previously (genes PHM, mael and CG6522, Park et al. 2011). The remaining enhancers showed mild induction or their induction was indistinguishable from controls. These data suggest that DIMM can indeed directly activate gene expression from a considerable fraction of enhancers identified by *in vivo* ChIP-chip.

Deep sequencing the LEAP transcriptome

The usefulness of a gene regulatory interactome can be greatly enhanced with information about the gene expression landscape operating in the same cells as the

transcription factor of interest. In order to obtain a transcriptional profile of LEAP cells, DIMM+ cells were labeled with GFP, and GFP+ cells from adult fly brains were FACS sorted. GFP+ cells and a group of randomly chosen GFP- cells of similar size (likely a stochastic mixture of neurons and glia) were sorted, their RNAs isolated and profiled by deep sequencing (RNA-Seq). All data analysis was done by comparing the DIMM+/GFP+ transcriptome directly to the DIMM-/GFP- transcriptome. This was a fair comparison because the samples were processed identically, in parallel and yielded similar numbers of RNA-Seq reads: 36.9 million 42-bp reads from the GFP+ sample and 39.6 million 42-bp reads from the GFP- sample. The majority of the reads mapped to the *Drosophila* genome (23.9 million / 36.9 million GFP+ reads and 32.3 million / 39.6 million GFP- reads). Furthermore, 34.8% of GFP+ reads were unique reads (one alignment in the genome) and 29.99% were multiple alignments (likely to genomic repeat regions). The GFP- sample had 56.33% uniquely aligned reads and 25.45% multiply aligned reads. The algorithm for comparing gene expression levels between RNA-Seq data standardizes all samples in terms of fragments per million base pairs sequenced (Roberts et al. 2011). Despite some differences in the total number of aligned reads between the GFP+ and GFP- samples, the two samples are directly comparable by Cufflinks.

After determining that the RNA-Seq datasets contained high quality sequence that aligned to the *Drosophila* genome, the DIMM+/GFP+ sample's transcriptome was compared to the DIMM-/GFP- transcriptome. Initial analysis showed that 4,676 isoform-specific transcripts out of 22,872 total transcripts were enriched in LEAP cells at a minimum 1.5-fold enrichment. The first specific measure examined was the enrichment of DIMM transcripts in the GFP+ against the GFP- sample. Indeed, DIMM transcripts were enriched at 138-fold in the GFP+/GFP- comparison. Next, I examined a group of genes known to be enriched in DIMM cells compared to other cell types. In all cases, the enrichment prediction matched the real data (Figure 6). For example: several of the seven Insulin-like peptides (Ilps) are known to be

enriched in LEAP cells, whereas other family members are expressed outside of the CNS (Park et al. 2008a; Brogiolo et al. 2001). The known LEAP cell-enriched Ilps were indeed enriched by RNA-Seq (Ilp5: 1.6-million fold, Ilp3: 688-fold, Ilp2: 489-fold, Ilp6: 26-fold; Figure 6A,C), whereas Ilps with known lack of CNS expression were not enriched (Ilp1: 0.3-fold enrichment, Ilp4 and Ilp7: no expression in either sample; Figure 6A,C). While DIMM overlaps with many peptides, the extent of DIMM overlap depends on the peptide (Park et al. 2008a). A few peptides have low or almost no overlap with DIMM (Park et al. 2008a). As expected, such peptides did not show significant RNA-Seq enrichment (Proctolin: 4% of larval LEAP cells and 1.3-fold enrichment; Ptth: 0% larval LEAP cells and 1.5-fold enrichment, Dh31: 5.3% of larval LEAP cells and 1.6-fold enrichment; Figure 6A). Half of Pigment dispersing factor (PDF+) neurons overlap with DIMM in the adult brain, which matches RNA-Seq data (Figure 6B; Park and Taghert 2009). Previous reports have indicated that while PHM overlaps extensively with DIMM, the overlap is not exclusive (Park and Taghert 2009). In accordance with this fact, RNA-Seq data showed that Phm was enriched 3.5-fold in LEAP cells over non-LEAP cells (Figure 7). CAT-4, another DIMM *bona fide* target was enriched at 2.77-fold (Park et al. 2011). Overall, these data confirm the high fidelity and quality of the RNA-Seq data set.

Integration of transcriptome (RNA-Seq) and interactome (ChIP-chip) results

The resolved LEAP cell transcriptome is an excellent resource that can be directly integrated with the ChIP-chip data set. The basic hypothesis here is that DIMM activates gene expression. Therefore, those LEAP-cell enriched genes that are adjacent to DIMM binding peaks likely represent the true targets of DIMM's transcriptional activity. DIMM binds to 156 regions in the genome, which corresponds to 284 genes and a total of 598 transcripts from those genes (Table 1). The LEAP transcriptome has 4,676 transcripts out of 22,872 total transcripts that are enriched at 1.5-fold or higher (Table 2). Of the 598 ChIP-chip-associated transcripts, 170 transcripts are amongst the 4,676 LEAP-cell-specific transcripts (Table 3,

Figure 7). This overlap is more than that due to chance alone (hypergeometric probability $P(x \geq 170) = 1.21 \times 10^{-5}$. As in previous analyses, PHM is included on the list of intersected genes and its inclusion could serve as another quality control measure (Park et al. 2011).

DISCUSSION

DIMM gene regulatory network

Scaling factors allow cells to scale subcellular resources (such as organelles) to become specialists in certain physiological functions or to adapt to changing daily conditions (Mills and Taghert 2011). These factors are conserved through evolution and once expressed, their expression persists for the lifetime of the cell (Mills and Taghert 2011). Identifying their mode of operation is an important task since their dysfunction might contribute to human diseases of adult onset (Mills and Taghert 2011). Furthermore, up until now, there has been a focus on understanding transcription factors acting in the early specification of progenitors and lineage development. While the information gleaned from such studies is beneficial to our understanding of how cells are constructed, we must also understand how cells acquire their mature properties. Studying scaling factors such as DIMM provides insight into this process.

Although I could have pursued a variety of biochemical, molecular and genetic approaches to better understand the mechanism of DIMM action, I felt that the best systematic approach was to define the full scope of DIMM targets on a genome-wide level. Thus, I identified sites of DIMM binding in the genome of adult NE cells *in vivo*. This allowed me to pursue identification of genes that DIMM targets for transcriptional activation. With the exception of a recent TFEB study, genome-wide identification of scaling factor direct targets has not been previously reported in literature (Palmieri et al. 2011). The ChIP-chip gene regulatory interactome that I identified in this study is one of the first genome-wide studies to identify direct targets of a scaling factor *in vivo*. The technical approach I pursued allowed for ChIP-chip to be

performed in a highly selected cell population (e.g. directed expression of the DIMM::MYC transgene only in 300 DIMM+ adult NE cells).

In order to focus my attention on DIMM target genes that contribute most highly to DIMM-dependent NE physiology, I decided to use extra-rigorous binding site identification methods. I first compared the DIMM ChIP sample and the negative control sample to their respective input samples. The negative control sample likely contains 'background' binding that is a common artifact of ChIP due to unexpected ''off-target effects'' of the antibodies, non-specific enrichment of chromatin through binding to unrelated immunoglobulins, as well as non-specific binding to agarose beads, reaction surfaces and other reagents (Sandmann et al. 2007). As a first step in data analysis, the coordinates of genomic regions enriched at statistically significant levels in the negative control were subtracted from the genomic regions bound by DIMM. This subtraction resulted in 2,914 regions enriched in the DIMM ChIP/Input sample but not in the NEG ChIP/Input sample (or 6,869 "bits" of regions with no such overlap). It is possible that a significant fraction of these peaks represent DIMM targets. In order to increase analysis stringency further, MAT results from DIMM ChIP tiling arrays were directly compared to results from negative control ChIP tiling arrays. This analysis yielded 258 DIMM-bound regions that were significantly enriched in the DIMM ChIP condition compared to the negative control ChIP. Finally, the intersection of the two sets of regions produced a set of 156 genomic regions that are the most likely enhancers that DIMM binds *in vivo* in adult NE cells.

By performing this rigorous bioinformatic identification of DIMM ChIP-chip binding sites, the certainty of uncovering true DIMM targets was raised at the expense of false negatives. Nevertheless, the estimated number of binding sites for DIMM (156 peaks) closely resembles the estimated number of lysosomal genes that are controlled by TFEB (on the order of 200 genes; Sardiello et al. 2009). It is unlikely that scaling factors, such as DIMM or TFEB control a large fraction of the genome. Furthermore, it is plausible that scaling factors, similar to other

transcription factors, lie at the top of a gene regulatory hierarchy involving multiple transcription factors acting in loops or cascades (Macquarrie et al. 2011). Therefore, a limited number of gene targets (on the order of 200) does not necessarily rule out more genes involved in scaling factor physiology, by means of activation by downstream transcription factors.

The assignment of genes to ChIP-chip peaks (Table 1) is not perfect due to many factors, including: close packing of genes in the genome, the incomplete annotation of transcription units, and the ability of enhancers to act over large distances that sometimes skip intermediate genes (Li et al. 2008). In fact, the c929-GAL4 element that overlaps extensively with DIMM expression is located 13 kilobases upstream of DIMM, in an alternative intron of another gene (*cryptocephal*; Hewes et al. 2003). Despite these caveats, it is believed that most transcription factors bind their gene targets in stereotypically close association to TSSs and long-range enhancers appear to be the exception rather than the rule (Arnosti 2003).

LEAP cell transcriptome

To the best of my knowledge, this high quality data set is the first NE transcriptome obtained by next generation sequencing. Although DIMM-expressing NE cells are highly diverse with respect to their peptidergic identity, they share in common a high secretory capacity and a robust RSP (Park and Taghert 2009). In *Drosophila*, these cells were previously subjected to limited peptidomic profiling (Yew et al. 2009), and a single identified set of DIMM neurons (the eight circadian pacemakers called large LNvs) was profiled transcriptionally (Kula-Eversole et al. 2010; Nagoshi et al. 2010). Nevertheless, there are currently no reports of RNA-Seq-derived transcriptomes of any known NE cell population. Unlike microarrays, deep sequencing allows for a more detailed examination of the transcriptome with the ability to distinguish allelic expression and transcriptional isoforms of genes (Wang et al. 2009). Due to the focus on using RNA-Seq to interpret ChIP-chip data, a detailed analysis of the RNA-Seq data is out of the scope of this chapter but will be pursued in detail later.

NE cells in mammals and insects represent the minority of CNS cells in terms of absolute numbers (Zupanc 1996; Park and Taghert 2009). Despite low absolute numbers, loss of DIMM expression from LEAP cells is not compatible with survival past embryonic stages (Hewes et al. 2003). There are several available methods of profiling gene expression in a cell population with sparse numbers: manual collection of individual cells, laser capture microdissection and FACS (Ozsolak and Milos 2011). While each method has advantages and disadvantages, the primary goal of the current study was to collect a highly purified cell population and maximize the amount of input material. Therefore, FACS was chosen as the method of choice due to its high speed and ability discriminate cellular fluorescence levels easily (Givan 2011).

Interactome – Transcriptome integration

A natural progression in data analysis was to integrate the DIMM gene regulatory interactome obtained by ChIP-chip and the LEAP cell transcriptome obtained by RNA-Seq of purified DIMM-expressing cells. When the list of 156 ChIP-chip regions that were mapped to 284 genes and a total of 598 transcripts was intersected with the list of genes enriched in LEAP cells, 170 transcripts (116 genes) fell into both categories. This likely represents a high quality, stringent list of direct DIMM targets. The observed overlap was greater than overlap due to chance, as shown by the cumulative hypergeometric probability. Thus, genes on this cross-sectional list are not only direct DIMM targets in LEAP cells by ChIP-chip, but their expression is enriched in LEAP cells compared to DIMM-negative cells.

The caveat to this intersectional assignment is that it assumes that DIMM acts as a transcriptional activator in all cases. This assumption is based on previous experiments that demonstrated that DIMM acts as an activator of PHM transcription (Park et al. 2008b). Nevertheless, certain bHLH proteins are known to act as repressors (Stevens et al 2008). Such repressive bHLHs were first identified in *Drosophila* (Stevens et al. 2008). In accordance with

this possibility, 50 out of 598 ChIP-chip associated transcripts are expressed in DIMM+/GFP+ cells at 50% of their level in the DIMM-/GFP- cells (and only 3 out of 50 are found at 10% of their level in DIMM-/GFP- cells). In contrast, 170 out of 598 ChIP-chip transcripts are enriched in DIMM+/GFP+ cells over DIMM-/GFP- cells, which gives credibility to the hypothesis that DIMM operates as a transcriptional activator in all cases. A precise answer to this question can be obtained by pursuing RNA-Seq of DIMM loss-of-function cells, which is one of the planned follow-up studies.

DIMM, like many transcription factors acts as an activator of transcription (Allan et al. 2005). By this logic, if a DIMM binding site is located in a dense milieu of TSSs and only one of the surrounding genes is enriched in LEAP cells, then, this increase the likelihood that DIMM is activating this particular gene and not its neighbors. Overall, transcriptome/interactome integration reduces the false positives in ChIP-chip gene regulatory data due to a reduction in gene assignment noise (Wyrick and Young 2002).

Classifying DIMM target genes into functional categories

The integrated data set consists of a variety of genes with certain subsets clustering in a few categories (Table 4). One such category are neuropeptide processing enzymes: the peptidylglycine hydroxylating alpha-monooxygenase *PHM*, the peptidyl-alpha-hydroxyglycine alpha-amidating lyase *PAL1*, the prohormone convertase *amontillado*, and the carboxypeptidase D *silver*. All these enzymes reside in LDCVs, where they are indispensible for the proper processing of neuropeptides. Their appearance on this high quality list confirms and extends the supposition that DIMM targets peptide biosynthetic enzymes (Park et al. 2008). It also provides important “quality control” assurance that the procedure is generating targets expected to be on the list. *Slamdance* is a *Drosophila* aminopeptidase with a bang-sensitive paralytic mutant phenotype (Zhang et al. 2002). Human aminopeptidase N, the mammalian orthologue of *slamdance*, is in involved neuropeptide processing and degradation (Montiel et al.

1997; Solhonne et al. 1987). Thus, the inclusion of *slamdance* as a likely DIMM target represents the first indication that it could be involved in neuropeptide processing in *Drosophila* as well as mammals.

Slamdance also establishes a link between neuropeptide signaling and epilepsy, which is a relatively new concept with a seminal paper reporting a knockout mutation of neuropeptide Y leading to epileptic phenotypes in the mouse (Baraban et al. 1997). Furthermore, expression of a PHM / Pal1 and a number of amidated neuropeptides increases after seizures induced by electroconvulsive shock, kainate, or pentylenetetrazole (Ma et al. 2002; Bhat et al. 2003; Baraban et al. 1997). Additional genes that fall into the endopeptidase/metallopeptidase category with possible role in neuropeptide processing or degradation are: *CG4933*, *Rpn9* and *Atg4*. This work shows that a whole battery of secretory processing enzymes is under transcriptional control of a single transcription factor, DIMM. Although the fact that DIMM controls processing enzymes may not be a surprising, it is, nevertheless, a completely novel finding. No single transcription factor has been revealed to possess the ability to control all or most neuropeptide processing factors.

