CIS-REGULATORY INTEGRATION OF INTRINSIC AND TARGET-DEPENDENT REGULATORS IS REQUIRED FOR TERMINAL DIFFERENTIATION OF *DROSOPHILA* NEURONS

by

ANTHONY JASON ERICH BERNDT

BHSc. Hons., University of Calgary, 2007

A THESIS SUBMITTED IN PARTIAL FULLFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Cell and Developmental Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

October 2015

ABSTRACT

Terminal differentiation of neurons often requires retrograde signals from the target cells they innervate, which trigger neural subtype-specific gene expression upon target contact. Target-derived BMP signaling and transcription factors including the LIM-Homeodomain transcription factor Apterous are required for FMRFa neuropeptide gene expression in *Drosophila* Tv4 neurons. We modeled the integrative mechanism of these extrinsic and intrinsic inputs at a Tv4 neuron-specific FMRFa enhancer. We show that Tv4-specific *FMRFa* expression requires two separable *cis*-elements, a BMP-response element (BMP-RE) that binds Mad, and a homeodomain response element (HD-RE) that binds Apterous. Strikingly, we find that concatemers of these two short (~30bp) *cis*-elements each independently drive spatial and temporal expression appropriate for Tv4-specific FMRFa. Thus, specific and robust expression is generated from the synergy of two low-activity heterotypic *cis*-elements that encode the same output from distinct inputs. We further examined the timing mechanism of *FMRFa* initiation, which models predict would be solely based on target contact. In contrast, we find that the timed downregulation of the COUP-TF I/II nuclear receptor Seven up functions to de-repress *HD-RE* and *BMP-RE* activity immediately prior to target contact. Thus, we reveal that the active suppression of neurotransmitter identity, prior to target contact, is an innate component of the target-dependent mechanism for timed gene activation.

Further examination of the FMRFa BMP-RE shows that the sequence of the Mad and Medea binding sites to be "perfectly wrong." It displays sequence similarity to both the previously characterized BMP Silencing Element (SE) and Activating Elements (AE) but with base substitutions that should abrogate Smad binding. Biochemical and reporter construct analysis demonstrate that the FMRFa BMP-RE is an atypical activating element that is specifically

attenuated in its ability to interact with the co-Smad Medea. *In vivo* reporter assays show that this attenuation is required to drive appropriate expression in the CNS. These findings represent only the third verified type of BMP-RE found in *Drosophila* leading us to call this class an AE2 element.

PREFACE

All worked was conducted in Dr. Douglas Allan's Laboratory at UBC.

Chapter 2: "*Cis*-regulatory integration of intrinsic and target-dependent regulators is required for terminal differentiation of neurons"

Work in this chapter is currently under review:
"Cis-regulatory integration of intrinsic and target-dependent regulators is required for terminal differentiation of neurons."
Anthony J.E. Berndt*, Jonathan C.Y. Tang*, Marc S. Ridyard, Tianshun Lian, Kathleen Keatings and Douglas W. Allan
*= co-first authorship

For this publication I performed all experiments except the genetic loss of function in Figure 9 and Figure 10 which were performed by a former lab Research Associate, Dr. Marc Ridyard. Jonathan Tang produced the YFP vector expression backbone and some of the initial point mutations in the FMRFa enhancer while a Master's student in the Allan Lab. I produced all deletion and concatamerization mutations and their associated data as well as identified the BMP response element and showed that Svp is a timing mechanism for FMRFa expression. Gel shift assays were performed by the lab technician, Tianshun Lian. I performed all data analysis. Kathleen Keatings assisted with dissections for the squeeze mutant verification and ectopic expression data found in supplemental figures. Douglas Allan and I conceived the experiments and methods of analysis, as well as co-wrote the manuscript.

Chapter 3: "Characterization of an atypical BMP response element."

Douglas Allan and I conceived of all the experiments and methods of analysis. I produced the pThunderbird GFP expression vector and I performed all *in vivo* experiments and analysis presented except for *brk*, *shn*, and *wit* loss of function in Figure 15. Tianshun Lian performed the EMSA assays. I performed all data analysis. Douglas Allan and I conceived the experiments and methods of analysis, as well as co-wrote the forthcoming manuscript.

Appendix A: "Mutation and initial Characterization of *Drosophila* LMX orthologues"
Douglas Allan and I conceived of all the experiments. I performed the transposon-based mutagenesis and crosses while a former lab Co-op student, Simaran Kalkat, assisted with PCR screening of the *CG32105* alleles. Undergraduate research assistants Shirley Guan and Raymond Cho assisted with the egg laying assays. Former lab graduate student, Monica Castellanos assisted with oviduct dissections.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS & ABBREVIATIONS

AE	Activating Element
ap	*apterous*
BMP	Bone Morphogenetic Protein
CGRP	Calcitonin Gene Related Peptide
ChIP	Chromatin Immunoprecipitation
CNS	Central Nervous System
cas	*castor*
col	*collier*
COUP-TF	Chicken Ovalbumin Upstream Promoter Transcription Factor
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
dac	*dachshund*
Dad	*Daughters against dpp*
Df	Deficiency
dILP	*Drosophila* Insulin-like peptide
dimm	*dimmed*
Dlg	Discs-large
DNO	Dorsal Neurohaemal Organ
Dpp	Decapentaplegic
DSB	Double-Strand Breaks
EGFP	Enhanced Green Fluorescent Protein
EJP	Excitatory Junction Potential
EMSA	Electrophoretic Mobility Shift Assay
eya	*eyes absent*
EYFP	Enhanced Yellow Fluorescent Protein
Flp	Flippase
Gbb	Glass bottom boat
GEF	Guanine Exchange Factor
GFP	Green Fluorescent Protein
grh	*grainy head*
hb	*hunchback*

HD	Homeodomain
HDR	Homology Directed Repair
HLH	Helix-Loop-Helix
HRP	Horseradish Peroxidase
Kr	*Kruppel*
Mad	*Mothers against dpp*
NHEJ	Non-Homologous End Joining
NMJ	Neuromuscular Junction
NPLP1	Neuropeptide-Like Precursor 1
PACAP	Pituitary Adenylate Cyclase Activating Polypeptide
pdm	*POU domain protein (also known as nubbin)*
RE	Response Element
RNA	Ribonucleic Acid
SE	Silencer Element
Seq	Sequencing (usually referring to high throughput methods)
sog	*short gastrulation*
sqz	*squeeze*
svp	*seven up*
TBH	Tyramine Beta-Hydroxylase
td	Tandem dimeric
TDC	Tyrosine decarboxylase
tkv	*thickveins*
twit	*target of wit*
UTR	Untranslated Region
VIP	Vasoactive Intestinal Peptide
VNC	Ventral Nerve Cord
wit	*wishful thinking*
WT	Wild Type

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Doug Allan. He took me in to his lab on a gamble... hopefully it has paid off for him. The past six years have pushed me to grow both as a scientist and as a person. I've been fortunate to have such a supportive mentor who has given me the freedom to explore some very hard questions and upheld an uncompromising approach to high quality science.

Secondly, I would like to thank the colleagues I have had the privilege of working with in the Allan lab over the years. In particular, Monica Castellanos, Tianshun Lian , Robin Vuilleumier, Kathleen Keatings, Kevin Eade, Jonathan Tang, Luba Veverytsa, and Marc Ridyard, Simaran Kalkat, Sarah Garner, Raymond Cho, Justin Fong, and Seo Am Hur. It's rare that you can find a lab where you want to show up not only because the science is good but because the people make you want to be there. In the greater LSI community: Andrew Wong, Stephano Brigidi, and Mary Pines- your feedback and insight as outsiders to the field has been invaluable. And, especially, Jacob Hodgson- thanks for always taking the time to "talk shop," the invaluable technical advice, and help with putting the ‘Ph’ in my Ph.D.

Thirdly, I would like to thank a very supportive committee: Don Moerman, Tim O'Connor, and Michael Gordon. Through some magic you found time in your busy schedules to give me helpful feedback and mentorship over the years.