Another category of genes under direct DIMM control includes fly orthologues of those mammalian genes whose protein products have been shown by proteomic studies to be constituents of LDCVs (Table 4; Gauthier et al. 2008; Wegrzyn et al. 2010). Genes in this category include: *14-3-3zeta*, *14-3-3epsilon*, *jaguar*, *Translationally controlled tumor protein* and *betaTubulin56D*. The two members of the 14-3-3 family of adaptor proteins are a highly conserved family of acidic molecules present in all eukaryotes (Aitken 2006). Both *14-3-3zeta* and *14-3-3epsilon* have been implicated in the response of the hypothalamic-neurohypophyseal system to osmotic stress (Gouraud et al. 2007). The expression of *14-3-3zeta* and *14-3-3epsilon* is increased in the rat supraoptic nucleus, a NE center, after three days of water deprivation. Furthermore, *14-3-3epsilon* has a crucial role in antimicrobial peptide secretion

(Shandala et al. 2011). *Jaguar*, a *Drosophila* myosin VI homologue is another gene with LDCV localization and known role in secretory vesicle fusion at the plasma membrane (Bond et al. 2011). *Jaguar* is a unique member of the myosin superfamily of actin-based motor proteins because it is the only myosin known to move towards the minus or pointed ends of actin filaments (Kisiel et al. 2011). *Drosophila* larvae lacking jaguar have altered synaptic vesicle localization: instead of peripherally located synaptic vesicles, *jaguar* mutants have diffusely located vesicle across the entire bouton area (Kisiel et al. 2011). After budding from the trans-Golgi network, LDCVs are transported to the secretion sites at the plasma membrane via microtubule-based transport systems (Park and Loh 2008). LDCVs are then loaded onto an actin/myosin system for distal transport through the actin cortex to just below the plasma membrane (Park and Loh 2008). Therefore, it is plausible that DIMM would activate the expression of those genes necessarily for LDCVs to properly traffic to their designated location at the cell periphery where they await release signals.

Transcription. In addition to secretory enzymes and LDCV constituents, DIMM also targets transcription factors, in particular members of the Atf/CREB superfamily of basic leucine zipper (bZIP) genes: the Drosophila Atf-4 orthologue *cryptocephal* (significant for being immediately adjacent to DIMM), the pro-secretory transcription factor *CrebA*, the stress-response gene *Atf-2*, as well as the circadian-implicated transcriptional repressor *vrille.* bZIP proteins contain a leucine zipper, an alpha-helical coil structure required for dimerization, as well as a neighboring basic domain needed for direct contact with DNA (Foulkes et al. 1997). Two non-bZIP transcription factors are the circadian regulator *clockwork orange*, as well as the Per/Arnt/Sim-domain containing bHLH *similar*. *Cryptocephal* (Atf-4) is important for metamorphosis and molting in *Drosophila*, and in oxidative stress, amino acid synthesis, autophagy, unfolded protein response and long-term memory in mammals (Hewes et al. 2000; Ameri and Harris 2008). Mammalian *Atf-2* and *Atf-4* regulate emotional behavior in the nucleus

accumbens by unknown mechanisms (Green et al. 2008). It is currently unknown how *cryptocephal* might play a role in the secretory capacity of NE cells.

CrebA is major and direct regulator of secretory capacity of a variety of cell types at the level of the Golgi apparatus (Fox et al. 2010). *CrebA* can upregulate expression of the general protein machinery required in all cells for secretion, as well as certain secreted proteins that are cell-type specific (Fox et al. 2010). In *Drosophila*, *CrebA* is important for elevated secretory activity in the salivary gland, as well as in the developing cuticle (Abrams and Andrew 2005). Thus *CrebA* is adaptable, in different cellular contexts, to promote a high capacity secretory state. *CrebA* appears to activate expression of genes encoding proteins required for ER targeting and translocation, and proteins that mediate transport between the ER and Golgi (Abrams and Andrew 2005). Thus, similar to other secretory cell types, it is plausible that NE cells need high levels of *CrebA*, which is recruited to the purpose by DIMM as a part of its general action mechanisms. Interestingly, there is very little overlap between genes thought to be *CrebA* targets and DIMM target genes that I have identified.

Furthermore, it may be notable that DIMM activates two factors that are known for the their potent ability to antagonize bHLH actions: *vrille* and *clockwork orange*. *Vrille* is a bZIP transcriptional repressor that acts by inhibiting expression of the core circadian bHLH gene *Clock*, as well as the circadian blue light photoreceptor *cryptochrome* (Glossop et al. 2003). As a part of the circadian system, VRILLE competes with a related bZIP protein PDP1 for binding to the *Clock* promoter. Whereas VRILLE represses *Clock* expression, PDP1 activates expression of *Clock* and thereby actively competes with *vrille* (Cyran et al. 2003). In addition to *vrille*, DIMM targets another repressive factor, the basic helix–loop–helix ORANGE family member *clockwork orange. Clockwork orange* forms its own negative feedback loop and directly suppresses the expression of other clock genes through the E-box sequences (Matsumoto et al. 2007; Kadener et al. 2007). It is possible that by targeting *vrille* and *clockwork orange*, DIMM uses complex

gene regulatory logic, beyond simple transcriptional activation. According to this line of thinking, the genetic logic of the mechanisms activated by the bHLH DIMM is similar that of the molecular clockwork instigated by the bHLH proteins *Clock* and *cycle*: both may use feedforward and / or feedback negative regulation by *vrille* and *clockwork orange* to modulate the amplitudes of the target gene expression.

Endocytosis. Genes involved in endocytosis and Notch signaling represent another category of DIMM targets: *nicastrin*, *Gp150*, *I'm not dead yet*, *TRAM* and *CG31064*. *Nicastrin* is a part of the gamma-secretase core complex required for cleavage of transmembrane proteins such as *Notch* and *presenillin* (Hass et al. 2009; Steiner et al. 2008). *Gp150* is an endosomal protein that regulates Notch activity (Li et al. 2004). Since the Notch protein itself is high in LEAP cells, it is possible that *Gp150* and *nicastrin* are performing a role in DIMM-dependent NE cell physiology in a manner that also involves *Notch* signaling in mature cells.

RNA-binding proteins. A separate family of DIMM targets includes genes that are implicated in RNA processing or metabolism: *alan shepard, ELL-associated factor, Ef1α48D, Not1, U4-U6-60K* and *eIF-4A*. *Alan shepard* is a single-stranded RNA-binding protein isolated in a genetic screen for gravitaxis mutants (Bjorum and Beckingham 2007). RNA-binding proteins such as *alan shepard* engage in pre-mRNA processing, which includes splicing, editing and polyadenylation of messages (Glisovic et al. 2008). At the structural level, RNA-binding proteins are highly modular, with one or more RNA-binding and auxiliary domains (Glisovic et al. 2008). DIMM binding frequently occurs in areas that are promoters of certain isoforms and introns of other isoforms of the same RNA-binding protein. Furthermore, RNA-Seq data show that only certain isoforms of RNA-binding proteins are enriched in LEAP cells by RNA-Seq, whereas other isoforms of the same gene are not enriched, or could even be repressed. Therefore, it is possible that DIMM targets certain isoforms of RNA-binding proteins with desired sequence-binding specificities for expression in LEAP cells.

RNA-binding proteins function in every aspect of RNA biology, from transcription, pre-mRNA splicing and polyadenylation to RNA modification, transport, localization, translation and turnover (Glisovic et al. 2008). Therefore, it is possible that RNA-binding proteins that DIMM targets play a major role in the ability of LEAP cells to correctly traffic or localize neuropeptides. Certain RNA-binding proteins are known to be potent repressors of translation of certain sequence-specific RNA targets (Horisawa et al. 2009). Since neuropeptides are translated in very large quantities inside NE cells, it is plausible that translation of other RNA messages needs to be reduced or halted from time to time in order to achieve necessary levels of neuropeptide translation. RNA-binding proteins could play a role of in translation of such RNA messages that are perhaps of secondary importance during particular physiologic states. Alternatively, the splicing of neuropeptide RNAs or their stability might need to be enhanced, therefore requiring RNA-binding proteins to carry this function out with high selectivity.

Before this work was initiated, Hamanaka et al. (2010) investigated what happened to photoreceptors, a class of conventional, non-peptidergic neurons, when they are forced to express DIMM. This work showed DIMM's remarkable ability to convert non-peptidergic cells into peptidergic cells possessing LDCVs, the morphologic correlate of the RSP. Therefore, we had a great insight into the ultrastructural changes that DIMM can set off once activated ectopically in a cell. What was lacking, however, was the molecular framework of how this remarkable cellular transformation is achieved. Collectively, therefore the results presented in this chapter provide a fundamental road map to define exactly how DIMM is able to organize, scale up and maintain an active RSP in NE cells.

REFERENCES

Abrams, E.W., and Andrew, D.J. (2005). CrebA regulates secretory activity in the Drosophila salivary gland and epidermis. *Development* **132**, 2743-2758.

Abruzzi, K.C., Rodriguez, J., Menet, J.S., Desrochers, J., Zadina, A., Luo, W., Tkachev, S., and Rosbash, M. (2011). *Drosophila* CLOCK target gene characterization: implications for circadian tissue-specific gene expression. *Genes Dev.* **25**, 2374-2386.

Ameri, K., and Harris, A.L. (2008). Activating transcription factor 4. *Int. J Biochem. & Cell Biol.* **40**, 14-21.

Aitken, A. (2006). 14-3-3 proteins: A historic overview. *Semin. Cancer Biol.* **16**, 162-172.

Allan, D.W., Park, D., St Pierre, S.E., Taghert, P.H., and Thor, S. (2005). Regulators acting in combinatorial codes also act independently in single differentiating neurons. *Neuron* **45**, 689-700.

Arnosti, D.N. (2003). Analysis and function of transcriptional regulatory elements: insights from Drosophila. *Annu. Rev. Entomol* **48**, 579-602.

Baraban, S.C., Hollopeter, G., Erickson, J.C., Schwartzkroin, P.A., and Palmiter, R.D. (1997). Knock-out mice reveal a critical antiepileptic role for neuropeptide Y. *J. Neurosci.* **17**, 8927-8936.

Bhat, R.V., Tausk, F.A., Baraban, J.M., Mains, R.E., and Eipper, B.A. (1993). Rapid increases in peptide processing enzyme expression in hippocampal neurons. *J. Neurochem.* **61**, 1315-1322.

Bjorum, S., Beckingham, K.M. (2007). Two Genes Affecting Drosophila Gravitaxis. A. Dros. Res. Conf. 48 : 638B.

Blankenberg, D., Taylor, J., Nekrutenko, A., and Galaxy, T. (2011). Making whole genome multiple alignments usable for biologists. *Bioinformatics* **27**, 2426-2428.

Bond, L.M., Peden, A.A., Kendrick-Jones, J., Sellers, J.R., and Buss, F. (2011). Myosin VI and its binding partner optineurin are involved in secretory vesicle fusion at the plasma membrane. *Mol. Biol. Cell* **22**, 54-65.

Borghese, L., Fletcher, G., Mathieu, J., Atzberger, A., Eades, W.C., Cagan, R.L., and Rørth, P. (2006). Systematic analysis of the transcriptional switch inducing migration of border cells. *Dev. Cell* **10**, 497-508.

Brand, A.H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-415.

Brogiolo, W., Stocker, H., Ikeya, T., Rintelen, F., Fernandez, R., and Hafen, E. (2001). An evolutionarily conserved function of the *Drosophila* insulin receptor and insulin-like peptides in growth control. *Curr. Biol.* **11**, 213-221.

Burbach, J.P., Luckman, S.M., Murphy, D., and Gainer, H. (2001). Gene regulation in the magnocellular

hypothalamo-neurohypophysial system. *Physiol. Rev.* **81**, 1197-1267.

Cao, Y., Yao, Z., Sarkar, D., Lawrence, M., Sanchez, G.J., Parker, M.H., Macquarrie, K.L., Davison, J., Morgan, M.T., Ruzzo, W.L.*, et al.* (2010). Genome-wide MyoD binding in skeletal muscle cells: a potential for broad cellular reprogramming. *Dev. Cell* **18**, 662-674.

Crivellato, E., Nico, B., Gallo, V.P., and Ribatti, D. (2010). Cell secretion mediated by granule-associated vesicle transport: a glimpse at evolution. *Anat. Rec. (Hoboken)* **293**, 1115-1124.

Cyran, S. A., Buchsbaum, A. M., Reddy, K. L., Lin, M.-C., Glossop, N. R. J., Hardin, P. E., Young, M. W., Storti, R. V., and Blau, J. (2003). vrille, Pdp1, and dClock form a second feedback loop in the Drosophila circadian clock. *Cell* **112**, 329–341.

Dafforn, A., Chen, P., Deng, G., Herrler, M., Iglehart, D., Koritala, S., Lato, S., Pillarisetty, S., Purohit, R., Wang, M.*, et al.* (2004). Linear mRNA amplification from as little as 5 ng total RNA for global gene expression analysis. *BioTechniques* **37**, 854-857.

Dannies, P. S. (1999) 'Protein hormone storage in secretory granules: mechanisms for concentration and sorting', *Endocr. Rev.* **20**(1): 3-21.

De Camilli, P., and Jahn, R. (1990). Pathways to regulated exocytosis in neurons. *Annu. Rev. Physiol.* **52**, 625-645.

Dougherty, J.D., Schmidt, E.F., Nakajima, M., and Heintz, N. (2010). Analytical approaches to RNA profiling data for the identification of genes enriched in specific cells. *Nucleic Acids Res.* **38**, 4218-4230.

Doyle, J.P., Dougherty, J.D., Heiman, M., Schmidt, E.F., Stevens, T.R., Ma, G., Bupp, S., Shrestha, P., Shah, R.D., Doughty, M.L.*, et al.* (2008). Application of a Translational Profiling Approach for the Comparative Analysis of CNS Cell Types. *Cell* **135**, 749-762.

Droit, A., Cheung, C., and Gottardo, R. (2010). rMAT--an R/Bioconductor package for analyzing ChIP-chip experiments. *Bioinformatics* **26**, 678-679.

Ezhkova, E., and Tansey, W.P. (2006). Chromatin immunoprecipitation to study protein-DNA interactions in budding yeast. *Methods Mol. Biol.* **313**, 225-244.

Finck, B.N., and Kelly, D.P. (2007). Peroxisome proliferator-activated receptor gamma coactivator-1 (PGC-1) regulatory cascade in cardiac physiology and disease. *Circulation* **115**, 2540-2548.

Flames, N., and Hobert, O. (2009). Gene regulatory logic of dopamine neuron differentiation. *Nature* **458**, 885-889.

Foulkes, N.S., Borjigin, J., Snyder, S.H., and Sassone-Corsi, P. (1997). Rhythmic transcription: the molecular basis of circadian melatonin synthesis. *Trends Neurosci.* **20**, 487–492.

Fox, R.M., Hanlon, C.D., and Andrew, D.J. (2010). The CrebA/Creb3-like transcription factors are major and direct regulators of secretory capacity. *J. Cell Biol.* **191**, 479-492.

Gauthier, D.J., Sobota, J.A., Ferraro, F., Mains, R.E., and Lazure, C. (2008). Flow cytometry-assisted

purification and proteomic analysis of the corticotropes dense-core secretory granules. *Proteomics* **8**, 3848-3861.

Givan, A.L. (2011). Flow cytometry: an introduction. *Methods Mol. Biol.* **699**, 1-29.

Glisovic, T., Bachorik, J., Yong, J., and Dreyfuss, G. (2008). RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* **582**, 1977–1986.

Glossop, N. R. J., Houl, J. H., Zheng, H., Ng, F. S., Dudek, S. M., and Hardin, P. E. (2003). VRILLE feeds back to control circadian transcription of Clock in the Drosophila circadian oscillator. *Neuron* **37**, 249–261.

Goecks, J., Nekrutenko, A., Taylor, J., and Team, G. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86.

Gouraud, S.S., Yao, S.T., Heesom, K.J., Paton, J.F.R., and Murphy, D. (2007). 14-3-3 proteins within the hypothalamic-neurohypophyseal system of the osmotically stressed rat: transcriptomic and proteomic studies. *J Neuroendocrinol.* **19**, 913-922.

Green, T.A., Alibhai, I.N., Unterberg, S., Neve, R.L., Ghose, S., Tamminga, C.A., and Nestler, E.J. (2008). Induction of activating transcription factors (ATFs) ATF2, ATF3, and ATF4 in the nucleus accumbens and their regulation of emotional behavior. *J. Neurosci.* **28**, 2025-2032.

Hamanaka, Y., Park, D., Yin, P., Annangudi, S.P., Edwards, T.N., Sweedler, J., Meinertzhagen, I.A., and Taghert, P.H. (2010). Transcriptional orchestration of the regulated secretory pathway in neurons by the bHLH protein DIMM. *Curr. Biol.* **20**, 9-18.

Han, H., Nutiu, R., Moffat, J., and Blencowe, B.J. (2011). SnapShot: High-Throughput Sequencing Applications. *Cell* **146**, 1044-1044.e1042.