DEDICATION

To Friends, Mentors, and Labmates who have always pushed me to do better today than I did yesterday.

To my Mom, Dad, and Brother for their unconditional support and encouragement. My success would not have been possible without you.

1 INTRODUCTION

1.1 Neuronal Differentiation

A central aim of developmental biology focuses on determining how the diversity of cell types found in multicellular organisms arise from a single fertilized egg. The nervous system displays potentially the highest degree of cellular diversity and specialization, easily evident with regard to cellular morphology or physiology. Of specific relevance to this thesis is the genetic and signaling mechanisms that direct the subtype-specific expression of terminal differentiation genes that define a neuron's function and identity. These so-called terminal differentiation genes include those genes that carry out the specialized functions of a neuron; the neuropeptides, receptors ion channels, neurotransmitter biosynthetic enzymes (Hobert et al., 2010). Throughout the thesis, I refer to the genetic and signaling mechanisms as intrinsic and extrinsic regulatory inputs, respectively. Intrinsic regulators refer to the TFs expressed by the neuroblast lineage and the postmitotic neuron. Extrinsic regulatory inputs refer to signaling pathways that alter gene expression, such as activated by hormones or morphogens.

The *Drosophila* central nervous system can be subdivided into the brain and the ventral nerve cord (VNC). I will discuss VNC neurogenesis here. A sheet of neuroectodermal cells located in the ventrolateral region of the embryo gives rise to the VNC. Within this neuroectodermal sheet, the combined action of maternal, gap, pair-rule and segment polarity genes classes patterns the anteroposterior axis to define segmental boundaries and intrasegmental identity. Each segment is further subdivided along the dorsoventral axis by the genes *ventral nervous system defective (vnd), intermediate neuroblasts defective (ind)* and *Drop (Msh*) (Reviewed in: (Skeath and Thor, 2003)). Together, these patterning genes lay down a

"checkerboard" organization that identifies positional identity throughout the entire VNC (See **Figure 1 A-B).** Neuroblasts (NBs) emerge from within this checkerboard-patterned neuroectoderm and acquire unique fates depending on their anteroposterior and dorsoventral position within this field of developmental cues (Reviewed in (Skeath and Thor, 2003)). In five sequential waves, NBs delaminate from the neuroectoderm, resulting in a nearly invariant pattern of 60 NBs per VNC segment (Bossing et al., 1996; Schmid et al., 1999), except for within abdominal segments A8-A10 which have fewer NBs (Birkholz et al., 2013).

Each NB acts as a CNS-specific stem cell. Generally, VNC NBs divide asymmetrically to produce another NB and a progenitor cell called a Ganglion Mother Cell (GMC), in so-called Type I proliferation mode. Very generally, the GMC divides once to produce two postmitotic neurons and/or glia depending on the GMC's unique developmental program. However, other neuroblast division programs exist, including type 0 and type II proliferation (Li et al., 2013). For example, the Tv cluster neurons discussed below arise from a type 0 proliferative mode (Baumgardt et al., 2014), as detailed below (**See Figure 1D**).

During NB lineage progression, the spatial patterning that diversifies neuroblast identities is further elaborated by temporal patterning cues that change throughout the neuroblast lineage. In the VNC, most neuroblast lineages express a temporal sequence of up to five 'temporal' transcription factor genes; termed temporal selectors (Allan and Thor, 2015). In their order of expression within lineages, these are Hunchback → Kruppel → Pdm (nubbin and Pdm2) → Castor → Grainy head (Hb→Kr→Pdm→Cas→Grh) (Brody and Odenwald, 2000; Isshiki et al., 2001). When a neuroblast divides asymmetrically into a GMC and another neuroblast, the temporal factor is retained in the GMC and usually the neurons born from that GMC, whereas

the newborn neuroblast may proceed to change the temporal factors(s) expressed (Brody and Odenwald, 2000; Tran and Doe, 2008) (**See Figure 1D**). Within a given neuroblast lineage, the expression of a specific temporal transcription factor corresponds to a change in the type(s) of neurons/glia generated. Within the well-studied neuroblast 7-1 lineage, *hb* is necessary and sufficient to specify U1 and U2 motoneuron fates, while Kr is necessary and sufficient to specify the U3 motoneuron fate (Isshiki et al., 2001). Between lineages, the same temporal transcription factor generates different neuronal types. For example, the Kr-expression GMC in the neuroblast 7-3 lineage produces a serotonergic interneuron, while the Kr-expressing GMC in the neuroblast 7-1 lineage produces the U3 motoneuron (Isshiki et al., 2001). The temporal specifiers *cas* and *grh* are expressed in the Tv cluster and play key roles in specifying the Tv cluster neurons (detailed in Section 1.4).

The temporal selector TFs are cross-regulatory. For example, in *hb* mutants, Kr expression can be observed precociously. Prolonged expression of Pdm can be observed in *cas* mutants. Gain-of-function studies have shown that precocious expression of a temporal selector factor (n) activates the next transcription factor in the cascade (n+1) while repressing the next one thereafter (n+2) (Isshiki et al., 2001) (See **Figure 1D** and **Figure 4A**). However, temporal selector genes are not always necessary for the expression of the next gene in the cascade, rather they modulate the timing of their expression. For example in *pdm* mutants where there is a delay in *cas* expression in some (NB3-1), but not all (NB7-1) lineages (Grosskortenhaus et al., 2006; Tran and Doe, 2008). By comparison *cas* is necessary for the expression of *grh* (Maurange et al., 2008).

An additional level of regulatory control of temporal selector factors exists. The Hb to Kr switch is regulated by 'switching transcription factors' Seven up (Svp), Distal antenna (*dan*) and

Distal antenna related (Kanai et al., 2005; Kohwi et al., 2011). These factors are required for the timely downregulation of Hb. Loss of *svp* function leads to duplication of early born Hb-expressing progeny (Kanai et al., 2005). This is discussed further in Sections 1.4 due to its relevance to Tv neuron generation.

While spatial and temporal selectors explain much of how neuroblasts generate distinct cell types, even this combination of neuronal identity coding is insufficient to explain the immense diversity of cells in the *Drosophila* CNS. One such additional layer of diversification is mediated by Notch signaling, which is common to many NB lineages. Generation of differences in daughter postmitotic cell fates is typically dependent on Notch signalling which acts antagonistically with the activity of Numb (Knoblich, 2008; Spana and Doe, 1996). Upon GMC division, Numb segregates into one daughter cell. The Notch ligand, Delta, is found in the underlying mesoderm and on NBs themselves (Spana and Doe, 1996). Daughter postmitotic cells with active Notch signalling (Numb-negative) acquire what is known as an A-cell fate. Repressed Notch signalling in the Numb-positive cell, the B-cell, leads to an alternative cell fate being adopted (Cau and Blader, 2009; Skeath and Thor, 2003). Lack of Notch signalling leads to both cells adopting a B-cell fate. For example, in the MP2 lineage, where two sibling daughter cells are generated, the vMP2 and dMP2, loss of Notch results in sibling cell transformation of the vMP2 in to a dMP2-like cell. In *Notch*; *numb* double mutants both MP2 daughter cells acquire the dMP2 fate (Spana and Doe, 1996). This suggests that the primary role Numb plays is in antagonizing Notch activity rather than acting in an instructive manner for cell identity.

One of the ways in which further diversification of neuronal identity is attained is through extrinsic cues from the axonal targets of neurons, which will be explored in the next section.

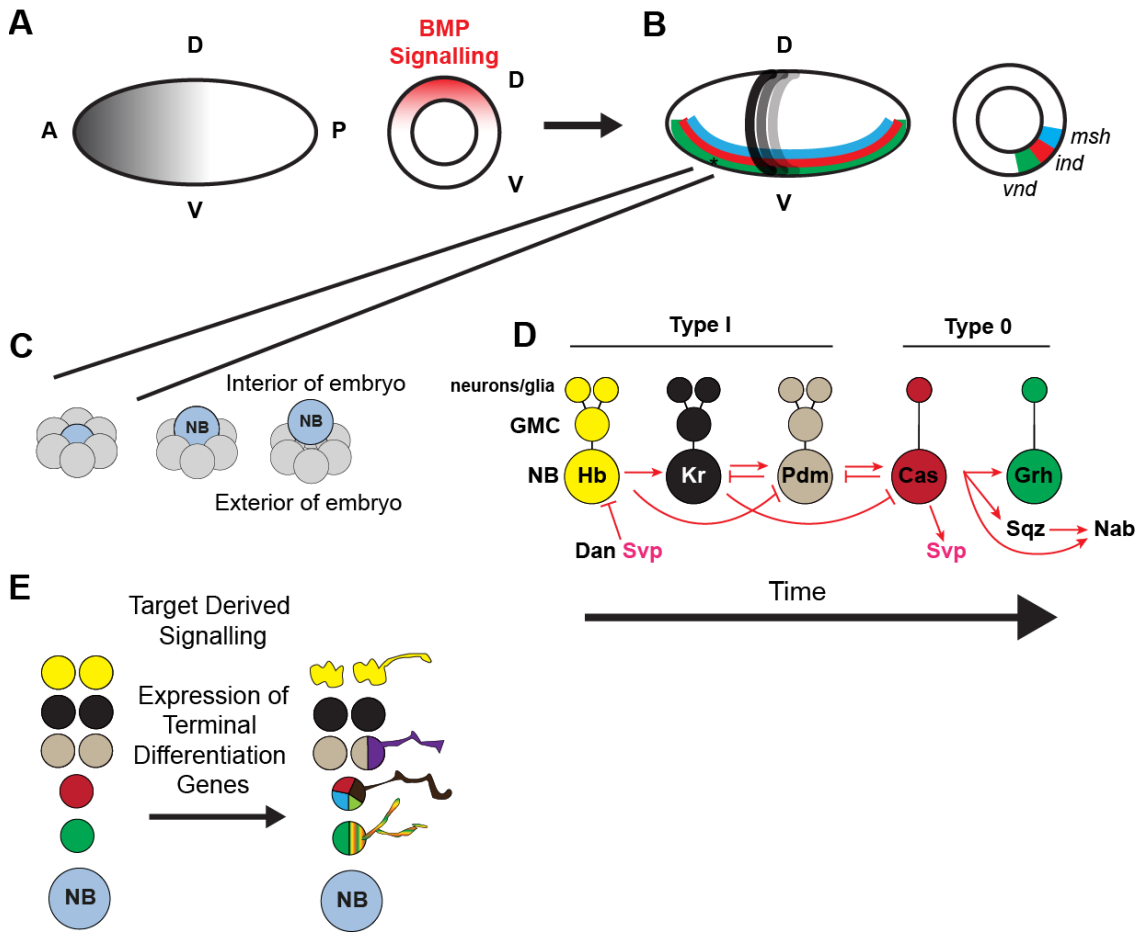


Figure 1. Development of the *Drosophila* VNC.

(**A**) Complex cascades of dorsoventral and anteroposterior patterning genes lead to the **(B)** expression of segment and segment-polarity (black-gray) and columnar (green-yellow-blue) genes in discrete stripes in each segment and along the neuroectoderm. (**C-D**) Neuroblasts (NB) are specified from the neuroectoderm and undergo asymmetric divisions to generate progenitors called Ganglion Mother Cells (GMC). Temporal selector genes lead to the diversification of cells that arise from within a neuroblast lineage. Dan, Svp, and Nab help to regulate temporal selectors and further subdivide temporal windows of differentiation. (**E**) Finally, in postmitotic cells, target derived signalling and subtype defining genes such as LIM-Homeodomain transcription factors act to further diversify and specialize cell function within the VNC.

1.2 Target Derived Signalling in Neuronal Differentiation

During differentiation, neurons receive extracellular signals from their axonal targets that regulate subtype-specific gene expression in neurons. Such target-derived signals are essential for neural survival, differentiation, and normal morphology, synaptic function and plasticity from *Drosophila* to vertebrates (da Silva and Wang, 2011; Deppmann and Ginty, 2006; Hippenmeyer et al., 2004; Marques, 2005; Salinas, 2005) (reviewed in (Zhu et al., 2008).

In vertebrates, target-dependent activation of terminal differentiation gene expression occurs through a number of different target-derived signals including neurotrophins (de Nooij et al., 2013; Patel et al., 2000), cytokines (Habecker et al., 1997), and transforming growth factor-β (TGFβ) ligands, such as activins, and BMPs (Bone Morphogenetic Protein) (Ai et al., 1999; Gomis et al., 2006). In *Drosophila*, only the BMP pathway has been found to activate terminal differentiation gene expression (Allan et al., 2003; Veverytsa and Allan, 2011). In the absence of target-derived signals, most neurons are generated, survive (except in a small number of cases (Zhu et al., 2008)) and are specified appropriately into their neuronal subtype; however, they are unable to terminally differentiate and function appropriately (Allan et al., 2003; Eade and Allan, 2009; Miguel-Aliaga et al., 2004). This was best demonstrated in the case of CCAP-neurons, in which the loss of appropriate neuropeptide expression as a result of loss of retrograde BMP-signaling, results in the lethal failure of ecdysis (Veverytsa and Allan, 2012; Veverytsa and Allan, 2013).

1.3 The BMP Pathway in *Drosophila*

The reader is referred to **Figure 2** for a summary of BMP signalling and orthologous proteins in other representative model systems.

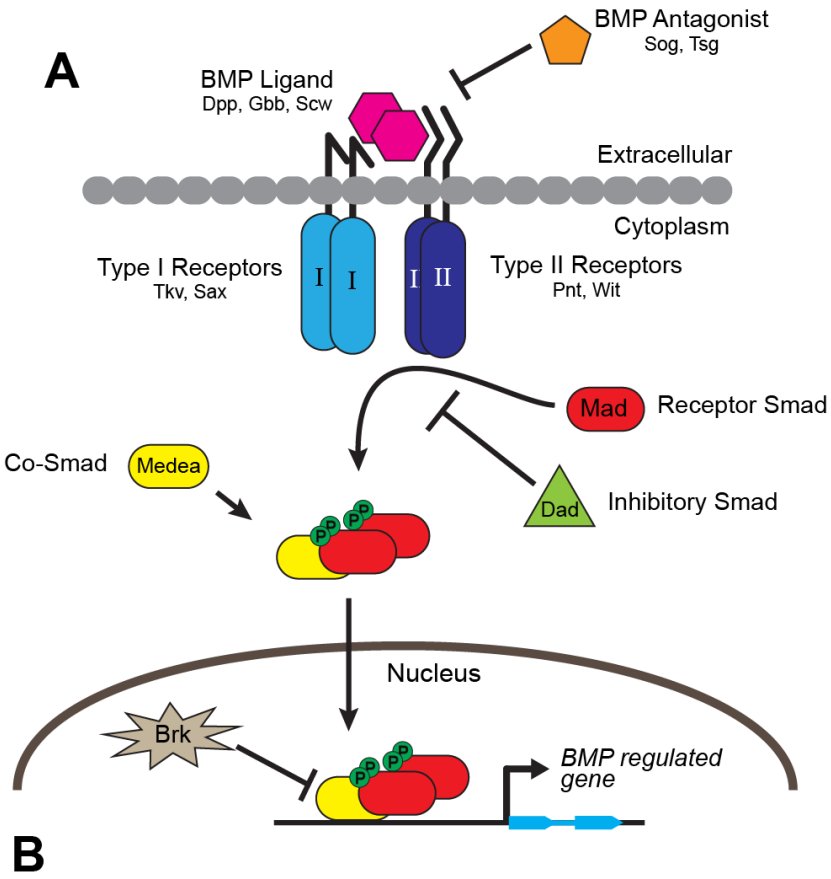
Target-derived BMP signaling is the only pathway known, to date, to be conserved from invertebrates to vertebrates for a role in target dependent neural differentiation, synaptic function and plasticity (Allan et al., 2003; Angley et al., 2003; Berke et al., 2013; Chou et al., 2013; Eade and Allan, 2009; Guha et al., 2004; Henriquez et al., 2011; Hodge et al., 2007; Kelly et al., 2013; Lopez-Coviella et al., 2005; Majdazari et al., 2013; McCabe et al., 2003; Miguel-Aliaga et al., 2008; Pavelock et al., 2007; Veverytsa and Allan, 2011; Xiao et al., 2013). The BMP pathway is grouped under the TGF-β superfamily of signalling pathways across metazoans. This superfamily includes TGF-β itself as well as Activin and Nodal (reviewed in (Kitisin et al., 2007; Peterson and O'Connor, 2014; Schmierer and Hill, 2007)). A common feature of these pathways is that a protein ligand dimer is bound by transmembrane receptor kinases that phosphorylate a downstream signalling molecule that translocates to the nucleus and acts as a sequence-specific transcription factor (See **Figure 2**). *Drosophila* has highly conserved representatives of the BMP and activin pathways (Peterson and O'Connor, 2014). As this thesis does not consider the activin pathway, I will introduce only the BMP pathway here.