Hass, M.R., Sato, C., Kopan, R., and Zhao, G.J. (2009). Presenilin: RIP and beyond. *Semin. Cell Dev. Biol.* **20**, 201-210.

He, H., Meyer, C.A., Shin, H.J., Bailey, S.T., Wei, G., Wang, Q.B., Zhang, Y., Xu, K.X., Ni, M., Lupien, M.*, et al.* (2010). Nucleosome dynamics define transcriptional enhancers. *Nat. Genet.* **42**, 343-U101.

Hendricks, T., Francis, N., Fyodorov, D., and Deneris, E.S. (1999). The ETS domain factor Pet-1 is an early and precise marker of central serotonin neurons and interacts with a conserved element in serotonergic genes. *J. Neurosci.* **19**, 10348-10356.

Hewes, R.S., Schaefer, A.M., and Taghert, P.H. (2000). The cryptocephal gene (ATF4) encodes multiple basic-leucine zipper proteins controlling molting and metamorphosis in Drosophila. *Genetics* **155**, 1711-1723.

Hewes, R.S., Park, D., Gauthier, S.A., Schaefer, A.M., and Taghert, P.H. (2003). The bHLH protein Dimmed controls neuroendocrine cell differentiation in *Drosophila*. *Development* **130**, 1771-1781.

Hindmarch, C., Yao, S., Beighton, G., Paton, J., and Murphy, D. (2006). A comprehensive description

of the transcriptome of the hypothalamoneurohypophyseal system in euhydrated and dehydrated rats. *Proc. Natl. Acad. Sci. USA* **103**, 1609-1614.

Horisawa, K., Imai, T., Okano, H., and Yanagawa, H. (2009). 3 '-Untranslated region of doublecortin mRNA is a binding target of the Musashi1 RNA-binding protein. *Febs Letters* **583**, 2429-2434.

Jiang, S.A., Campusano, J.M., Su, H., and O'dowd, D.K. (2005). *Drosophila* mushroom body Kenyon cells generate spontaneous calcium transients mediated by PLTX-sensitive calcium channels. *J. Neurophysiol.* **94**, 491-500.

Johnson, C.L., Kowalik, A.S., Rajakumar, N., and Pin, C.L. (2004). Mist1 is necessary for the establishment of granule organization in serous exocrine cells of the gastrointestinal tract. *Mech. Dev.* **121**, 261-272.

Johnson, W., Li, W., Meyer, C., Gottardo, R., Carroll, J., Brown, M., and Liu, X. (2006). Model-based analysis of tiling-arrays for ChIP-chip. *Proc. Natl. Acad. Sci. USA* **103**, 12457.

Kadener, S., Stoleru, D., McDonald, M., Nawathean, P., and Rosbash, M. (2007). Clockwork Orange is a transcriptional repressor and a new Drosophila circadian pacemaker component. *Genes Dev.* **21**, 1675–1686.

Kim, T., Gondré-Lewis, M.C., Arnaoutova, I., and Loh, Y.P. (2006). Dense-core secretory granule biogenesis. *Physiology (Bethesda)* **21**, 124-133.

Kisiel, M., Majumdar, D., Campbell, S., and Stewart, B.A. (2011). Myosin VI contributes to synaptic transmission and development at the Drosophila neuromuscular junction. *BMC. Neurosci.* **12**.

Kolodziej, K.E., Pourfarzad, F., de Boer, E., Krpic, S., Grosveld, F., and Strouboulis, J. (2009). Optimal use of tandem biotin and V5 tags in ChIP assays. *BMC Mol. Biol.* **10**, 6.

Kramer, J.M., Kochinke, K., Oortveld, M.A.W., Marks, H., Kramer, D., de Jong, E.K., Asztalos, Z., Westwood, J.T., Stunnenberg, H.G., Sokolowski, M.B.*, et al.* (2011). Epigenetic Regulation of Learning and Memory by *Drosophila* EHMT/G9a. *Plos Biol.* **9**.

Kula-Eversole, E., Nagoshi, E., Shang, Y., Rodriguez, J., Allada, R., and Rosbash, M. (2010). Surprising gene expression patterns within and between PDF-containing circadian neurons in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **107**, 13497-13502.

Küppers-Munther, B., Letzkus, J.J., Lüer, K., Technau, G., Schmidt, H., and Prokop, A. (2004). A new culturing strategy optimises *Drosophila* primary cell cultures for structural and functional analyses. *Dev. Biol.* **269**, 459-478.

Kurn, N., Chen, P., Heath, J.D., Kopf-Sill, A., Stephens, K.M., and Wang, S. (2005). Novel isothermal, linear nucleic acid amplification systems for highly multiplexed applications. *Clin. Chem.* **51**, 1973-1981.

Lee, T.I., Johnstone, S.E., and Young, R.A. (2006). Chromatin immunoprecipitation and microarray-based analysis of protein location. *Nat. Protoc.* **1**, 729-748.

Lehman, J.J., Barger, P.M., Kovacs, A., Saffitz, J.E., Medeiros, D.M., and Kelly, D.P. (2000).

Peroxisome proliferator-activated receptor gamma coactivator-1 promotes cardiac mitochondrial biogenesis. *J. Clin. Invest.* **106**, 847-856.

Li, Y.X., Fetchko, M., Lai, Z.C., and Baker, N.E. (2004). Scabrous and Gp150 are endosomal proteins that regulate notch activity (vol 130 pg 2819, 2004). *Development* **131**, 5213-5213.

Li, X.-Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D.A., Iyer, V.N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C.L.*, et al.* (2008). Transcription factors bind thousands of active and inactive regions in the *Drosophila* blastoderm. *PLoS Biol.* **6**, e27.

Liu, Y., Mziaut, H., Ivanova, A., and Solimena, M. (2009). beta-Cells at the crossroads: choosing between insulin granule production and proliferation. *Diabetes Obes. Metab.* **11 Suppl 4**, 54-64.

Liu, T., Ortiz, J.A., Taing, L., Meyer, C.A., Lee, B., Zhang, Y., Shin, H., Wong, S.S., Ma, J., Lei, Y.*, et al.* (2011). Cistrome: an integrative platform for transcriptional regulation studies. *Genome Biol.* **12**.

Ma, X.M., Mains, R.E., and Eipper, B.A. (2002). Plasticity in hippocampal peptidergic systems induced by repeated electroconvulsive shock. *Neuropsychopharmacology* **27**, 55-71.

Matsumoto, A., Ukai-Tadenuma, M., Yamada, R. G., Houl, J., Uno, K. D., Kasukawa, T., Dauwalder, B., Itoh, T. Q., Takahashi, K., Ueda, R., et al. (2007). A functional genomics strategy reveals clockwork orange as a transcriptional regulator in the Drosophila circadian clock. *Genes Dev.* **21**, 1687–1700.

Macquarrie, K.L., Fong, A.P., Morse, R.H., and Tapscott, S.J. (2011). Genome-wide transcription factor binding: beyond direct target regulation. *Trends Genet.* **27**, 141-148.

McGuire, S.E., Le, P.T., Osborn, A.J., Matsumoto, K., and Davis, R.L. (2003). Spatiotemporal rescue of memory dysfunction in *Drosophila*. *Science* **302**, 1765-1768.

Menet, J.S., Abruzzi, K.C., Desrochers, J., Rodriguez, J., and Rosbash, M. (2010). Dynamic PER repression mechanisms in the *Drosophila* circadian clock: from on-DNA to off-DNA. *Genes Dev.* **24**, 358-367.

Mills, J.C., and Taghert, P.H. (2011). Scaling factors: Transcription factors regulating subcellular domains. *Bioessays*.

Montiel, J.L., Cornille, F., Roques, B.P., and Noble, F. (1997). Nociceptin/orphanin FQ metabolism: Role of aminopeptidase and endopeptidase 24.15. *J. Neurochem.* **68**, 354-361.

Moore, A.W., Barbel, S., Jan, L.Y., and Jan, Y.N. (2000). A genomewide survey of basic helix-loop-helix factors in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **97**, 10436-10441.

Nagoshi, E., Sugino, K., Kula, E., Okazaki, E., Tachibana, T., Nelson, S., and Rosbash, M. (2010). Dissecting differential gene expression within the circadian neuronal circuit of *Drosophila*. *Nat. Neurosci.* **13**, 60-68.

Nitabach, M.N., and Taghert, P.H. (2008). Organization of the *Drosophila* circadian control circuit. *Curr. Biol.* **18**, R84-93.

Orlando, V. (2000). Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. *Trends Biochem. Sci.* **25**, 99-104.

Ozsolak, F., and Milos, P.M. (2011). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* **12**, 87-98.

Palmieri, M., Impey, S., Kang, H.J., di Ronza, A., Pelz, C., Sardiello, M., and Ballabio, A. (2011). Characterization of the CLEAR network reveals an integrated control of cellular clearance pathways. *Hum. Mol. Genet.* **20**, 3852-3866.

Park, J.J., and Loh, Y.P. (2008). How peptide hormone vesicles are transported to the secretion site for exocytosis. *Mol. Endocrinol.* **22**, 2583-2595.

Park, D., Veenstra, J.A., Park, J.H., and Taghert, P.H. (2008a). Mapping peptidergic cells in *Drosophila*: where DIMM fits in. *PLoS ONE* **3**, e1896.

Park, D., Shafer, O.T., Shepherd, S.P., Suh, H., Trigg, J.S., and Taghert, P.H. (2008b). The *Drosophila* basic helix-loop-helix protein DIMMED directly activates PHM, a gene encoding a neuropeptide-amidating enzyme. *Mol. Cell. Biol.* **28**, 410-421.

Park, D., and Taghert, P.H. (2009). Peptidergic neurosecretory cells in insects: organization and control by the bHLH protein DIMMED. *Gen. Comp. Endocrinol.* **162**, 2-7.

Park, D., Hadzic, T., Yin, P., Rusch, J., Abruzzi, K., Rosbash, M., Skeath, J.B., Panda, S., Sweedler, J.V., and Taghert, P.H. (2011). Molecular organization of *Drosophila* neuroendocrine cells by dimmed. *Curr. Biol*. **21**:1515-24

Pin, C.L., Bonvissuto, A.C., Konieczny, S.F. (2000). Mist1 expression is a common link among serous exocrine cells exhibiting regulated exocytosis. *Anat. Rec.* **259**, 157–167.

Ramsey, V.G., Doherty, J.M., Chen, C.C., Stappenbeck, T.S., Konieczny, S.F., and Mills, J.C. (2007). The maturation of mucus-secreting gastric epithelial progenitors into digestive-enzyme secreting zymogenic cells requires Mist1. *Development* **134**, 211-222.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306.

Roberts, A., Pimentel, H., Trapnell, C., and Pachter, L. (2011). Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics* **27**, 2325-2329.

Robinson, J.T., Thorvaldsdottir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24-26.

Sandmann, T., Jakobsen, J., and Furlong, E. (2007). ChIP-on-chip protocol for genome-wide analysis of transcription factor binding in Drosophila melanogaster embryos. *Nat. Protoc* **1**, 2839-2855.

Saeed, A.I., Bhagabati, N.K., Braisted, J.C., Liang, W., Sharov, V., Howe, E.A., Li, J.W., Thiagarajan, M., White, J.A., and Quackenbush, J. (2006). TM4 microarray software suite. *Methods Enzymol.* **411**,

134-+.

Sardiello, M., Palmieri, M., di Ronza, A., Medina, D.L., Valenza, M., Gennarino, V.A., Di Malta, C., Donaudy, F., Embrione, V., Polishchuk, R.S.*, et al.* (2009). A gene network regulating lysosomal biogenesis and function. *Science* **325**, 473-477.

Scharrer, B. (1987). Neurosecretion: beginnings and new directions in neuropeptide research. *Annu. Rev. Neurosci.* **10**, 1-17.

Schmid, I., Hausner, M.A., Cole, S.W., Uittenbogaart, C.H., Giorgi, J.V., and Jamieson, B.D. (2001). Simultaneous flow cytometric measurement of viability and lymphocyte subset proliferation. *J. Immunol. Methods* **247**, 175-186.

Shandala, T., Woodcock, J.M., Ng, Y., Biggs, L., Skoulakis, E.M.C., Brooks, D.A., and Lopez, A.F. (2011). Drosophila 14-3-3ε has a crucial role in anti-microbial peptide secretion and innate immunity. *J. Cell Sci.* **124**, 2165-2174.

Simon, I., Barnett, J., Hannett, N., Harbison, C.T., Rinaldi, N.J., Volkert, T.L., Wyrick, J.J., Zeitlinger, J., Gifford, D.K., Jaakkola, T.S., et al. (2001). Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* **106**, 697-708.

Solhonne, B., Gros, C., Pollard, H., and Schwartz, J.C. (1987). Major localization of aminopeptidase M in rat brain microvessels. *Neuroscience* **22**, 225-232.

Southall, T.D., and Brand, A.H. (2007). Chromatin profiling in model organisms. *Brief. Funct. Genomic. Proteomic.* **6**, 133-140.

Steiner, H., Fluhrer, R., and Haass, C. (2008). Intramembrane Proteolysis by gamma-Secretase. *J. Biol. Chem.* **283**, 29627-29631.

Südhof, T.C. (2007). Neurotransmitter Release. *Handb. Exp. Pharmacol.* **184**, 1-21.

Stevens, J.D., Roalson, E.H., and Skinner, M.K. (2008). Phylogenetic and expression analysis of the basic helix-loop-helix transcription factor gene family: genomic approach to cellular differentiation. *Differentiation* **76**, 1006-1022.

Tian, X., Jin, R.U., Bredemeyer, A.J., Oates, E.J., Błazewska, K.M., McKenna, C.E., and Mills, J.C. (2010). RAB26 and RAB3D are direct transcriptional targets of MIST1 that regulate exocrine granule maturation. *Mol. Cell. Biol.* **30**, 1269-1284.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111.

Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* **10**, 57-63.

Wegrzyn, J.L., Bark, S.J., Funkelstein, L., Mosier, C., Yap, A., Kazemi-Esfarjani, P., La Spada, A.R., Sigurdson, C., O'Connor, D.T., and Hook, V. (2010). Proteomics of dense core secretory vesicles reveal distinct protein categories for secretion of neuroeffectors for cell-cell communication. *J.*

Proteome Res. **9**, 5002-5024.

Wyrick, J.J., and Young, R.A. (2002). Deciphering gene expression regulatory networks. *Curr. Opin. Genet. Dev.* **12**, 130-136.

Yue, C., Mutsuga, N., Verbalis, J., and Gainer, H. (2006). Microarray analysis of gene expression in the supraoptic nucleus of normoosmotic and hypoosmotic rats. *Cell. Mol. Neurobiol.* **26**, 959-978.

Yew, J.Y., Wang, Y., Barteneva, N., Dikler, S., Kutz-Naber, K.K., Li, L., and Kravitz, E.A. (2009). Analysis of neuropeptide expression and localization in adult *Drosophila melanogaster* central nervous system by affinity cell-capture mass spectrometry. *J. Proteome Res.* **8**, 1271-1284.

Zhang, H., Tan, J., Reynolds, E., Kuebler, D., Faulhaber, S., and Tanouye, M. (2002). The Drosophila slamdance gene: a mutation in an aminopeptidase can cause seizure, paralysis and neuronal failure. *Genetics* **162**, 1283-1299.

Zhu, L.J., Gazin, C., Lawson, N.D., Pages, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC. Bioinformatics* **11**.

Zupanc, G.K. (1996). Peptidergic transmission: from morphological correlates to functional implications. *Micron* **27**, 35-91.

ACKNOWLEDGEMENTS

I would like to thank the following members of the Taghert lab for helping out with brain dissections: Paul Taghert, Dongkook Park, Laura Duvall and SeolHee Im. I would also like to thank Mr. Chris Holley and Bill Eades, of the Siteman Flow Cytometry Core for providing excellent technical assistance with FACS sorting. I would also like to thank Dayna Oschwald and Mark Johnson of the Genome Technology Access Center at Washington University for processing samples for deep sequencing and analyzing data. This work was funded by NIH P01 NS044232, P30-NS045713 and the Howard Hughes Medical Institute (to MR), and NINDS grant NS21749 (to PHT). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We also thank the Bloomington Stock Center and the VDRC for flies.