A dimer of BMP ligands binds to a heterotetrameric complex of two type I and two type II BMP receptors. In *Drosophila*, ligands include *Decapentaplegic* (*dpp*), *glass bottom boat* (*gbb*), and *screw* (*scw*) (vertebrate BMP2/4, BMP 5/8 (Schmierer and Hill, 2007)). The *Drosophila* Type I BMP receptors are *thickveins* (*tkv*), *saxophone* (*sax*), and the Type II receptors are *punt* (pnt) and *wishful thinking* (*wit*) (Huminiecki et al., 2009). Ligand activation of this receptor complex activates kinase activity of the Type I receptors. This triggers signal transduction by phosphorylating the C-terminal serines of the receptor Smad (R-Smad), Mothers against Decapentaplegic (Mad) (vertebrate Smad 1/5/8) (Massague and Chen, 2000). These are named R-Smads after the founding members *C.elegans* Sma-2/3/4 genes (Savage et al., 1996)and

Drosophila Mad (Sekelsky et al., 1995) and because they are phosphorylated by the Type I BMP receptors. Two phospho-Mad (pMad) subunits bind the co-Smad, Medea (vertebrate Smad 4), to form the Smad complex (Gao and Laughon, 2006; Gao et al., 2005). The Smad complex translocates to the nucleus where Mad and Medea act as sequence-specific transcription factors.

BMP signalling can be regulated at multiple steps in vertebrates and invertebrates. Extracellular antagonists can bind to BMP ligands to prevent receptor interaction. For example, Dpp ligand availability is shaped in *Drosophila* by the BMP antagonists *short gastrulation* (*sog*) and *twisted gastrulation* (*tsg*) along the dorsoventral axis. The conserved inhibitory Smad in *Drosophila*, *Daughters against dpp* (*Dad*) (vertebrate Smad 6/7) competes with Mad for binding to the activated type I receptors to inhibit Smad phosphorylation (Inoue et al., 1998; Kamiya et al., 2008). *Dad* itself is a BMP-activated gene, thus completing a regulatory negative feedback loop (Weiss et al., 2010).

Drosophila also utilizes a repressor of BMP-responsive genes called *brinker* (*brk*), that is not conserved in vertebrates or *C. elegans*, and in addition a conserved co-repressor, *schnurri* (*shn*) (Jin et al., 2006; Yao et al., 2006). These regulators of BMP regulated genes will be discussed in Section 1.6.4 in much greater detail, since their activity is contingent on the exact sequence of Smad DNA binding motif.



BMP Signalling Component	***Drosophila***	**Mouse**	***C. elegans***
BMP Ligand	Dpp, Gbb, Scw	BMP2/4/5/8	dbl-1
BMP Antagonist	Sog, Tsg	Chordin, TWSG1	?
Type II Receptor	Pnt, Wit	ACTRIIB	daf-4
Type I Receptor	Tkv, Sax	BMPRIB	sma-6
Receptor Smad	Mad	Smad1/5/8	sma-2/3
Co-Smad	Medea	Smad4	sma-4
Inhibitory Smad	Dad	Smad6/7	tag-68
Brk	Brk	No obvious ortholog	No obvious ortholog

Figure 2. Summary of BMP signalling in *Drosophila*

(**A**) *Drosophila* BMP Pathway (**B**) Orthologs in Mouse and *C. elegans*

1.3.1 BMP Signalling Regulates Neuronal Development

Throughout neuronal development, BMP signalling plays important roles in many different cellular contexts to specify and differentiate tissue and cell fates, in both vertebrates and invertebrates (Hegarty et al., 2013; Liu and Niswander, 2005) (**Figure 3**). In both invertebrates and vertebrates, blockade of BMP signaling is critical in the patterning and induction of the neuroectoderm from other ectodermal fates (Umulis et al., 2009). In brief, BMP ligands and antagonists are produced on opposing dorsoventral sides of the early embryo; in vertebrates BMP ligand are high ventrally and low dorsally, with BMP antagonists such as chordin expressed high dorsally and low ventrally. In the *Drosophila* embryo, similar oppositional gradients are formed (O'Connor et al., 2006), only that their orientation is opposite to that of vertebrates during initial pattern of the nervous system (Holland et al., 2013) (**Figure 3A&B**). The BMP ligand, Dpp, is expressed on the dorsal side of the embryo while the BMP ligand antagonist, Short gastrulation (Sog) (antagonist) is produced ventrally (O'Connor et al., 2006). These opposing gradients result in a gradient of pMad activity from high dorsally to progressively less ventrally. Similar to CNS formation in vertebrates, the *Drosophila* CNS forms in the region of low BMP signalling. However, in vertebrates low BMP is seen dorsally and in *Drosophila*, low BMP is seen ventrolaterally. Dorsally, high Dpp signalling specifies the amnioserosa and dorsal ectoderm (See **Figure 3A**). After initial CNS formation, BMP signalling plays a further role in patterning the vertebrate neural tube. Numerous BMP ligands are secreted from the dorsal roof-plate of the developing neural tube to function as a dorsalizing morphogenic signal that acts in apposition to a sonic hedgehog (Shh) ventralizing morphogenic signal secreted from the notochord and floor-plate. These apposed morphogens are critical to the establishment of domains of transcription factor expression within neuroglial progenitors of the neural tube that specify their differential fates (Dasen, 2009).

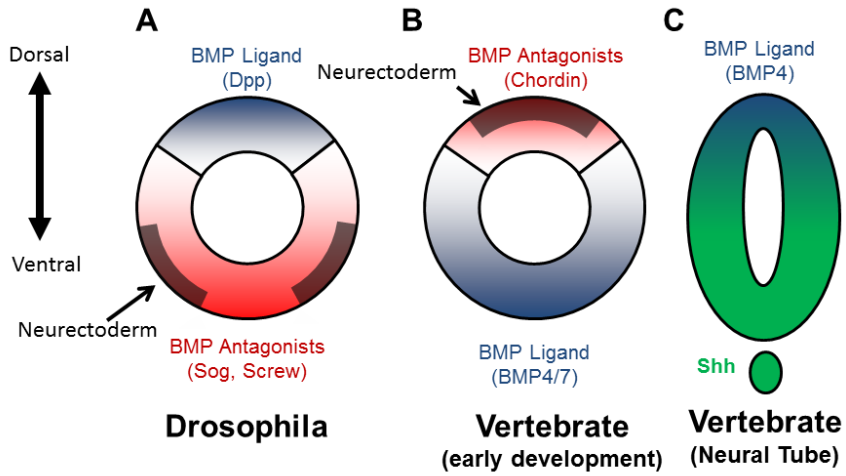


Figure 3. Summary of the early role of BMP in patterning the CNS.

(**A**) Early *Drosophila* patterning (**B**) Early vertebrate patterning. (**C**) Dorsoventral patterning of the vertebrate neural tube.