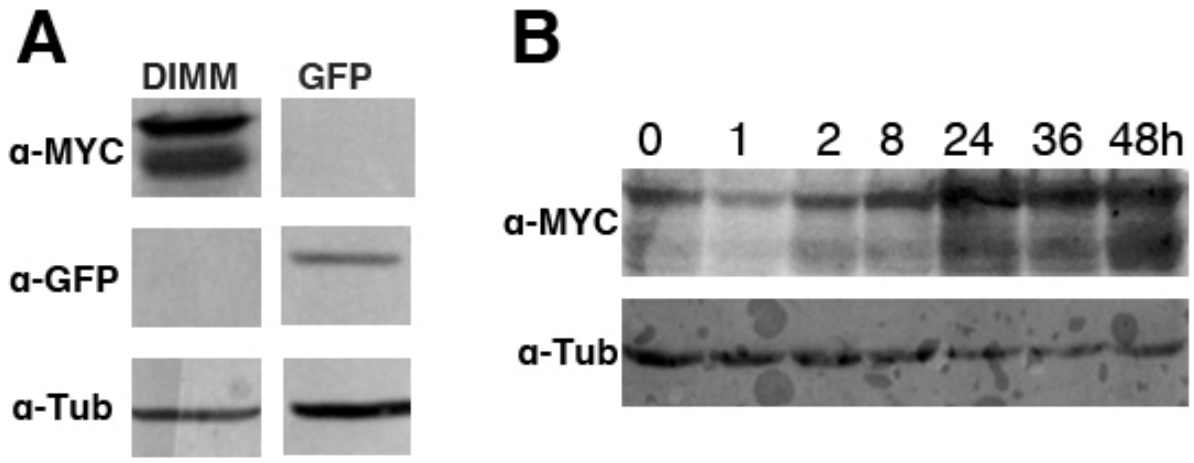

Figure 1. Detection of DIMM::MYC induction in adult heads by Western blotting. (A) Flies overexpressing DIMM::MYC (left panel, labeled “DIMM”) or GFP (right panel, labeled “GFP”) in DIMM+ (c929-GAL4-expressing cells) developed normally at the restrictive temperature (18°C). Following eclosion, both genotypes were shifted to 30°C for 3-5 days. This allowed for a specific induction of DIMM::MYC expression in DIMM::MYC-overexpressing heads (top left panel) and lack thereof in GFP-overexpressing heads (top right panel). In contrast to DIMM::MYC induction, GFP was induced in GFP-overexpressing heads (middle right panel) but not in DIMM::MYC-expressing heads (middle left panel). Adult head extracts were analyzed by Western blotting with indicated antibodies. Tubulin was used as a loading control (bottom). Genotype: yw/w; c929-GAL4/UAS-DIMM::MYC; tub-GAL80ts/+ (left panel) and yw/w; c929-GAL4/UAS-2xEGFP; tub-GAL80ts/+ (right panel). (B) Time course of DIMM::MYC induction following the shift from restrictive to permissive temperature. Genotype: yw/w; c929-GAL4/UAS-DIMM::MYC; tub-GAL80ts/+. Tubulin was used as a loading control. Hours after shift to 30°C are indicated above each lane. Adult head extracts were analyzed by Western blotting with indicated antibodies.

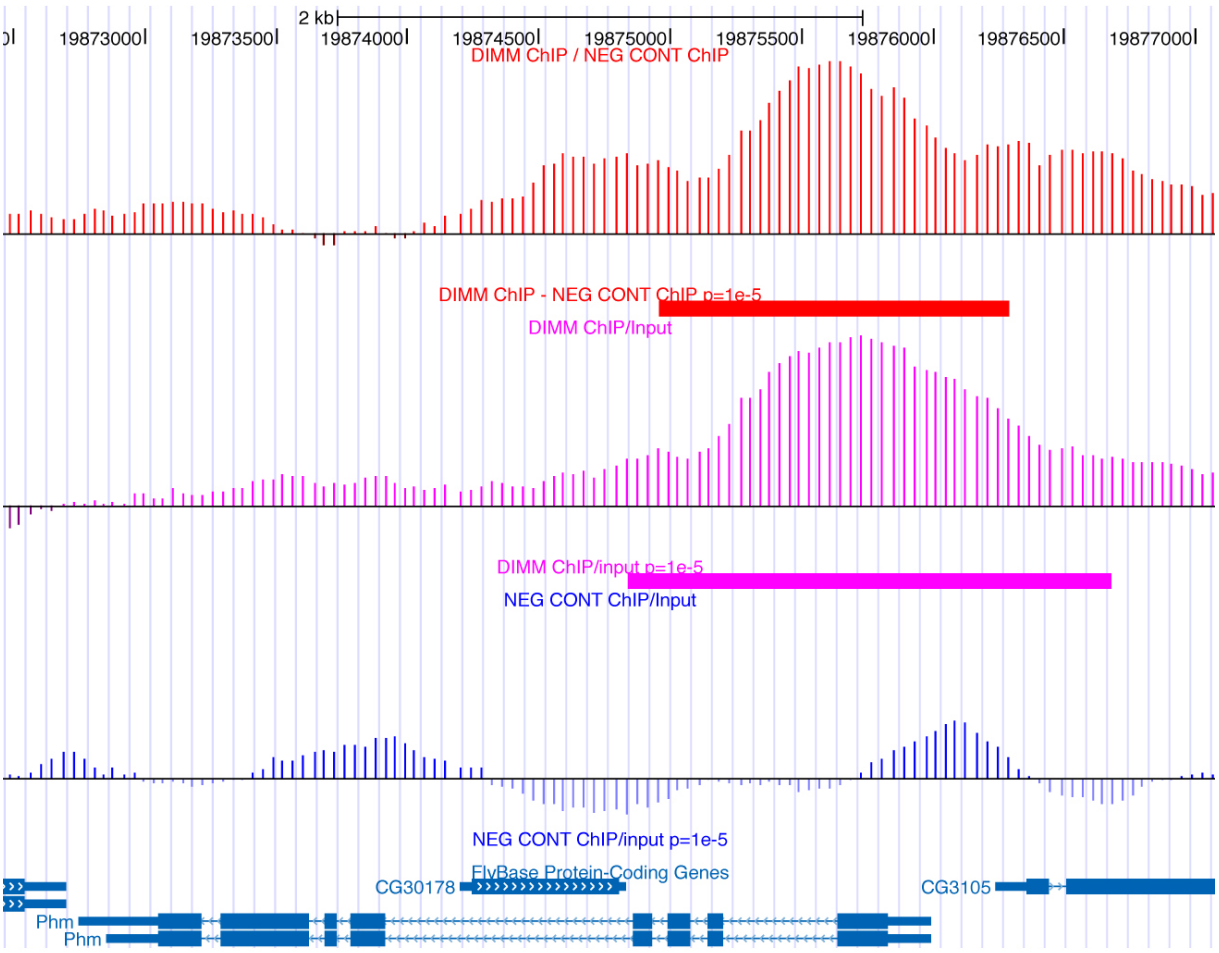

Figure 2. DIMM Binding to its known enhancer in the first intron of PHM by ChIP-chip. MYC antibody ChIP-chip was performed in flies expressing a DIMM::MYC transgene in normally DIMM+ cells. The negative control strain, lacked the DIMM::MYC transgene, but was otherwise identical to the experimental strain. ChIP and input DNA were analyzed using Affymetrix tiling arrays and MAT. The PHM locus is displayed inside the UCSC Genome Browser. In order to rigorously identify true DIMM binding, three types of MAT comparisons were used. Genomic enrichment in the DIMM::MYC ChIP DNA was compared to enrichment in its input (DIMM ChIP/Input, shown in pink). Negative control ChIP DNA enrichment was compared to enrichment of its input (NEG CONT ChIP/Input, shown in dark blue). Finally, ChIP DNA

enrichment in the DIMM::MYC samples was directly compared to DNA enrichment in the negative control (DIMM ChIP / NEG CONT ChIP, shown in red). In order to identify true DIMM binding, peaks from NEG CONT/Input were first subtracted from DIMM ChIP/Input peaks. Secondly, the resulting list of binding regions was then intersected with regions identified in the DIMM ChIP / NEG CONT ChIP comparison. In order for a peak to pass this cutoff, genomic regions in the subtracted list had to have at minimum 1 bp overlap with genomic regions in the DIMM ChIP / NEG CONT ChIP list.

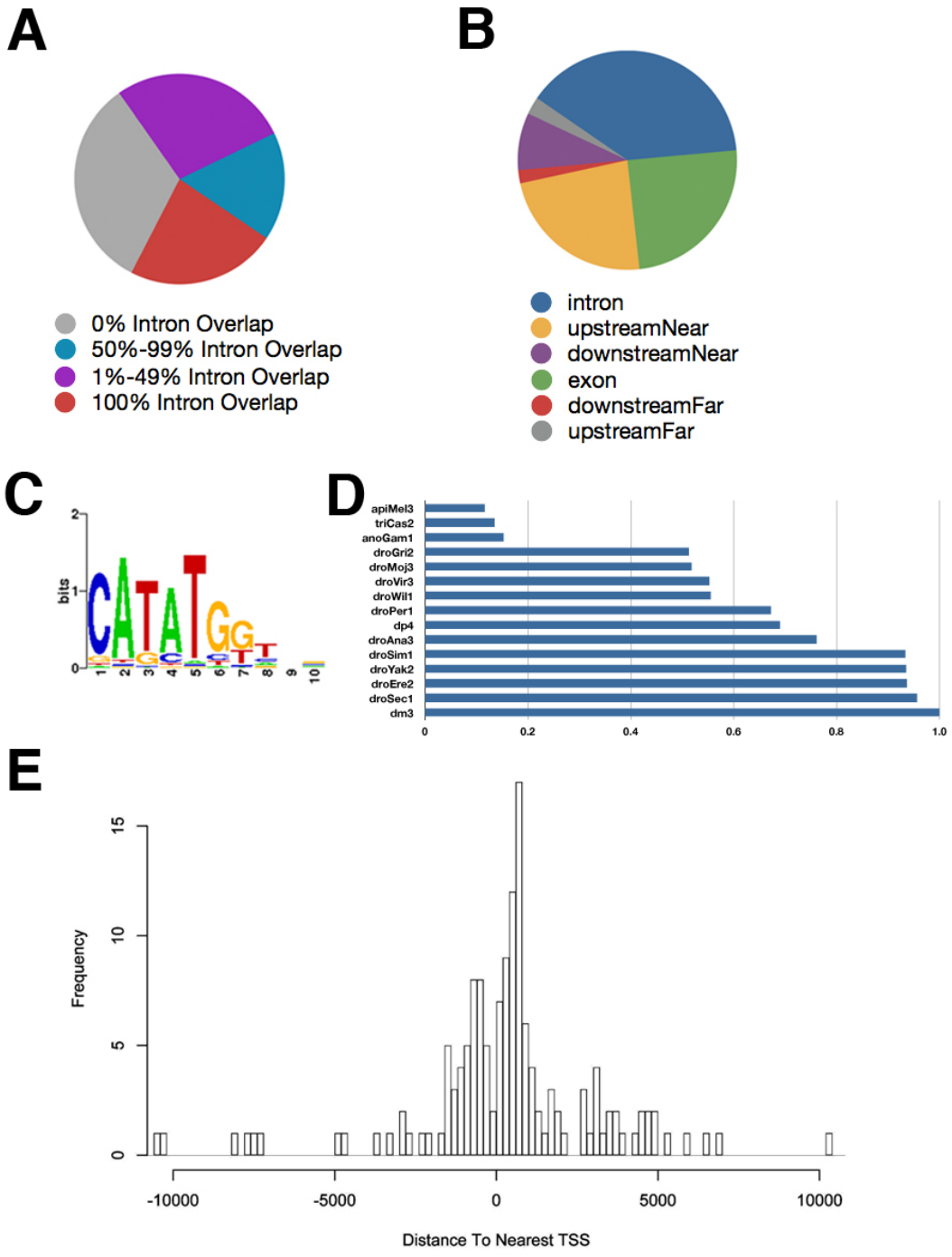

Figure 3. Genomic annotation, motif identification and conservation of DIMM binding peaks. (A) Overlap of DIMM ChIP-chip peaks with introns calculated as percentage of their length. (B) PinkThing software annotation of ChIP-chip peaks against *Drosophila* gene structural elements (exons, introns, promoter and downstream elements. (C) CATATG E-box motif enriched in

DIMM ChIP-chip peaks found in introns (p-value 0.004273). (D) Conservation analysis (Multi Alignment blocks on 15 insect species) of ChIP-chip peaks between *D. melanogaster* (dm3, shown on the bottom) and 14 other insect species. (E) Distance of ChIP-chip peaks from transcription start sites.

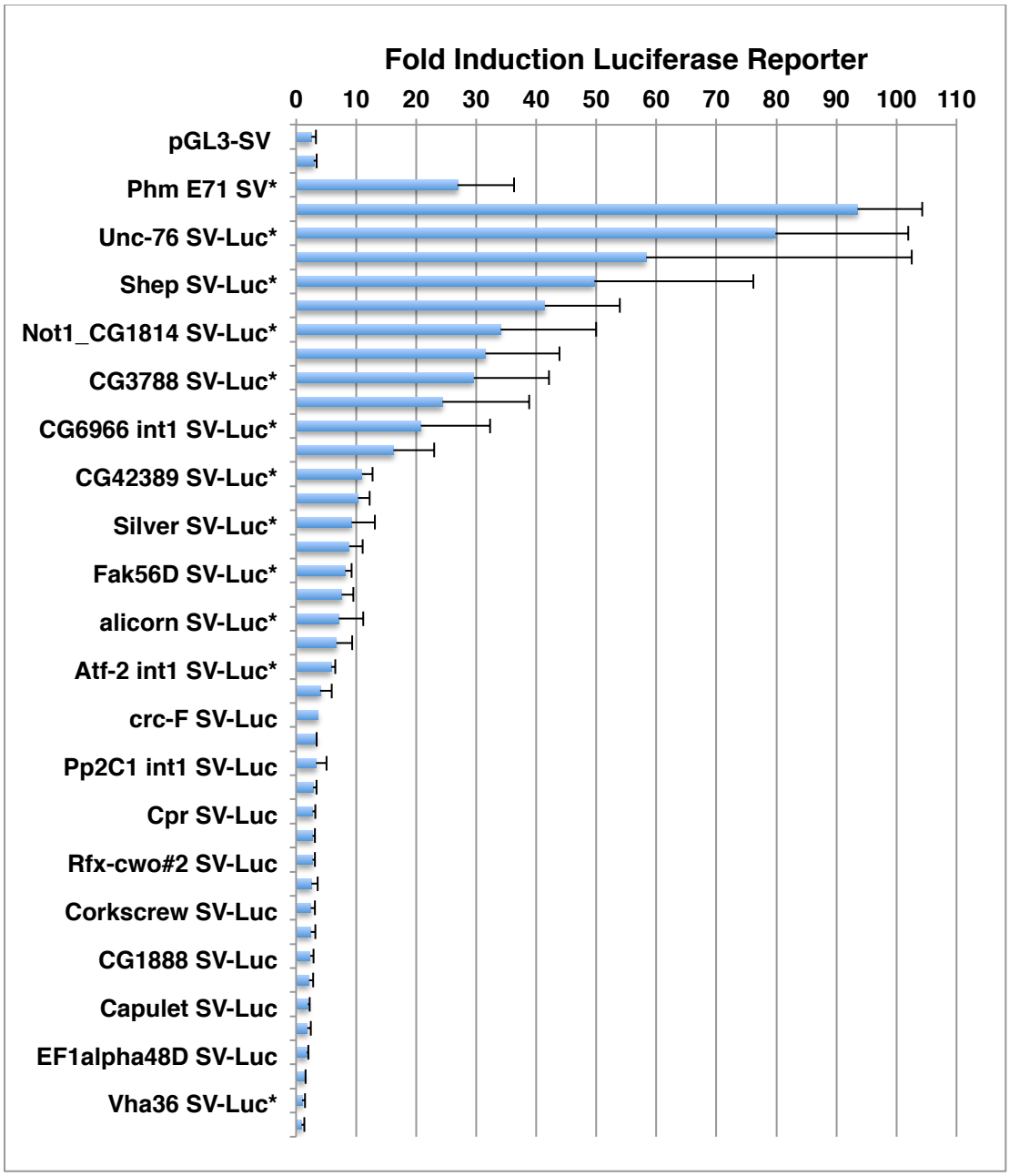

Figure 4. DIMM transactivation of 39 genomic fragments with significant DIMM ChIP-chip binding in a *Drosophila* BG3-c2 neuronal cell line. Luciferase reporter was placed downstream of a mini-SV promoter and a ChIP-chip derived enhancer. Fold ratios represent Luciferase levels with DIMM co-transfection divided by those without. Histograms represents means and SEMs of at least three independent replicate assays except for the crc-F fragment. *p-val<0.05 student t-test