1.3.2 Target Derived BMP Signalling in the Terminal Differentiation of Postmitotic Neurons

As nervous system development progresses, BMP signalling is again used to refine the expression of genes and control neuronal identity in postmitotic neurons. BMP-signaling within the fly CNS is activated by Gbb acting at a presynaptic BMP-Receptor complex composed of Wit, Tkv, and Sax receptors that phosphorylates Mad (Allan et al., 2003; McCabe et al., 2003).

In the developing embryonic and larval *Drosophila* nervous system, null mutants in *wit* result in a total loss of BMP signaling (Aberle et al., 2002), and are lethal (Marques et al., 2003; Veverytsa and Allan, 2011). The retrograde trafficking of target-derived BMP signalling represents a common mechanism for the terminal differentiation of neurosecretory neurons and for normal synaptic function of motoneurons. It is required for the transcriptional activation of the neuropeptides FMRFa in Tv4 neurons (Allan et al., 2005; Allan et al., 2003; Eade and Allan, 2009; Miguel-Aliaga et al., 2004), Dilp7 in the dMP2 neurons (Miguel-Aliaga et al., 2008), and CCAP, MIP, and Bursicon β in CCAP neurons (Veverytsa and Allan, 2011; Veverytsa and Allan, 2012). In the case of CCAP neurons, BMP signaling is required for expression of a set of neuropeptides that are essential for pupal ecdysis, and in the absence of BMP signaling in CCAP-neurons, there is a failure of pupal ecdysis that results in lethality at the pharate adult pupal stage- pharate refers to the stage at which metamorphosis is complete but the adult animal is still in the pupal casing (Veverytsa and Allan, 2011). Such diversity in neuronal subtype-specific BMP-responsiveness is also seen in mammals. For example, retrograde BMP signaling is essential for determining TF expression profiles in trigeminal neurons (Hodge et al., 2007) regulation of levels of PACAP (pituitary adenylate cyclase-activating polypeptide) and VIP (vasoactive intestinal peptide) expression in rat superior cervical ganglia post-ganglionic

sympathetic neurons (Pavelock et al., 2007) and calcitonin gene related peptide (CGRP) in cutaneous afferent sensory neurons (Ai et al., 1999; Bucelli et al., 2008).

Disruption or loss of BMP-ligand, receptor or Smads leads to a severe reduction in the growth of synaptic boutons at the neuromuscular junction and decreased neurotransmission. In *wit* mutants, evoked excitatory junctional potential (EJP) is reduced by 90% due to 55% fewer boutons and reduced miniature excitatory junction potential amplitude and frequency (Aberle et al., 2002; Marques et al., 2002). These phenotypes have been partially accounted for by discovery of two BMP-dependent genes acting pre-synaptically in motoneurons. The Rac guanine exchange factor (GEF) *trio* displays target-derived BMP signalling dependence. Loss of *trio* in *wit* mutants contributes to the observed reduction in neuromuscular junction growth and restoration of *trio* expression in *wit* and *Mad* mutants is able to only partially rescue the neuromuscular junction growth phenotype (Ball et al., 2010). Neurotransmission phenotypes seen in *wit* mutants are not recapitulated in *trio* mutants. However, other studies had found that the *target of wit* (*twit)* gene, encoding a Ly-6 family glycophosphatidylinositol (GPI) anchored protein, is downregulated in *wit* mutants. Importantly, *twit* mutants exhibit a decrease in miniature excitatory synaptic potential amplitude (Gao et al., 2005), and rescue of *twit* expression in *wit* mutants is able to partially (but not fully) rescue this phenotype (Gao et al., 2005). Suppression of neuromuscular junction growth and function that phenocopies loss of *wit* or *Mad* gene function can also be achieved by Gal4-UAS induced expression of Mad^{1}, a dominant negative allele of *Mad* that contains a point mutation in the MH1 DNA binding domain but otherwise behaves like wild-type *Mad*, demonstrating that Smad DNA binding and Smad-dependent transcription is the primary mediator of BMP function in fly neurons (Berke et al., 2013).

1.4 Specification and Differentiation of FMRFa-Expressing Tv4 Neurons

The specification and differentiation of *Drosophila* Tv cluster neurons from their neuroblast represents one of the most thoroughly understood models of neuronal subtype development with one of the best described transcription factor regulatory networks. *The reader is referred to* ***Figure 4*** *throughout the following section.* Here, I will introduce the transcriptional networks that specify the Tv cluster neuronal subtypes with a primary focus on the Tv4 neuron, which expresses the neuropeptide *FMRFa*.

1.4.1 Specification of the Tv Cluster

The Tv neurons are a discrete cluster of four neurons in each thoracic VNC hemisegment that is derived from a single neuroblast lineage, but each neuron ultimately differs in gene expression profile, neurotransmitter expression and morphology. The Tv cluster neurons are generated by the neuroblast 5-6T lineage (NB 5-6T) (Baumgardt et al., 2007). This neuroblast undergoes the archetypal VNC Hb → Kr → Pdm → Cas → Cas/Grh transitions in temporal selectors (**Figure 4A,B**). Through the Hunchback to Cas transitions, the NB undergoes archetypal Type I divisions (**Figure 4B**). However, during the Cas/Grh window, the lineage switches to Type 0 divisions. The four Tv neurons, Tv1 through Tv4, numbered in the order of their birth, are born during this Cas/Grh window and represent the final neurons of the lineage (Baumgardt et al., 2009; Baumgardt et al., 2007).

In the NB 5-6T lineage, Notch is required for the proper transition from Type I NB division to the Type 0 NB divisions that occur during Tv-neuron generation (Baumgardt et al., 2009; Ulvklo et al., 2012). Notch intracellular activity increases in the NB5-6T by the time of the Type I to Type 0 switch; appearing at Stg. 13 and increasing in intensity through the period of Tv-neuron generation until the programmed cell death of the NB5-6T neuroblast. Notch activation

through this period is required for normal Tv neuron numbers. Loss of Notch signalling results in supernumerary Ap neurons. Intriguingly, this is not due to the premature appearance of Ap neurons through the lineage, nor shifts in NB5-6T temporal progression, nor extra NB5-6T divisions. Instead, it is due to the continuation of Type I divisions through the Tv-neuron generation phase, and thereby the generation of two neurons per NB division, instead of just one if Type 0 NB divisions would normally have occurred (Ulvklo et al., 2012).

Tv1 and Tv4 can be easily discriminated by differential expression of the neuropeptides Nplp1 and FMRFa, respectively, by late embryonic stages (Baumgardt et al., 2009; Baumgardt et al., 2007). In *cas* mutants, all Tv neurons are lost. In *grh* mutants, only the Tv4-neuron is not born, the last neuron to be generated by the lineage. The helix-loop-helix (HLH) factor Collier (Col) is expressed in the NB and in newly born Tv neurons and is a critical determinant of Tv neuron specification and differentiation; in *col* mutants the Tv cluster is absent (Baumgardt et al., 2007).

1.4.2 Early Differentiation of Tv Neurons

As the Tv1-4 neurons are generated between embryonic stages 14-15, the expression of numerous transcriptional regulators becomes induced, including the LIM-Homeodomain transcription factor *apterous* (*ap*), and the transcriptional cofactor *eyes absent* (*eya*) (Miguel-Aliaga et al., 2004). Subsequently, Tv1 and Tv4 neurons then initiate the expression of the basic helix-loop-helix (bHLH) transcription factor *dimmed* (*dimm*) (Allan et al., 2005) under the control of *ap*, *eya* in both cells, and also by *col* in Tv1 (Baumgardt et al., 2009; Baumgardt et al., 2007) (**Figure 4C&E**). Dimmed is a pro-neurosecretory determinant (Hewes et al., 2003; Mills and Taghert, 2012). Dimmed activates or up-regulates multiple genes that process pre-pro-peptides and increases the expression of genes required for secretory machinery biogenesis such

as dense core vesicles (Hadzic et al., 2015; Park et al., 2011). Tv2, Tv3, and Tv4 also express the transcriptional co-factor *dachshund* (*dac*) and the zinc finger transcription factors *squeeze* (*sqz*) and *nab*. Upon their induced expression, *sqz* and *nab* subsequently repress *col* expression, thereby extinguishing *col* expression in Tv2-4 soon after neuronal birth, and leaving only Tv1 to maintain col expression into adulthood (Baumgardt et al., 2009; Baumgardt et al., 2007; Eade et al., 2012). Loss of *sqz* or *nab* leads to mis-specification of the Tv4 neuron and duplication of Tv1-like Nplp1-expressing neurons. In this context *sqz* and *nab* act as a further level of temporal regulatory factors further dividing up the *cas/grh* temporal window (Allan et al., 2005; Baumgardt et al., 2009) (**Figure 1D**).

1.4.3 Terminal Differentiation of Tv4 Neurons

The collective action of *ap, eya, dac,* and *dimm* are all required for the appropriate expression of FMRFa during development (details provided in **Figure 4C**). In addition to the role of these transcriptional regulators in FMRFa initiation, a target-derived signal was also found to be required. During embryonic stages 15-17, Tv4 neurons each extend their axon towards the midline, where they turn dorsally to innervate the dorsal neurohaemal organ (DNO), a glial structure that protrudes from the dorsal midline of the VNC (Allan et al., 2003). Once Tv4 axons reach the dorsal neurohaemal organ, during mid-stage 17, Tv4-neurons are able to gain access to the BMP ligand, Gbb. Gbb activates BMP signalling through the *wit, tkv* and *sax* receptors (Allan et al., 2003). Upon this target contact and BMP pathway activation at mid-late stage 17, phospho-Mad (pMad) is observed in the Tv4 neuron's nucleus (See **Figure 4C&E**). ***Active BMP signalling is absolutely required for the initiation of the FMRFa neuropeptide in the Tv4 neuron.*** This target derived BMP-signalling is dependent on dynein-based retrograde trafficking within the axon (Allan et al., 2003; Eade and Allan, 2009); FMRFa expression and pMad

immunoreactivity are lost when a dominant negative form of *Glued*, a member of the p150 dynein/dynactin complex, is expressed in the Tv4 neuron (Allan et al., 2003; Eade and Allan, 2009). The target-derived signal that is required for FMRFa expression appears to be solely mediated by the BMP signal. If Tv4 axons are misguided away from the midline, by overexpression of UAS-*roundabout* (UAS-*robo*), pMad and FMRFa expression are lost. The resulting loss of FMRFa expression can be rescued by co-expression of UAS-*gbb* with *UAS-robo* (Allan et al., 2003). Similarly, FMRFa loss of expression in *wit* mutants is rescued by expression of constitutively activated *tkv* and *wit* receptors (Allan et al., 2003). Therefore, activation of the BMP pathway appears to be the mediator of dorsal neurohaemal organ-dependent activation of FMRFa in the TV4 neurons.

Taken together, the transcription factors *cas, col, grh, sqz,* and *nab* can be viewed as specification transcription factors, because in their absence the neuron is not specified normally, or is transfated to a different neuronal subtype. In contrast, *ap, eya, dac, dimm* and BMP signalling appear to regulate the differentiation of Tv neurons, in that in their absence, the neuron is still present but fails to express its normal repertoire of terminal differentiation genes. However, depending on the exact criteria by which neurons are considered "differentiated", these factors can also be viewed as part of a continuous process of progressively finer demarcation of cell function and identity. *In Chapter 4, I will re-visit the idea of cell subtype identity in light of findings in Chapter 2, which call into question the demarcation of 'specification transcription factors' and 'terminal differentiation' genes.*

At the time of initiating this thesis, it was unclear as to how the Tv4-neuron interprets both the intrinsic transcription factors and the extrinsic BMP signalling to give appropriately timed FMRFa expression. Unresolved questions included: Do all these factors affect FMRFa transcription, and if so, do they do so directly or indirectly? How are these inputs interpreted by the *cis*-regulatory regions of the *FMRFa* gene to generate appropriate spatial (Tv4-specific) and temporal (late stage 17-specific) initiation?

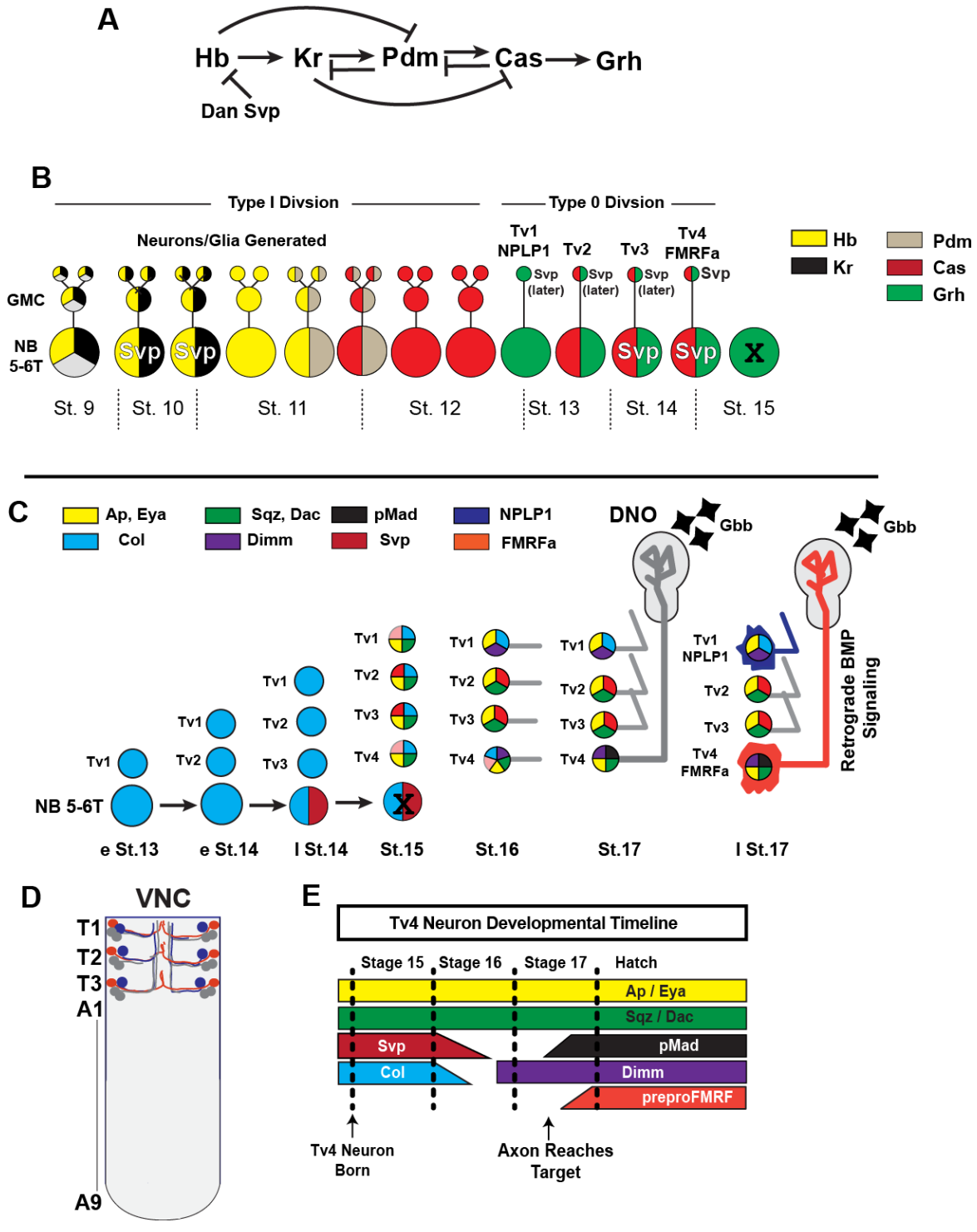


Figure 4. Development of the Tv4 neurons.