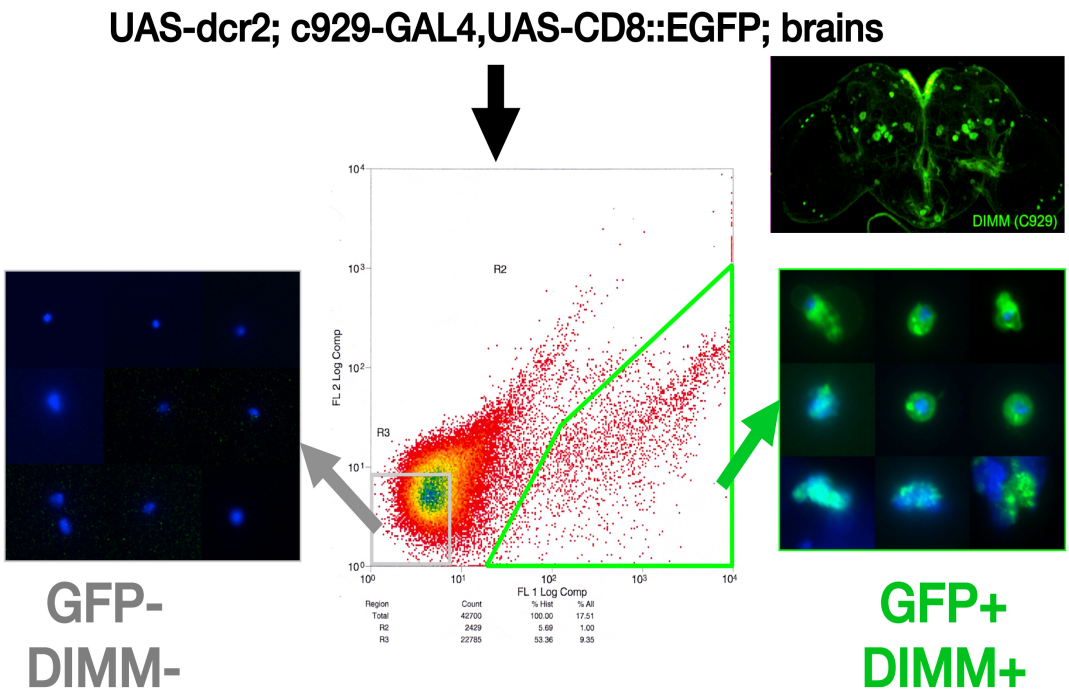

Figure 5. FACS sorting of DIMM+ and DIMM- cells. Brains were dissected from UAS-dcr2; c929-GAL4, UAS-CD8; flies and dissociated into single cells. DIMM+ cells (as marked by c929>GFP expression) and randomly selected GFP- cells were FACS sorted. Top left panel shows the distribution of c929>GFP signal in an adult brain. Examples of GFP+ cells (bottom right panel) and GFP- cells (bottom left panel) are shown.

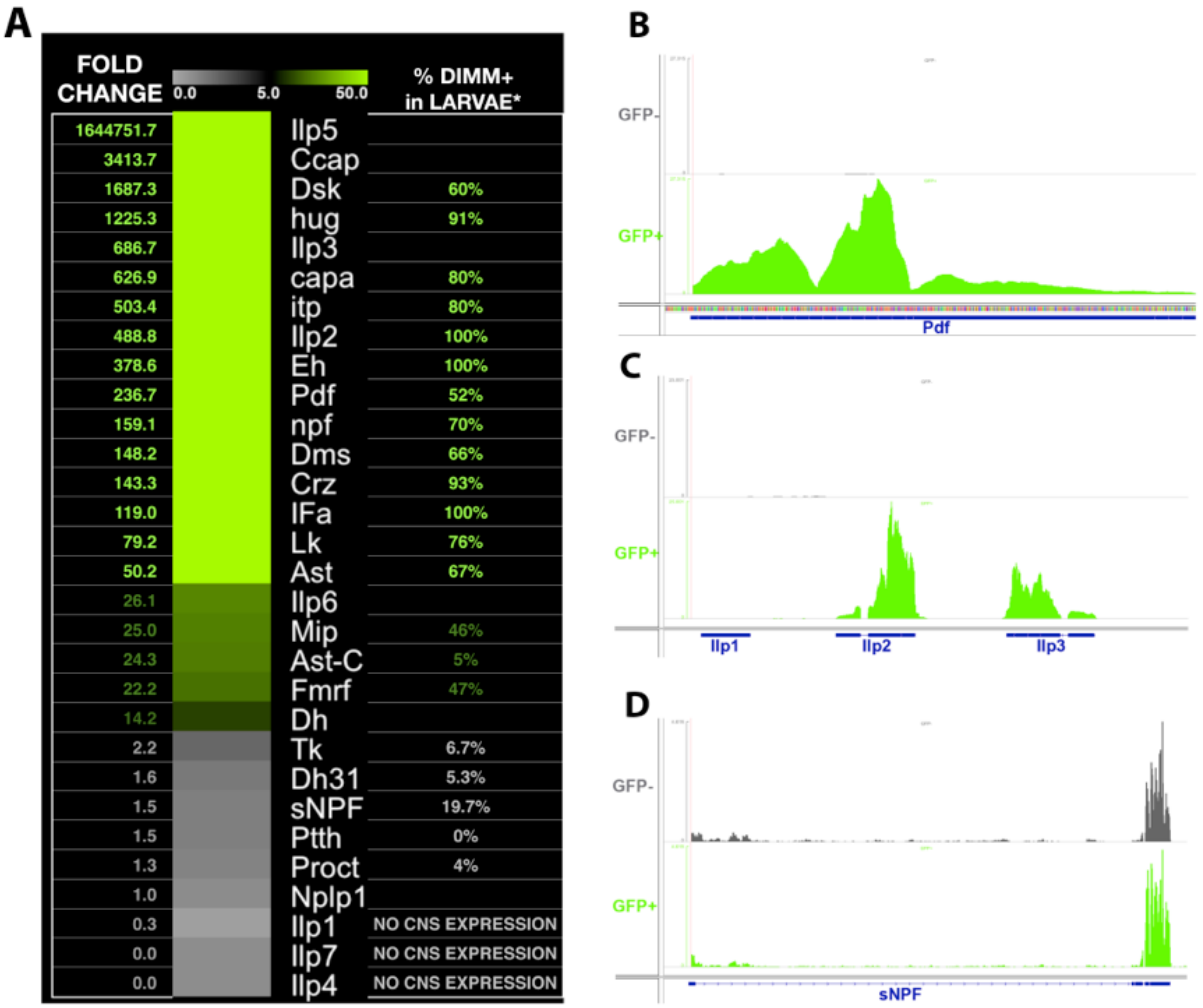

Figure 6. The neuroendocrine transcriptome of LEAP cells revealed by deep sequencing. (A) Enrichment of known *Drosophila* neuropeptides by RNA-Seq in the GFP+/GFP- sample is depicted as fold change (left) and in the form of a heat map (right). Neuropeptide names are indicated in the center of the figure. Overlap of DIMM+ cells with some of the neuropeptides in the larval CNS is given on the right (Park et al. 2008a). (B-D) Integrated Genomics Viewer was used to visualize RNA-Seq tracks from GFP+ and GFP- cells. (B) Pigment dispersing factor (PDF) locus. (C) Ilp1, Ilp2 and Ilp3 loci. (D) short Neuropeptide F (sNPF) locus.

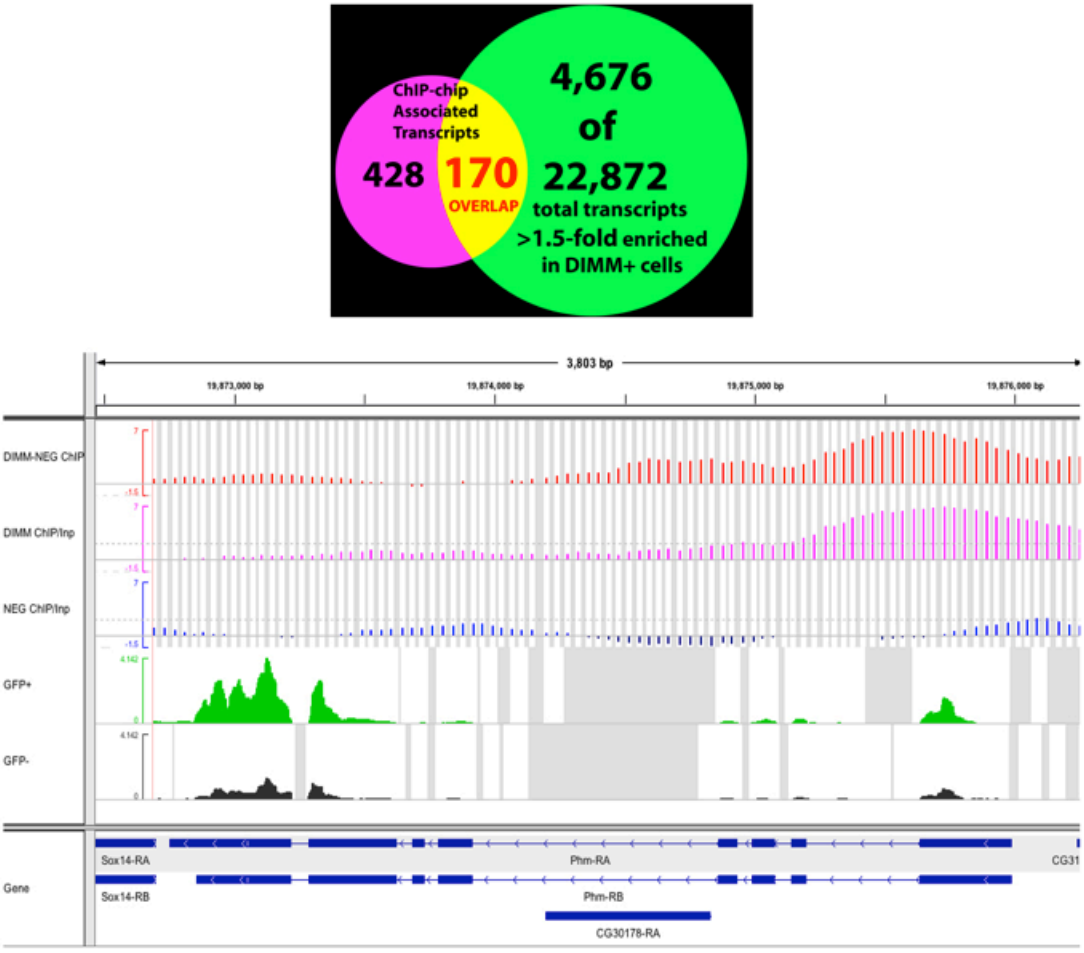

Figure 7. Intersection of DIMM targets detected by ChIP-chip and transcripts enriched in DIMM+ LEAP cells. Venn diagram depicts the intersection of DIMM interactome and LEAP transcriptome genes (top, not drawn to scale). Integrated Genomics Viewer view of the PHM gene locus (bottom). The top three tracks depict ChIP-chip results (DIMM ChIP/Input, pink; NEG CONT ChIP/Input, blue; DIMM ChIP / NEG CONT ChIP, red). The bottom two tracks depict RNA-Seq signal: DIMM+/GFP+ signal is shown in green, and DIMM-/GFP- signal is shown in gray. PHM gene locus structure is shown at the bottom of the window in blue.

Table 1. List of 156 DIMM binding sites identified by ChIP-chip. Genomic coordinates given based on the dm3 genomic scaffold. ChIP-chip sites were obtained by subtracting NEG CONT ChIP/Input sites from DIMM ChIP/Input sites, and by intersecting the resulting sites with the DIMM ChIP / NEG CONT ChIP sites. Please see attached / uploaded Microsoft Excel spreadsheet file.

Table 2. List of 284 DIMM ChIP-chip binding region-associated genes. Please see attached / uploaded Microsoft Excel spreadsheet file.

Table 3. Intersection of the 598 ChIP-chip-associated transcripts and the 4,676 LEAP-cell-specific transcripts. Each entry is described by: gene symbol, CG#, Flybase Gene ID, Flybase Transcript ID and enrichment in DIMM+/GFP+ cells compared to DIMM-/GFP- cells (reported as fold change). Please see attached / uploaded Microsoft Excel spreadsheet file.

Table 4. ChIP-chip / RNA-Seq-derived genes grouped into functional categories based on known gene function reported in literature. Please see attached / uploaded Microsoft Excel spreadsheet file.

Chapter 4.

Assessing the contribution of individual DIMM targets to DIMM-dependent LEAP cell physiology by an RNAi-based genetic screen for sleep defects

I performed the sleep experiments in this chapter: genetic crosses, behavioral monitoring and data analysis with some technical help from Neely Williams.

INTRODUCTION

LEAP cells express neuropeptides, many of which function in modulating various behaviors (Nässel 2002). DIMM-expressing neuropeptides control metabolism and a variety of behaviors, such as ecdysis and sleep (Park et al. 2008a; Kim and Rulifson 2004; O'Brien and Taghert 1998; Sheeba et al. 2008; Shang et al. 2008). I have chosen to evaluate the functional consequences of DIMM target knockdown using sleep as a regulated behavior in *Drosophila* that has been an established paradigm (Shaw et al. 2000). A subset of LEAP cells, the large LNv neurons have been implicated in the control of sleep and light-mediated arousal (Sheeba et al. 2008; Shang et al. 2008). When DIMM+ large LNv and DIMM- small LNv neurons are chronically hyperexcited, nocturnal activity is increased (Sheeba et al. 2008). Furthermore, when the authors hyperexcited LEAP cells but excluded PDF-positive Large and Small LNv neurons, flies exhibited increased daytime and nighttime locomotor activity with decreased sleep (Sheeba et al. 2008). Another study showed that activation of DIMM- small LNvs had no effect on sleep or arousal (Shang et al. 2008). This study provides the best evidence the effect of PDF neuron hyperexcitation on sleep and arousal is attributed solely to the DIMM+ large LNv neurons and not to the DIMM- small

A direct connection between DIMM and sleep has not been established in the literature and is reported for the first time in this chapter. When DIMM expression is reduced in small and large LNv neurons, this leads to an increase in sleep compared to controls (Figure 1A). Therefore, a functional genetic screen for sleep defects was conducted by RNAi-mediated gene knockdown specifically in LEAP cells. This allowed for interrogation of the intersected interactome / transcriptome data set by evaluating individual gene targets in a functional, RNA interference (RNAi) based screen. This would be relatively feasible considering the ease of conducting genetics in *Drosophila*.