(**A**) Summary of the temporal cascade of temporal selectors and their regulatory relationships in the NB5-6T lineage, also showing the role of Svp as a switching factor (see Section 1.5 for more detailed discussion). (**B**) Expression of the temporal selector TFs in NB5-6T throughout lineage progression. 'X' indicates death of neuroblast. Svp protein expression is not detected in the neuroblast or daughter cells when Tv1,2 and 3 are initially born, but is detectable by the end of St.14/early St.15. (**C**) Expression profile of TFs involved during the specification and differentiation of Tv-cluster neurons. The Tv4 neuron begins to extend its axon towards the dorsal neurohaemal organ (DNO) at mid-St.17, where it gains access to the BMP ligand Gbb. Strong pMad immunoreactivity and FMRFa expression is detected shortly thereafter. At St.15 Svp is detectable in Tv1 and Tv4 (light red), albeit at lower levels than Tv2 and Tv3 (dark red). In Tv4, Svp down regulates during St. 16 in the Tv4 neuron and is lost before pMad immunoreactivity is observed at mid to late St. 17. (**D**) Schematic of the *Drosophila* VNC showing location of the Tv clusters in the thoracic segments. (**E**) Timeline of gene expression specifically within the Tv4 neuron.

1.5 COUP-TF Family Factors and the Timing of Neuronal Identity Specification

COUP-TFI/II (chicken ovalbumin upstream promoter-transcription factor I/II) family factors are expressed in a highly dynamic manner in both vertebrate and invertebrate nervous systems during development (Qiu et al., 1994; Qiu et al., 1997). COUP-TFs play key roles in neurogenesis and gliogenesis (Naka et al., 2008; Yamaguchi et al., 2004), neuronal migration (Tripodi et al., 2004) as well as axon guidance (Qiu et al., 1997) in vertebrates. The *Drosophila* Coup TFI/II ortholog, *svp* is required to act as a temporal switching factor in many lineages, including in the NB5-6T lineage, to transition from a Hb to a Kr temporal selector window (Benito-Sipos et al., 2011; Kanai et al., 2005; Kohwi and Doe, 2013). Studies in mice have shown that this role is conserved; for example, the transition from earlier neurogenic to later gliogenic competency windows of neural stem cells is inhibited in *coup-tfI/-II* loss of function models (Naka et al., 2008). In late stages of NB5-6T lineage progression, *svp* further functions as a sub-temporal switching factor to determine different neuronal identities within the window of expression of a single temporal selector, *cas* (Naka et al., 2008) (details provided in Section 1.5.1).

1.5.1 Role of Seven Up in Specifying Tv4 Neurons

See **Figure 4** for developmental expression pattern of Svp expression in Tv4 neurons.

Seven-up (*svp*) plays two roles in specifying the Tv4 neurons at different stages of development. Early in the NB5-6T lineage, *svp* (~stg 10) is expressed as a transient pulse in the neuroblast and is required to promote the switch from Hb→Kr by suppressing *hb* (Benito-Sipos et al., 2011). In *svp* loss of function mutants, *hb* is not down regulated and later born neurons of the NB5-6T lineage, including the Tv cluster, are lost in most thoracic segments. Later in development, (St 14), Svp is specifically expressed in the neuroblast and in the postmitotic Tv1-4

neurons, but is rapidly down-regulated in Tv1 and Tv4 (Benito-Sipos et al., 2011). Svp continues to be expressed in Tv2 and Tv3 into late embryonic stages (Benito-Sipos et al., 2011). Loss of *svp* leads to supernumerary Nplp1^{+} Tv1-like cells and a total loss of Tv4 FMRFa-expressing cells. Examination of Tv1-4-specific markers indicated that first part of the Cas window is extended, while the last part in which Grh is upregulated is lost. Moreover, *svp* is required to down regulate *col* in Tv2-4, in order that the correct Tv cell fates are discriminated. Ectopic (extended later than normal during development) expression of *svp* leads to *col* down-regulation and loss of both Nplp1 and FMRFa expression, in addition to loss of *dimm* expression in Tv1 and Tv4. Taken together, *svp* acts like a 'sub-temporal factor', similar to *sqz* and *nab* (Baumgardt et al., 2009), in that it plays a critical role in generating unique Tv1-Tv4 neuronal identities, within a single temporal selector expression window (Benito-Sipos et al., 2011).

1.6 *Cis*-Regulatory Modules and Enhancers

Transcription of genes by RNA polymerase II yields messenger RNA intermediates that are translated into protein. The regulation of transcriptional activity is essential to proper gene expression levels in time and space, a fundamental process required to make genetically identical cells different during development in multicellular organisms. Gene transcription is regulated by both promoter sequence composition, where the core transcriptional machinery acts (i.e. RNA pol II, TFIIB, the Mediator complex), and additional sequences that are more distal from the promoter. These more distal *cis*-regulatory regions (CRMs) contain information on the same strand of DNA (in *cis*) as the gene is encoded. For the purposes of this introduction and thesis work I will focus on gene regulation by CRMs; however, it would be overly simplistic to not recognize that cell-type specific promoters exist and play a role in gene regulation and that,

preferential enhancer-promoter pairing between enhancers and promoters is a well-known phenomenon (Choi and Engel, 1988; Li and Noll, 1994).

Transcriptional enhancers, commonly simply referred to as enhancers, contain specific DNA sequence motifs with which DNA-binding proteins interact, in order to ultimately influence transcription of their target gene. Enhancers were first defined as *cis*-acting DNA sequences that increase transcription in an orientation-independent manner and independent of their distance relative to the transcriptional start site. This flexibility is highlighted by the fact that enhancers can be found tens of kilobases 5' and 3' from their target genes' transcriptional units, or also within intronic sequences (even within the introns of other neighbouring genes) (Evans et al., 2012; Lodish, 2004). Enhancers can vary greatly in size, typically from a few hundred bases to kilobases (Bulger and Groudine, 2011; Calo and Wysocka, 2013; Evans et al., 2012). From a technical standpoint, however, these 'functional' definitions may be dependent on technical restrictions due to the sensitivity of reporter systems used to characterize the enhancer's activity, and the degree to which a 'minimal enhancer' (the smallest enhancer fragment sufficient to generate a specific pattern in time and space), was dissected out of a general regulatory region. Since highly trimmed sequences are known to drive ectopic, very weak, or restricted expression patterns compared to a larger enhancer region it is often unclear if important suppressor or activator elements have been removed or if division of a regulatory region has now split up a 'minimal enhancer' (Evans et al., 2012). These aspects of enhancer 'dissection' are discussed in light of my own findings in Chapter 4.

A single gene may be regulated by multiple enhancers that act in a modular manner, whose combined activity generates the overall expression pattern of the gene. A classic model of this enhancer modularity has been the *even-skipped* gene in *Drosophila*. The expression pattern of

even-skipped in the early embryo (7 regularly spaced dorsoventrally extending stripes) is controlled by independent enhancers that appear to act in a modular manner; each enhancer region controls the expression domain of just one or two of the seven stripes (Goto et al., 1989; Small et al., 1992).

1.6.1 Integration of *Trans*-Acting Factors at an Enhancer

A primary focus of decades of research has been the binding and function of sequence-specific transcription factors at enhancers. In fact, a tacit definition of an enhancer for some time has been a cluster of transcription factor binding sites located somewhere 'near' the gene of interest that regulates its promoter activity. From early analysis, transcription of a gene requires that enhancers integrate both positive and negative regulatory inputs, and that many different combinations of activators and repressors of transcription can be bound at an enhancer. Individual transcription factors are often restricted to specific cell types, present during specific stages of development, and/or only present in response to an environmental cue. By virtue of the selective sequence-specific binding of specific transcription factors, an enhancer, therefore, integrates developmental and environmental information to regulate the expression of a gene in a biologically-appropriate and responsive manner.