MATERIALS AND METHODS

Sleep

Sleep assays were performed according to standard procedures (Andretic and Shaw 2005). Conditions were standardized to avoid any inconsistencies since behavior can be affected by many factors. Virgin females of genotype *w,UAS-dcr2; c929-GAL4;* were crossed to males from the Vienna *Drosophila* RNAi Center (VDRC; Dietzl et al. 2007). The screen was carried out with the VDRC KK library, which was made by inserting UAS-RNAi casettes into a single attP genomic site on chromosome II. As controls*, w,UAS-dcr2; c929-GAL4;* virgins were crossed to the *yw; attP;* males. The *yw; attP;* stock was originally used to create the VDRC KK RNAi library and is isogenic with the UAS-RNAi lines. Two or more crosses were done for each genotype in each biological replicate in order to be able to collect 20 female offspring within 48 hours. For each cross, 15 virgin females were crossed to 5 males at 25°C in standard *Drosophila* vials. Males and females were allowed to mate for 5 days, and were then transferred out of the vials. Daughters from each cross were collected within 12 hours of eclosion and were in almost all cases virgins. After collection, virgin flies were housed in regular fly vials with ~20 flies per vial for three days under 12 hour light / 12 hour dark (LD) conditions. In all cases, flies were raised and aged on the same food medium (standard *Drosophila* medium). On day four, flies were placed in 65 mm x 5 mm glass tubes (VWR glass) containing standard *Drosophila* medium. Thus, flies were raised and their sleep tested while being fed the same medium. Flies were then transferred to behavior incubators, where they spent one day acclimating to LD conditions. Light emitting diodes (LEDs) were used as a light source. Light intensity was tightly controlled to be in the range of 40 – 50 Lux for all flies. Locomotor activity was collected with second generation *Drosophila* Activity Monitoring (DAM2) System monitors in one-minute bins (Trikinetics, Waltham, MA). Fly sleep was measured for five days, and analysis was done on the first day of behavior. Although data consistency from day one to day five is relatively high, a minority of flies

died during recording (from desiccation due to an accidental breach of the wax plug integrity, or other unknown factors). Therefore, the first full day of behavior is used for all sleep analyses. Sleep analysis was done with Excel macros provided by Dr. Paul Shaw (Washington University). Per convention, sleep was measured as bouts of uninterrupted inactivity lasting 5 minutes (Andretic and Shaw 2005). Daytime sleep is defined as the total amount of sleep during the first twelve hours of the day (this is the period during which lights are on). Daytime sleep is defined as the total amount of sleep during the first twelve hours of the day (this is the period during which lights are on). Sleep data shown in Figure 1A were provided by Dongkook Park and were obtained by crossing *UAS-dcr2; pdf-GAL4;* females to *UAS-dimm RNAi* males. Two different DIMM RNAi lines from the VDRC library were tested: a KK line with precisely inserted UAS-RNAi, as well as DIMM RNAi from the GD library, which was created by random insertion transgenesis (Dietzl et al. 2007).

RESULTS

Functional genetic screen of integrated DIMM targets

The 170 transcripts that were identified by integrating DIMM ChIP-chip targets and the LEAP transcriptome are encoded by a total of 116 genes (Chapter 3 Table 3). Therefore, this list of 116 “integrated” genes likely represents true DIMM targets. In addition to *in vivo* binding and expression, DIMM can directly transactivate at least 12 of these target genes in luciferase *in vitro* transactivation assays (Chapter 3 Figure 4). Nevertheless, it is currently not known how DIMM target genes might affect the physiology of LEAP cells. Therefore, the role of these genes in DIMM-dependent LEAP cell functions was tested in a genetic screen for sleep defects. When DIMM expression is reduced in PDF-expressing LNv neurons, flies exhibit increased daytime sleep (Figure 1A). This effect can be solely attributed to DIMM+ large LNv neurons, which are a subset of LEAP cells (Sheeba et al. 2008; Shang et al. 2008). DIMM- small LNv neurons do not

express DIMM and likely have no role in sleep (Park et al. 2008; Sheeba et al. 2008; Shang et al. 2008).

In order to test the effect of DIMM target gene knockdown on sleep, RNAi was used to trigger gene knockdown in LEAP cells and the effect of this manipulation on sleep was monitored. Of the 116 DIMM target genes identified by integrating ChIP-chip and RNA-Seq data, 62 were available from the VDRC KK RNAi library. These 62 isogenic RNAi lines were created by precise P-element insertion were tested 2 – 4 times as independent biological replicates in sleep assays. Results showed that manipulating many of DIMM target genes had an effect on one or more sleep parameters. Various sleep parameters examined included: sleep latency (time to the first sleep episode of the night), length of sleep episodes (bouts) during the day and night, length of the longest sleep episode during the day and night, as well as the number of sleep episodes during the day and night (Andretic and Shaw 2005). The most obvious effect of DIMM knockdown in PDF+ neurons is an increase in daytime sleep. Therefore, this was the first parameter examined in the sleep screen (Figure 1B). Indeed, knockdown of many of the high confidence DIMM targets in LEAP cells showed a tendency towards increased daytime sleep, which partially phenocopied DIMM loss-of-function sleep phenotype (Figure 1A). This effect was not significant by a one-way ANOVA with a Bonferroni correction except in the case of Tango14 knockdown. A third of the crosses tested had daytime sleep changes that were significant by an unpaired student t-test (p-value 0.05). By visual inspection, RNAi lines clearly cluster in the right half of the graph, whereas control flies (shown in black) sleep ~100 minutes during the first twelve hours of the day (Figure 1B, 1C). RNAi manipulations with some of the most extreme deviations in daytime sleep were: *Tango14*, *CG15394*, *jaguar* (Myosin VI), and *Rpn9* (Figure 1C). Additionally, many genes had strong effects on other sleep parameters (data not shown).

DISCUSSION

When genome-wide studies yield lists of genes implicated in a certain process, it is commonly useful to evaluate the validity of such genes in a functional screen. Since LEAP cells, and in particular the large LNv DIMM+ neurons, have been implicated in sleep control, I chose to conduct an RNAi-based screen for sleep defects induced by loss of DIMM target gene expression in LEAP cells. The sleep screen showed that knockdown of a significant portion of the DIMM target genes identified by interactome/transcriptome integration produced sleep defects. In particular, a third of tested RNAi lines phenocopied DIMM RNAi sleep defects with statistical significance.

There are several explanations available for explaining why all RNAi lines did not phenocopy DIMM RNAi sleep defects. Many DIMM target genes could function redundantly with several other proteins, so it is unlikely that knockdown of any one gene would be able to phenocopy the DIMM loss-of-function sleep phenotype. Indeed, only a few genes produced a sleep phenotype across all the various sleep parameters such as daytime and nighttime sleep, sleep latency, the number and length of sleep episodes. Nevertheless, many genes produced defects in at least one of these categories compared to controls. Furthermore, due to the inherent variability in sleep time results that exists even after the most rigorous controls, it is entirely possible that finishing the screen with 4 biological replicates for each tested genotype could lead to increased ability to determine statistical significance (Andretic and Shaw 2005).

Another potential reason for why some of the tested 62 lines failed to show statistically significant sleep changes could be due to the GAL4 line used in the sleep study. Although c929-GAL4 overlaps greatly with DIMM-expressing neurons, this collection of cells is highly heterogeneous with respect to neuropeptide expression. C929-GAL4 cells express at least 17 different neuropeptides, and likely more (Park et al. 2008). It is possible that some neuropeptides could be sleep-promoting, whereas others, such as PDF, could be wake-

promoting (Parisky et al. 2008). Therefore, interfering with synthesis, accumulation, storage or trafficking of sleep-promoting and wake-promoting neuropeptides could have neutralizing effects on sleep. In other words, sleep defects induced by RNAi-based knockdown of DIMM target genes in one subgroup of c929-GAL4 cells could be cancelled out by another subgroup of c929-GAL4 cells. This could be correct by perhaps using pdf-GAL4 for a repeat sleep screen. This GAL4 line is only expressed by ~20 PDF-expressing peptidergic neurons that have clearly been implicated in wake-promoting activity (Sheeba et al. 2008; Shang et al. 2008).

ACKNOWLEDGEMENTS

I would like to thank Neely Williams for generous help with the sleep screen, as well as Dongkook Park for providing data in Figure 1A, and Weihua Li for help with the circadian / sleep apparatus. I would also like to thank Dr. Paul Shaw for his help with analyzing sleep data.

REFERENCES

Andretic, R., and Shaw, P.J. (2005). Essentials of sleep recordings in *Drosophila*: moving beyond sleep time. *Meth Enzymol* **393**, 759-772.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S.*, et al.* (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151-U151.

Kim, S.K., and Rulifson, E.J. (2004). Conserved mechanisms of glucose sensing and regulation by *Drosophila* corpora cardiaca cells. *Nature* **431**, 316-320.

Nässel, D.R. (2002). Neuropeptides in the nervous system of *Drosophila* and other insects: multiple roles as neuromodulators and neurohormones. *Prog. Neurobiol.* **68**, 1-84.

O'Brien MA, Taghert P.H. (1998) A peritracheal neuropeptide system in insects: release of myomodulin-like peptides at ecdysis. *J. Exp. Biol.* **201** (Pt 2): 193–209.

Parisky, K. M., Agosto, J., Pulver, S. R., Shang, Y., Kuklin, E., Hodge, J. J. L., Kang, K., Kang, K., Liu, X., Garrity, P. A., et al. (2008). PDF cells are a GABA-responsive wake-promoting component of the Drosophila sleep circuit. *Neuron* **60**, 672–682.

Park, D., Veenstra, J.A., Park, J.H., and Taghert, P.H. (2008). Mapping peptidergic cells in *Drosophila*: where DIMM fits in. *PLoS ONE* **3**, e1896.

Shang, Y., Griffith, L.C., and Rosbash, M. (2008). Light-arousal and circadian photoreception circuits intersect at the large PDF cells of the *Drosophila* brain. *Proc. Natl. Acad. Sci. USA* **105**, 19587-19594.

Shaw, P.J., Cirelli, C., Greenspan, R.J., and Tononi, G. (2000). Correlates of sleep and waking in *Drosophila* melanogaster. *Science* **287**, 1834-1837.

Sheeba, V., Fogle, K.J., Kaneko, M., Rashid, S., Chou, Y.-T., Sharma, V.K., and Holmes, T.C. (2008). Large ventral lateral neurons modulate arousal and sleep in *Drosophila*. *Curr. Biol.* **18**, 1537-1545.

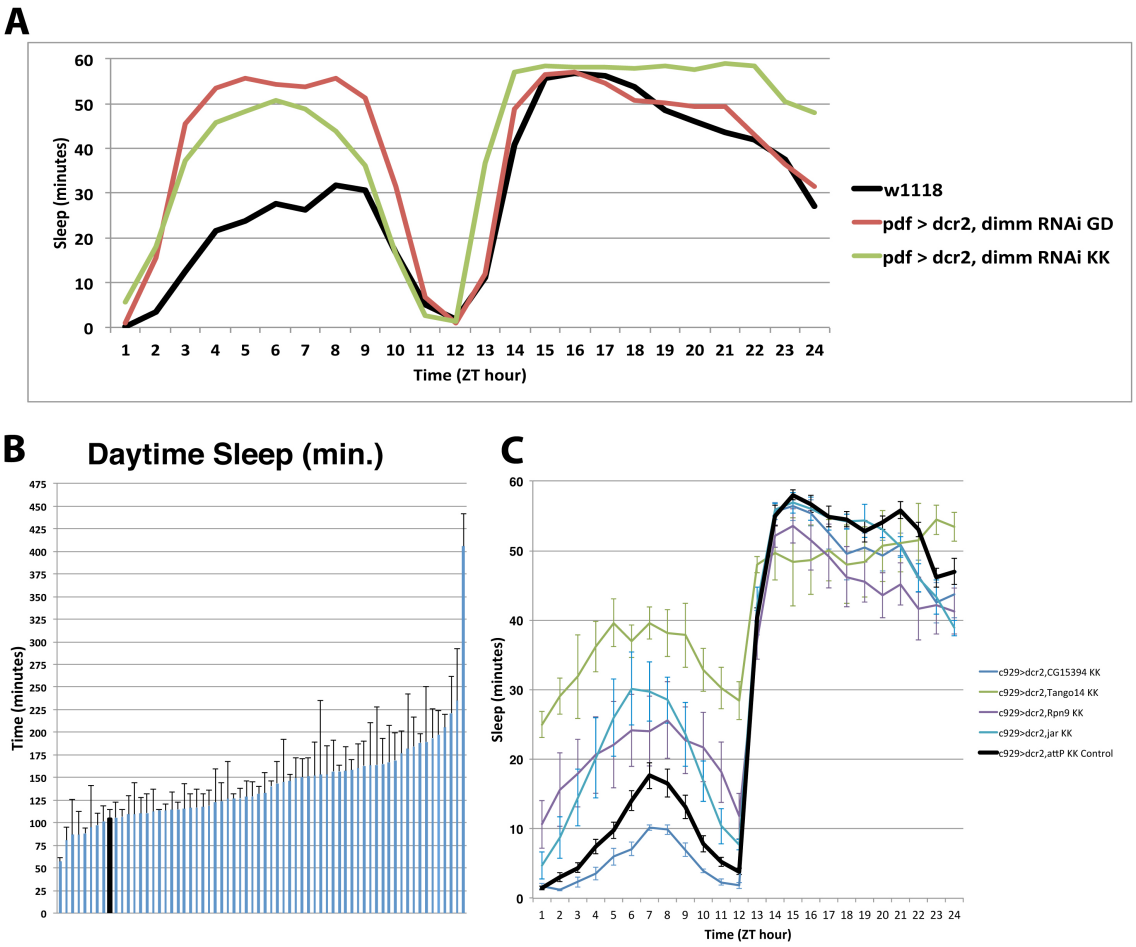

Figure 1. Genetic screen for sleep defects induced by RNAi of DIMM targets derived from integrated interactome / transcriptome analysis (A) DIMM RNAi in PDF-expressing LNv cells increased daytime sleep compared to w^{1118} controls. Pdf-GAL4 was used to knock down DIMM expression with two different dimm RNAi constructs from the Vienna *Drosophila* RNAi Center (VDRC). (B and C) DIMM targets were knocked down by RNAi expressed in DIMM+ cells. UAS-dcr2; c929-GAL4; females were crossed to UAS-RNAi males from the VDRC attP-based KK Library or the UAS-attP empty vector controls. (B) Knockdown of many of the integrated DIMM targets in DIMM-expressing cells shows a tendency towards increased daytime sleep, which partially phenocopies DIMM loss-of-function sleep phenotype. Daytime sleep values are displayed for 62 tested KK VDRC RNAi lines (blue) and the attP control progeny (shown in black

color). Histograms represents means and SEMs of at least two and in most cases three or four independent biological replicate assays (C) Sleep per hour depicted over a circadian day for selected genotypes from the same analysis as in (A). Control shown in black.

Chapter 5.

Concluding remarks about the role of the scaling factor DIMM in LEAP cell physiology

THESIS SUMMARY

This thesis presents two approaches to understanding the mechanism that DIMM uses to instruct cells to produce LDCVs, scale up RSP components and develop most or all competencies required of a NE cell except for selection of neuropeptide cargo. In Chapter 2, I contributed molecular and bioinformatic efforts to a larger, multidisciplinary study. This study aimed to understand how DIMM functions by trying to identify a few crucial DIMM targets and to confirm the fidelity of these targets by multiple assays. The contribution of these individual gene targets to DIMM-dependent physiology was assessed by cellular, molecular and proteomic assays. This study provided the first specific clues about how DIMM functions in a molecular context. In Chapter 3, I took a genome-wide approach to understanding the mechanism of DIMM function. I first attempted to identify all direct genomic targets of DIMM by ChIP-chip. In order to correlate genome binding to gene activation, I profiled gene expression in LEAP cells by RNA-Seq. I then merged these two data sets by identifying ChIP-chip targets that were enriched in LEAP cells. The integration of these two data sets allowed me to obtain the most direct measure of DIMM direct gene targets. In Chapter 4, I carried a pilot genetic screen for sleep behavioral defects caused by RNAi-based knockdown of DIMM target gene expression. These results show preliminary effects on sleep that need to be further investigated. In sum I produced a body of information detailing precisely how DIMM operates in LEAP cells by identifying the genes that it regulates, their expression levels and preliminarily, some functional consequences of their knockdown. Here I discuss the overall scope of the work, including what I see as limitations and caveats, as well as a summary of the major conclusions and my opinions on potentially useful future directions.

Section on Limitations.

Limitations of tagged DIMM::MYC ChIP-chip.

In Chapter 3, I carried out tagged ChIP-chip to identify direct DIMM binding targets. This approach is a modification of traditional ChIP-chip, which is carried out with native antibodies against the transcription factor of interest to immunoprecipitate DNA-protein adducts after crosslinking. I resorted to tagged ChIP-chip after originally attempting ChIP with an anti-DIMM antibody that was developed in the laboratory, and which has shown specificity in tissue staining and Western blots. Unfortunately, this "native" antibody did not perform well in ChIP-chip in my hands. According to my reading of the literature, this is not an uncommon occurrence, as ChIP antibodies must be able to capture fixed epitopes in solution. I therefore adapted my approach by using the GAL4-UAS system to express a single copy of a DIMM transgene tagged at the C-terminal with 6 copies of the human hexapeptide MYC tag. I used *c929*-GAL4 in combination with the TARGET system in order to achieve precise, spatiotemporally controlled induction of DIMM::MYC only in DIMM+ adult neurons (McGuire et al. 2003).

Application of affinity-tagged transgenes is a frequently-used strategy for carrying out ChIP-chip *in vivo* (Kolodziej et al. 2009; Mazzoni et al. 2011). Tagged ChIP-chip is used when a ChIP-grade antibody is not available or when a native antibody recognizes several isoforms of a transcription factor, or also when it recognizes heterologous molecules (Harada and Nepveu 2012). Nevertheless, as with any approach, including native antibody ChIP-chip, there are drawbacks to tagged ChIP-chip. One major potential source of error (false positives) is genome-binding artifacts arising from overexpression of the transcription factor being studied. The hypothesis here is that increasing expression levels of the transcription factor under study, beyond those found in normal cells, leads to increased competition with other transcription factors for similar

binding sites in the genome. This, in turn, could cause promiscuous binding to occur at those loci not normally bound by the transcription factor in question. While this is a valid point, several experiment-specific conditions need to be examined. The main issues are those of space and time where the ChIP-chip transgene is induced. Is transgene induction global, e.g., in the whole brain, or is it more limited, as I performed my study in Chapter 3? Second, is the transgene induced from the birth of the animal, when chromatin structure is being established, or is the transgene induced in fully differentiated, adult cells, as in Chapter 3? In the latter example, the animal develops chromatin boundaries normally, and its cells express normal levels of transcription factors, until the transcription factor transgene is turned on in adult, post-mitotic cells. Overexpression artifacts are more likely to happen with broad overexpression that begins at birth and is maintained for the life of the animal. In order to avoid such artifacts, I selectively induced DIMM::MYC in adult cells that normally express DIMM.

Another issue is the potential for off-target binding by a transcription factor that is expressed at supra-physiological levels. For example, it may compete with other similar transcription factors (in this case, other members of the Atonal superfamily of bHLHs) for related target sites. In theory, extraneous binding would be weaker than endogenous binding because affinity for sites normally bound by other bHLHs is presumably lower than the affinity for the sites that the particular bHLH under study binds on its own. False positives could also occur by binding to sites that match the native binding site but are normally not bound by any transcription factors. Since bHLHs are known to bind six-base pair long E-boxes (Powell and Jarman 2008), such sequences occur throughout the genome in all parts of a gene, as well as intergenically at a rate of approximately once every 2 kB. It is possible that some of these sites are found in open chromatin due to stochasticity or unrelated transcription. Hence, an overexpressed transcription factor

could bind to such sites more frequently than the same transcription factor at physiologic levels. Finally, it is possible that the affinity tag that is placed at the N or C-terminal of a transcription factor could interfere with the native function of the transcription factor in question, or it could cause some neomorphic interactions. The UAS-DIMM::MYC transgene used in Chapter 3 is capable of rescuing the *dimm* null mutants (Hewes et al. 2003). Therefore, the possibility that the MYC hexapeptide interferes with the normal function of DIMM is less likely, due to the successful transgene rescue of the mutant.

A recent study examined results obtained from tagged ChIP versus native antibody ChIP against the bHLH Olig2 at great length (Mazzoni et al. 2011). Comparisons between native antibody and tagged ChIP-seq showed that endogenous Olig2 and inducible Olig2-V5 ChIP-seq experiments were in agreement (Mazzoni et al. 2011). Furthermore, the binding-site distribution found in both experiments was also highly concordant (Mazzoni et al. 2011). Comparing the read counts at enriched peaks showed that only 0.2% and 1.1% were differently enriched in the native Olig2 and Olig2-V5 ChIP experiments, respectively (Mazzoni et al. 2011). This is the first study to systematically compare tagged ChIP and native ChIP results. The authors used a doxycycline-inducible system in differentiating embryonic stem cells and generated 24 tagged lines. Interestingly, the authors observed that the efficiency and homogeneity of transgene induction declined in postmitotic neurons (Mazzoni et al. 2011). My DIMM ChIP-chip study (Chapter 3) likely does not suffer from the decreased postmitotic efficiency because ChIP-chip was carried out *in vivo,* in cells that were already expressing normal levels of DIMM. Therefore, the chromatin state of DIMM-positive cells already allowed for DIMM expression and DIMM target expression, with DIMM cell specification and progenitor delineation having occurred normally during DIMM::MYC-unaffected development. Furthermore, DIMM is normally expressed only in postmitotic

neurons, with DIMM expression not commencing until the neuroblast daughter cell exits the cell cycle in the embryo (Allan et al. 2005; Hewes et al. 2003). Nevertheless, even though the experimental setup that I used in Chapter 3 differs from that of Mazzoni et al. (2011), both sets of experiments were carried out on tagged transcription factors that were induced from transgenes. Mazzoni et al. (2011) were fortunate to have ChIP-grade native antibodies available in addition to the tagged transgenes. The authors were therefore able to compare tagged ChIP-seq results directly to native antibody ChIP-seq results and observed a remarkable overlap in results.

The above referenced study is an example of an excellent validation study that demonstrates the utility of tagged ChIP in identifying transcription factor binding (Mazzoni et al. 2011). Nevertheless, each system is different, so proper scrutiny would require that each tagged ChIP experiment be verified by native antibody ChIP. In most cases, this is not possible, because tagged ChIP is usually employed because native antibody ChIP failed or a native antibody is not available. In the case of DIMM, there currently is no ChIP-grade native antibody available. Therefore, DIMM::MYC ChIP-chip results cannot be compared against DIMM native antibody ChIP-chip. It is possible that some of the binding peaks identified by DIMM ChIP-chip could represent overexpression artifacts and not true DIMM binding sites. This possibility cannot be completely excluded, no matter how careful the experimental design is.

Even with native antibody ChIP-chip, there is no guarantee that all identified binding sites represent true events. In most cases, authors test the performance of a certain number of binding sites in transcription factor luciferase reporter assays, such as the one employed in Chapter 3. This *in vitro* assay directly tests the ability of an activating transcription factor to activate luciferase expression from an enhancer fragment located on an episome. An alternative technique is the electrophoretic mobility

shift (EMSA) assay, which is even more devoid of a cellular context and can have a high rate of false positives. Ultimately, gene expression studies are another way to deduce the validity of ChIP-chip results. Optimally, a gene expression profile of the cell type in which the ChIP-chip transgene was induced should be obtained. Additionally, it is beneficial to have a loss-of-function gene expression profile, obtained from the cell type of interest lacking expression of the transcription factor under study. In Chapter 3, I presented the LEAP cell transcriptome but we currently do not have a DIMM loss-of-function LEAP cell transcriptome. The normal transcriptome is useful in assessing the correlation of ChIP-chip peaks with gene expression. As outlined in Chapter 3, the basic assumption is that DIMM acts as an activator of gene expression. Therefore, DIMM binding in the proximity of a gene should correlate with the enrichment of that gene in LEAP cells.

FACS sorted cells may be damaged

Preparation for FACS sorting requires that cells be disaggregated from each other and dissociated into single cells or clumps of a few cells (Givan 2011). During FACS, cells are exposed to mechanical damage from high pressures and speeds. As a consequence, all information about tissue architecture and cell distribution is lost after FACS sorting (Givan 2011). Strictly speaking, such information is not required for gene expression profiling or ChIP-chip studies. These studies do not examine the morphology of cells, but rather their RNA or nuclear contents. Nevertheless, dissociation of cells in preparation for FACS can damage cells even before FACS is started. When fragile cells are put through a FACS machine that generates high pressures and speeds, the cells of interest can be further damaged. FACS can produce strong shearing forces, making it tricky to sort more fragile, adult neurons (Hempel et al. 2007). Whether or not this affects gene expression depends on the tissue and cell type, dissociation method, type and

speed of FACS sorting and other factors. Cell viability after FACS sorting can be checked by various vitality dyes that enter damaged cells that have initiated apoptosis. In order to assess cell vitality, I used a vital dye in Chapter 3, which showed that the majority of LEAP cells are alive after sorting. Although cells may be intact mechanically, it is still possible that they are quite altered due to FACS sorting. The key parameter here is the amount of time between tissue harvesting and the sorting endpoint, which is RNA capture in cell lysis buffer. This last step essentially freezes the cellular RNA contents in time by lysing cells and neutralizing RNase activities. In my experiments, I tried to keep this time as short as possible, to minimize any cell damage effects of cell sorting.

The major argument I can offer against the possibility that the quality of the RNA-seq results I obtained were compromised by the FACS methods is that the DIMM-positive collection of cells produced the heavy enrichment of DIMM and neuropeptides that was expected from our prior understanding of DIMM cell biology. For example, there are certain neuropeptides that are expressed exclusively in DIMM-positive LEAP neurons, whereas others are expressed in DIMM-negative ones exclusively, and yet others in both sets (Park et al. 2008). My results were highly concordant with these three general categories. Hence I submit that the possibility for a biasing of my results (due to disruption of gene expression by cell isolation) appears low.

Section on Major Conclusions

A delimited and defined number of transcription factors downstream of DIMM

In Chapter 3, I showed that only a few transcription factors are found amongst DIMM targets. One transcription factor family appeared to be overrepresented: members of the Atf/CREB family of bZIP genes. The fact that several bZIP genes are DIMM targets is intriguing in light of the fact that DIMM only activates six transcription factors, and those six come from three transcription factor families. There are an estimated 451

sequence-specific transcription factors in *Drosophila* (The *Drosophila* Transcription Factor Database FlyTF.org, v1.0). Of 451 transcription factors, there are only 16 bZIP genes and 4 of the 16 bZIPs are DIMM targets. Compared to the relatively small bZIP gene family, other transcription factor families contain many more members: bHLHs (53 members), homeodomain factors (73 members) and zinc finger transcription factors (91 members). The fact that four of six DIMM-targeted transcription factors are bZIPs is likely telling of DIMM function. ATF/CREB bZIP factors control a wide variety of physiologic processes, but what they share in common is transcriptional control of stress-response genes (Persengiev and Green 2003).

One member of the mammalian ATF/CREB family, Atf2 has been shown to interact with three beta cell enriched transcription factors to cause synergistic activation of the insulin promoter (Han et al. 2011). This is interesting in light of the fact that data presented in this thesis strongly argue against DIMM acting as a transcriptional activator of neuropeptide expression. Drawing parallels to Atf2 in the mammalian beta cell, it is possible that some of the ATF/CREB transcription factors that DIMM activates could contribute to the transcriptional activation of individual neuropeptide genes. Therefore, while DIMM does not activate neuropeptides, some of its downstream transcription factor targets could do so combinatorially with other factors. Another mammalian ATF/CREB factor, Atf-1, acts as a potent repressor of cyclic AMP-responsive element (CRE)-mediated transcription. Atf-1, also known as Inducible cAMP early repressor (ICER) plays an important role in the mammalian neuroendocrine and circadian systems, where it regulates transcription of the neuropeptide melatonin (Foulkes et al. 1996; Foulkes et al. 1997). ICER also coordinates reaction to hypothalamic-pituitary-adrenal axis stimulation by inhibiting Corticotropin Releasing Hormone (CRH) transcription (Della Fazia et al. 1998; Borlikova and Endo 2009). The fact that several mammalian and fly

ATF/CREB factors play roles in the (neuro)endocrine system provide more support for the role of DIMM as the super-organizer of such factors, and thereby the secretory capacity of a cell.

One view of scaling factor action is that such factors do not activate many transcription factors. If a scaling factor activated many downstream transcription factors, its effects would depend on the timing and ability of the downstream transcription factors to exert precise and coordinated effects on their genomic targets. This would delay action of the master scaling factor itself and would therefore not be a very effective way to exert its function. Instead, a scaling factor can act more immediately by activating 'terminal' genes needed for specific functions in a subcellular compartment, such as LDCVs. In addition to these "terminal" genes, a scaling factor would activate only a few key transcription factors that are also responsible for the function of a particular subcellular compartment.

If DIMM had to control a large number of downstream transcription factors, its function would require a coordinated effort of multiple downstream DNA-binding activities that would have to produce a timely effect on the transcriptome of the cell. In contrast, I propose that if DIMM activates a few downstream factors and many direct RSP participants, this would allow for a more dynamic and direct function in the RSP. Some of the downstream transcription factors that DIMM activates are also known transcriptional repressors. This would be another point of regulation of the transcriptome. Instead of acting directly as a transcriptional repressor, DIMM could activate specific transcriptional repressors that would then inhibit certain pathways or genes, whose function might directly interfere with the expansion and operation of a robust RSP. The study of Hamanaka et al. (2010) showed that DIMM promoted the peptidergic cell fate at the expense of the classic neurotransmitter cell fate. The molecular arm of this repression

was never identified and work presented in Chapter 3 provides the first clues as to how this might be achieved. Furthermore, it is possible that DIMM activates several transcriptional repressors that function in negative feedback loops, similar to some members of the clock system (Nitabach and Taghert 2008). Two of the six transcription factors targeted by DIMM are already known to participate in negative feedback loops (Matsumoto et al. 2007; Kadener et al. 2007).

Neuropeptides are not downstream of DIMM

Analysis of DIMM targets in Chapter 3 shows that neuropeptides are not direct DIMM targets. Allan et al. (2005) had put forward the hypothesis that DIMM performs two related functions in individual DIMM cells: that DIMM directly activates its exclusive gene targets (like PHM) and that DIMM cooperates with different sets of other transcription factors to produce codes specific to activate different neuropeptide genes. Hamanaka et al. (2010) proposed an alternative hypothesis – that DIMM does not participate in neuropeptide (cargo) gene activation, but only in the support of the "peptidergic secretory machinery." My systematic analysis now fully supports and validates the latter hypothesis. Work in Chapter 3 is one of the first demonstrations that scaling factors such as DIMM act on whole subcellular compartments instead of on the specific cargo inside those compartments. This is well in agreement with the idea of scaling factor action. Scaling factors are thought to act on whole subcellular compartments, regardless of the particular cell subtype. The idea is that many heterogeneous types of cells are found even within a single tissue. For example, DIMM-positive cells express at least 17 different neuropeptides, and likely more (Park et al. 2008). Therefore, LEAP cells are highly heterogeneous with respect to their peptidergic identity, but they share the ability to make, store and release large amounts of neuropeptides. DIMM ensures this ability, but does not act as the master selector of peptidergic identity. Instead, there are likely

several other transcription factors that act combinatorially to select appropriate neuropeptides for expression in the dedicated cell type. This ensures that all LEAP cells have the cellular capacity for neuropeptide secretion, and can appropriately select the correct neuropeptide(s) for expression, and ultimately, for their dedicated NE function.

DIMM activates the major neuropeptide biosynthetic enzymes

Results from Chapter 3 also demonstrated that DIMM activates all major neuropeptide biosynthetic enzymes. Although this finding is somewhat intuitive (meaning it would have been difficult to explain if it had not been true), demonstrating it, and demonstrating the extent to which it is true among all possible genes encoding modifying enzymes remains a valuable contribution. In the case of peptidergic neurons, the biosynthetic and processing enzymes are located in the trans-Golgi and also packaged within immature LDCVs, where they operate directly on peptide precursors. Hamanaka et al. (2010) elegantly demonstrated DIMM's capacity to instruct cells to make ectopic LDCVs. DIMM, along with Mist1, is the only known transcription factor able to scale up LDCVs inside a cell. Based on Chapter 3 results, one can predict that the ectopic LDCVs produced inside photoreceptors contain all major neuropeptide biosynthetic enzymes (Hamanaka et al. 2010). The number of known or putative enzymes controlled by DIMM is five, and includes both ones whose roles in neuropeptide biosynthesis are well-established (like Prohomone convertase 2) as well as ones not previously described as such (e.g., *slamdance*; Zhang et al. 2002).

Unexpected roles of DIMM targets: endocytosis and RNA regulation

Amongst unexpected findings in Chapter 3, the inclusion of genes involved in endocytosis and RNA metabolism as direct DIMM targets were particularly novel. The endocytosis genes are intriguing because they are the first link between DIMM and processes that could be occurring after massive neuropeptide release (suggested to

normally occur in LEAP cells). Changes occurring at the cell membrane during prolonged and copious release of neuropeptides are not trivial as the cellular contents and architecture are rapidly changing. Such a cell likely needs to replenish numerous structural components, as well as major signaling factors, and orchestrate a highly choreographed re-capture of LDCV membrane from the plasma membrane for recycling. From work on classic neurons, it is known that endocytosis is one way of replenishing critical components that are lost during synaptic transmission (Südhof 2007). Following exocytosis, synaptic terminals must rapidly replenish their vesicle pool by locally recycling synaptic vesicles (Zhang 2003). Recent work has shown that in many synapses, exocytosis of neurotransmitter is coupled to endocytosis, and that synapses have evolved a specialized apparatus of scaffolding proteins to comply with these demands (Haucke et al. 2011). Exocytosed synaptic vesicle membrane proteins and lipids are recycled at the periactive zone that surrounds the release site, in order to restore functional synaptic vesicle pools for reuse and to ensure long-term functionality of the synapse (Haucke et al. 2011). It is therefore entirely possible that cellular sites of NE cargo release have similar requirements for the replenishment of basic building blocks necessary for their type of neurotransmission. DIMM could play a role in this process by directing transcription of proteins involved in endocytosis of specific cargo, which may be particularly important for the proper functioning of LEAP cells. By virtue of this significant target list, my work is the first to call attention to the fundamental importance of LDCV endocytosis in the biology of peptidergic neurons.

RNA regulation is another process that was not previously implicated to be included in DIMM action mechanism. Several RNA-regulatory factors were identified as major DIMM targets in Chapter 3. How regulation of various RNA species affects LEAP cell physiology is entirely unknown at this point. There are several examples of bHLH

transcription factors activating RNA-binding proteins with key roles in the processes controlled by the bHLH in question. One such example is the bHLH MyoD and its RNA-binding protein target Seb4/RBM24 (Li et al. 2010). MyoD activates RBM24 in the developing *Xenopus* embryos and loss of RBM24 produces a phenotype similar to a MyoD dominant negative mutant (Li et al. 2010). Therefore, the MyoD target gene RBM24 appears to be required for expression of myogenic genes in the frog embryo (Li et al. 2010). Another example comes from the mammalian bHLH NeuroD and its RNA-binding protein target SRp38 (Liu and Harland 2005). NeuroD activates SRp38 and SRp38 inhibits neuronal differentiation at a step between neurogenin and neuroD activity (Liu and Harland 2005). SRp38 inhibits neural differentiation in the *Xenopus* embryo but it does not affect neural induction or competence (Liu and Harland 2005). SRp38 is a known mitotic splicing repressor, but it may also act by regulating ribosome biogenesis, via its binding to the 28S rRNA (Liu and Harland 2005). The examples of RBM24 and SRp38 illustrate that RNA-binding proteins that are transcriptionally activated by certain bHLHs may play important roles in biologic processes controlled by the bHLH in question. What is interesting about DIMM-targeted RNA-binding proteins is the fact that LEAP cells are a mature, post-mitotic lineage. Therefore, the RNA-binding proteins that DIMM targets do not function in the context of a developing, highly mitotic lineage. Nevertheless, the intracellular environment in NE cells might be similarly dynamic because NE cells are required to integrate many environmental stimuli and to produce long-lasting effects on the whole organism. It is possible that due to the highly complex nature of LDCV formation, storage, traffic and release, particular RNAs must be precisely modified with respect to their half-lives, biochemical modifications, translation and localization.

The value of establishing all major gene targets for a regulatory transcription factor

Establishing a particular transcription factor's gene regulatory interactome and interpreting these results in the context of a cell-specific transcriptome are important for various reasons. First, this will help us understand how a particular transcription factor acts inside a cell in order to accomplish its various functions. Specifically, for a scaling factor, this will allow us to understand how whole subcellular compartments are constructed inside of cells. Second, interactome / transcriptome identification is helpful because we currently lack complete models of transcription factor action in specific cell types, as opposed to whole tissues. Publicly supported group projects such as modENCODE and ENCODE have understandably focused on profiling whole tissues. Cell-type specific models of transcription factor actions are useful especially in evolutionary contexts: do Drosophilids use more or less complicated gene regulatory hierarchies to accomplish similar tasks as mammals? Does a fly scaling factor accomplish the same scaling effect as a homologous mammalian scaling factor by using the same types of tools, or by using completely different tools? If so, do both types of scaling factors arrive at their goal by completely different methods? Or is the gene regulatory logic required for scaling conserved between mammals and flies? Although these questions might seem esoteric, they have real life applications. For example, beta cells have LDCVs and a robust RSP. Attempts at understanding beta cell biology have been ongoing for decades, but we are still unable to effectively repair a defective RSP inside these cells. Diabetes is rapidly escalating as a 21st century epidemic, yet, we do not know understand in great detail how beta cells are constructed to make insulin.

Section on Future Directions.

Elucidating LEAP cell LDCV content and uniformity by ultrastructural methods

Transmitters are stored in vesicles and granules that have transmitter-specific morphological characteristics. For example, glutamate-filled and acetylcholine-containing vesicles are round, small (~35 nm in diameter) and display a clear core (Watson and Schürmann 2002; Honda and Semba 1995). GABA-containing vesicles are often small and clear, but flattened (Fabian-Fine et al. 2000; Hamori et al. 1990). Amine containing granules are slightly larger and have a characteristic dense core (Shkolnik and Schwartz 1980). Peptide-containing granules are significantly larger (80-200 nm) and also display a dense core (Borgonovo et al. 2006; Crivellato et al. 2005; Crivellato et al. 2006; Edwards 1998). Overall, LDCVs tend to be at least twice the size of small synaptic vesicles (Bruns et al. 2000). Given the association of DIMM with the biology of peptide LDCVs (Hamanaka et al. 2010), it is natural to inquire about the relationships between the individual genes that DIMM activates and the production, stabilization, trafficking, accumulation, release and or endocytosis of LDCVs. If DIMM instructs cells how to scale up the LDCV compartment, an obvious question pertains to the nature of the LDCVs formed. If DIMM acts through a single mechanism, the implicit assumption is that it will instruct the cells to produce a single type of LDCV. Since LDCVs are a subcellular structure, they can only be effectively visualized by electron microscopy (EM). It could therefore be useful to assess the consequences of DIMM action at the ultrastructural level by obtaining EM images of purified LEAP cells. The number, size and characteristics of LDCVs could be assessed in detail from EM images. If the prediction is true, LEAP cells should have a fairly homogenous pool of LDCVs that are produced in response to DIMM action. An interesting correlate is that peptide-containing granules in non-DIMM cells may have a distinct category of LDCVs (morphologically, biochemically, physiologically distinct). On the other hand, LEAP cell LDCVs could be heterogeneous if their underlying peptide cargo has important consequences on the

size, shape and characteristics of the LDCVs that house the cargo. Just such a relationship between cargo content and LDCV morphology has been proposed for the case of granules containing the Atrial Natriuretic Factor (Baertschi et al. 2001).

Identifying DIMM loss-of-function transcriptome

In Chapter 3, I identified the normal LEAP cell transcriptome. While this is a useful first step in assigning ChIP-chip peaks to genes, its value will be even grater once it can be compared to the gene expression profile of LEAP cells that have acutely lost DIMM expression. This loss-of-function gene expression profile would be a snapshot of changes in transcript levels that occur with loss of DIMM support and help interpret the full list of DIMM targets identified in Chapter 3. The loss-of-function transcriptome would allow one to check whether or not LEAP cells lose expression of identified DIMM target genes when they lose DIMM expression. Furthermore, it is likely that numerous other genes will have perturbed levels, as an indirect / downstream consequence of DIMM loss. Changes in the expression of these genes would tell us about what downstream processes DIMM controls. This would be helpful in case that DIMM acts not only as a scaling factor, but also in other, as of yet, unknown ways. Furthermore, the DIMM loss-of-function gene expression would be helpful in identifying any genes that DIMM might repress. In this case, such genes would be expected to be upregulated upon DIMM loss.

REFERENCES

Allan, D.W., Park, D., St Pierre, S.E., Taghert, P.H., and Thor, S. (2005). Regulators acting in combinatorial codes also act independently in single differentiating neurons. *Neuron* **45**, 689–700.

Baertschi, A.J., Monnier, D., Schmidt, U., Levitan, E.S., Fakan, S., and Roatti, A. (2001). Acid Prohormone Sequence Determines Size, Shape, and Docking of Secretory Vesicles in Atrial Myocytes. *Circ. Res.* **89**, e23–e29.

Borgonovo, B., Racchetti, G., Malosio, M., Benfante, R., Podini, P., Rosa, P., and Meldolesi, J. (1998). Neurosecretion competence, an independently regulated trait of the neurosecretory cell phenotype. *J. Biol. Chem.* **273**, 34683-34686.

Borlikova, G., and Endo, S. (2009). Inducible cAMP Early Repressor (ICER) and Brain Functions. *Mol. Neurobiol.* **40**, 73–86.

Bruns, D., Riedel, D., Klingauf, J., and Jahn, R. (2000). Quantal release of serotonin. *Neuron* **28**, 205–220.

Crivellato, E., Nico, B., Bertelli, E., Nussdorfer, G.G., and Ribatti, D. (2006). Dense-core granules in neuroendocrine cells and neurons release their secretory constituents by piecemeal degranulation (review). *Int. J. Mol. Med.* **18**, 1037-1046.

Crivellato, E., Nico, B., and Ribatti, D. (2005). Ultrastructural evidence of piecemeal degranulation in large dense-core vesicles of brain neurons. *Anat. Embryol.* **210**, 25-34.

Della Fazia, M.A., Servillo, G., Foulkes, N.S., and Sassone-Corsi, P. (1998). Stress-induced expression of transcriptional repressor ICER in the adrenal gland. *FEBS Lett.* **434**, 33–36.

Edwards, R.H. (1998). Neurotransmitter release: variations on a theme. *Curr. Biol.* **8**, R883–5.

Fabian-Fine, R., Meinertzhagen, I.A., and Seyfarth, E.A. (2000). Organization of efferent peripheral synapses at mechanosensory neurons in spiders. *J. Comp. Neurol.* **420**, 195–210.

Foulkes, N.S., Borjigin, J., Snyder, S.H., and Sassone-Corsi, P. (1996). Transcriptional control of circadian hormone synthesis via the CREM feedback loop. *Proc. Natl. Acad. Sci. USA* **93**, 14140–14145.

Foulkes, N.S., Borjigin, J., Snyder, S.H., and Sassone-Corsi, P. (1997). Rhythmic transcription: the molecular basis of circadian melatonin synthesis. *Trends Neurosci.* **20**, 487–492.

Givan, A.L. (2011). Flow cytometry: an introduction. *Methods Mol. Biol.* **699**, 1–29.

Hamanaka, Y., Park, D., Yin, P., Annangudi, S.P., Edwards, T.N., Sweedler, J., Meinertzhagen, I.A., and Taghert, P.H. (2010). Transcriptional orchestration of the regulated secretory pathway in neurons by the bHLH protein DIMM. *Curr. Biol.* **20**, 9-18.

Hamori, J., Takacs, J., and Petrusz, P. (1990). Immunogold electron microscopic demonstration of glutamate and GABA in normal and deafferented cerebellar cortex: correlation between transmitter content and synaptic vesicle size. *J. Histochem. Cytochem.* **38**, 1767–1777.

Harada, R., and Nepveu, A. (2012). Chromatin affinity purification. *Methods Mol. Biol.* **809**, 237–253.

Haucke, V., Neher, E., and Sigrist, S.J. (2011). Protein scaffolds in the coupling of synaptic exocytosis and endocytosis. *Nat. Rev. Neurosci.* **12**, 127–138.

Hempel, C.M., Sugino, K., and Nelson, S.B. (2007). A manual method for the purification of fluorescently labeled neurons from the mammalian brain. *Nat. Protoc.* **2**, 2924–2929.

Honda, T., and Semba, K. (1995). An ultrastructural study of cholinergic and non-cholinergic neurons in the laterodorsal and pedunculopontine tegmental nuclei in the rat. *Neuroscience* **68**, 837–853.

Kadener, S., Stoleru, D., McDonald, M., Nawathean, P., and Rosbash, M. (2007). Clockwork Orange is a transcriptional repressor and a new Drosophila circadian pacemaker component. *Genes Dev.* **21**, 1675–1686.

Kolodziej, K.E., Pourfarzad, F., de Boer, E., Krpic, S., Grosveld, F., and Strouboulis, J. (2009). Optimal use of tandem biotin and V5 tags in ChIP assays. *BMC Mol. Biol.* **10**, 6.

Li, H.-Y., Bourdelas, A., Carron, C., and Shi, D.-L. (2010). The RNA-binding protein Seb4/RBM24 is a direct target of MyoD and is required for myogenesis during Xenopus early development. *Mech. Dev.* **127**, 281–291.

Liu, K.J., and Harland, R.M. (2005). Inhibition of neurogenesis by SRp38, a neuroD-regulated RNA-binding protein. *Development* **132**, 1511–1523.

Matsumoto, A., Ukai-Tadenuma, M., Yamada, R. G., Houl, J., Uno, K. D., Kasukawa, T., Dauwalder, B., Itoh, T. Q., Takahashi, K., Ueda, R., et al. (2007). A functional genomics strategy reveals clockwork orange as a transcriptional regulator in the Drosophila circadian clock. *Genes Dev.* **21**, 1687–1700.

Mazzoni, E.O., Mahony, S., Iacovino, M., Morrison, C.A., Mountoufaris, G., Closser, M., Whyte, W.A., Young, R.A., Kyba, M., Gifford, D.K., et al. (2011). Embryonic stem cell–based mapping of developmental transcriptional programs. *Nat. Meth.* **8**, 1056–1058.

McGuire, S.E., Le, P.T., Osborn, A.J., Matsumoto, K., and Davis, R.L. (2003). Spatiotemporal rescue of memory dysfunction in *Drosophila*. *Science* **302**, 1765-1768.

Nitabach, M.N., and Taghert, P.H. (2008). Organization of the Drosophila circadian control circuit. *Curr. Biol.* **18**, R84–93.

Park, D., and Taghert, P.H. (2009). Peptidergic neurosecretory cells in insects: organization and control by the bHLH protein DIMMED. *Gen. Comp. Endocrinol.* **162**, 2-7.

Persengiev, S.P., and Green, M.R. (2003). The role of ATF/CREB family members in cell growth, survival and apoptosis. *Apoptosis* **8**, 225–228.

Powell, L.M., and Jarman, A.P. (2008). Context dependence of proneural bHLH proteins. *Curr. Opin. Genet. Dev.* **18**, 411–417.

Shkolnik, L.J., and Schwartz, J.H. (1980). Genesis and maturation of serotonergic vesicles in identified giant cerebral neuron of Aplysia. *J. Neurophysiol.* **43**, 945–967.

Südhof, T.C. (2007). Neurotransmitter Release. *Handb. Exp. Pharmacol.* **184**, 1-21.

Watson, A.H.D., and Schürmann, F.-W. (2002). Synaptic structure, distribution, and circuitry in the central nervous system of the locust and related insects. *Microsc. Res. Tech.* **56**, 210–226.

Zhang, B. (2003). Genetic and molecular analysis of synaptic vesicle recycling in Drosophila. *J Neurocytol.* **32**, 567–589.

Zhang, H., Tan, J., Reynolds, E., Kuebler, D., Faulhaber, S., and Tanouye, M. (2002). The Drosophila slamdance gene: a mutation in an aminopeptidase can cause seizure, paralysis and neuronal failure. *Genetics* **162**, 1283-1299.