When sequence-specific transcription factors are bound to enhancers they recruit non-DNA-binding cofactors. These non-DNA binding cofactors regulate recruitment of basal transcriptional machinery and RNA polymerase II either directly or indirectly via the Mediator complex (reviewed in (Malik and Roeder, 2005) and (Szutorisz et al., 2005)). Some of the non-DNA-binding co-factors include chromatin modifying factors that include histone acetyltransferases and members of the SWI/SNF family that regulate the removal of nucleosomes and change the local chromatin structure around the enhancer (reviewed in (Bulger

and Groudine, 2011) and (Calo and Wysocka, 2013)). Recent evidence from genome wide RNA-seq studies suggests that enhancers themselves can also act as transcriptional start sites for non-coding RNAs that may play a wide variety of regulatory roles. The transcripts, however, are not stabilized to the degree that is observed for protein coding transcripts by virtue of a lack of such processes as polyadenylation, and therefore appear to be much more transient in the cell (Core et al., 2014). This thesis does not directly test any hypothesis that involves chromatin modifiers or non-coding RNA, so further detailed discussion is beyond the scope of this work.

1.6.1.1 Models of CRM Function

Two predominant models of enhancer structure and organization have emerged since detailed CRM analysis was undertaken, beginning in the 1980s. The 'enhanceosome' focuses on the 3-dimensional organization of the DNA with the transcription factors that bind it, while the 'information display' (billboard) model is primarily concerned with the number of positive and negative transcription factor binding sites in an enhancer, with less concern about their arrangement along the DNA strand (Arnosti and Kulkarni, 2005; Kulkarni and Arnosti, 2003). See **Figure 5** for a summary of the enhanceosome and information display models of CRM function.

The enhanceosome model is supported by findings that changes in spacing between transcription factor binding sites can dramatically decrease enhancer activity because synergistic activity or cooperativity in transcription factor binding is disrupted. One highly informative example of the enhanceosome model is the activation of the IFN-β gene in response to viral infection (Ford and Thanos, 2010; Lotz et al., 2004). Assembly of several factors at the enhancer is required to activate the gene. These factors physically interact and/or sit at closely apposed sites on the DNA. The presence of the composite surface they present is required to recruit a

cofactor that drives strong transcription of the gene (Merika et al., 1998). The biochemical cooperativity mechanistically explains the binary on/off response of this enhancer, since a complete complex must be present for transcription to occur.

The information display model simply requires the correct balance of positive and negative inputs, without strict requirements of spacing and organization of transcription binding sites along the length of the enhancer (Arnosti and Kulkarni, 2005) (**Figure 5B**). The very well established stripe 2 module of the *even-skipped* enhancer is an example of the information display model. The *even-skipped* strip 2 enhancer is active only in a narrow band of cells where Bicoid and Hunchback (activating factors in this context) are present, but where expression levels of repressive factors Krüppel, Giant and Sloppy-paired are low or absent (Andrioli et al., 2002; Small et al., 1992). The transcription factor binding sites can be re-arranged without significant loss of enhancer function (Arnosti et al., 1996). This flexibility is also suggested by poor conservation in the arrangement of positively-activating binding sites for Bicoid and Hunchback across multiple sequenced *Drosophila* species (Ludwig et al., 1998). Moreover, activator sites can be 'swapped'; replacement of positively acting Hunchback binding sites with Gal4 binding sites in the presence of maternally provided Gal4-Sp1 protein rescues enhancer function (Arnosti et al., 1996).

In both of the above examples, enhancers act as simple computers. A collection of inputs are resolved by biological processing into single output which is an instruction to the basal transcriptional machinery to clear the promoter and transcribe the gene. The two models discussed above most likely represent extremes on a spectrum with many enhancers potentially demonstrating characteristics of each. To date, only a few enhancers have been analyzed at a

resolution through deletions, point mutations, and rearrangements, to allow for conclusive categorization of known enhancers into either aforementioned category, or unique categories.

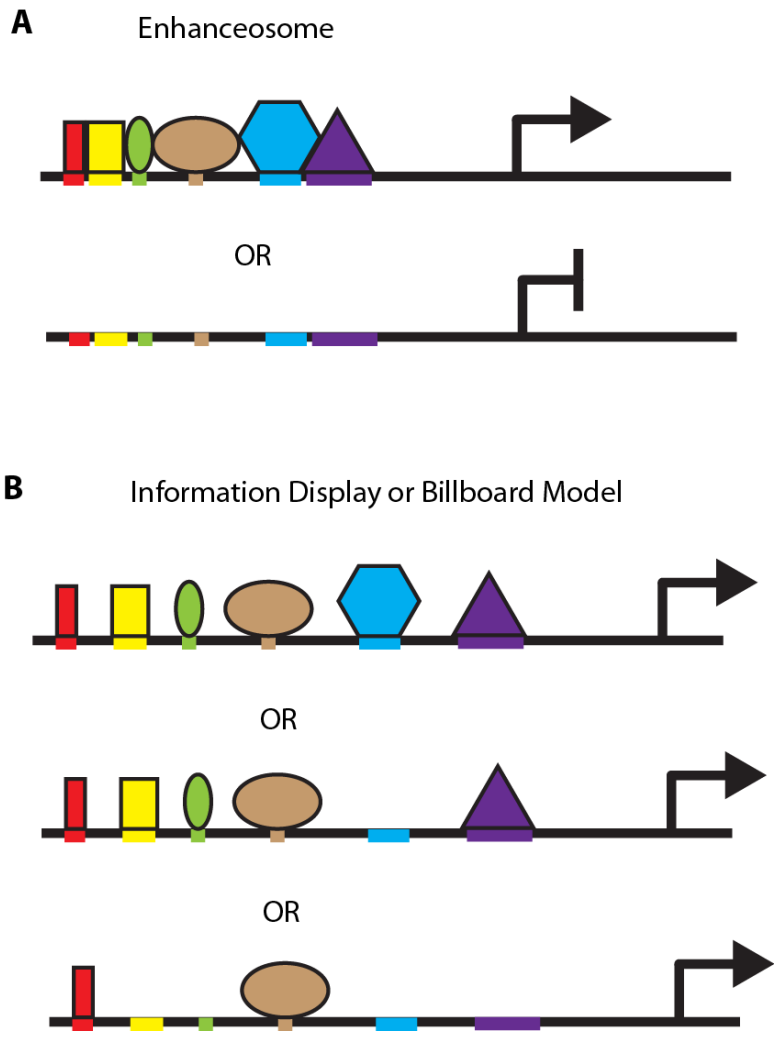


Figure 5. Summary of the two major models of CRM function

(**A**) Enhanceosome model of CRM activity, showing highly cooperative DNA binding, stringent transcription factor binding site motif organization and a unitary output where the integrated activity of all transcription factors is required simultaneously. (**B**) Information Display or Billboard Model of Enhancer activity. Additive or cooperative binding is possible. Less stringent requirements on transcription factor binding site order and distance can utilize a subset of factors to activate transcription.

1.6.2 *Cis*-Regulatory Regions of the FMRFa Gene

Work previously performed in the Taghert lab had identified numerous non-overlapping enhancers that independently activate transcription in different FMRFa-expressing cells (Benveniste and Taghert, 1999; Schneider et al., 1993). A single 446-bp region situated 467 bp upstream from the *FMRFa* transcriptional start site was identified as containing sufficient information to trans-activate expression specifically only in the six Tv4 neurons. No other regions were found that were necessary or sufficient for Tv4-neuron expression. Partial deletional analysis showed that the Tv4 FMRFa enhancer contained three key activation regions and two repressor regions (**Figure 6**) (Benveniste and Taghert, 1999). Activation regions were determined to exist if they were necessary for Tv4 specific reporter expression. Repressor regions were defined as sequences that lead to ectopic reporter expression when deleted (Benveniste and Taghert, 1999). It should be noted that the reporter constructs used in these studies were random p-element genomic integrations, rather than being site-directed to a specific genomic locus. Thus, the influence of positional effects cannot be fully discounted in the interpretation of these data, in spite of an average of three or more unique insertions being analyzed for each transgene. Within the activation domains, three putative LIM-homeodomain transcription factor sites were identified that would match an Ap-binding motif. The results of mutant analysis of these putative Ap-sites suggested that all three Ap-sites contribute equally, in an additive manner, to enhancer function. Moreover, Ap was shown in electrophoretic mobility shift assays to bind all these putative Ap sites *in vitro*. **Figure 6** summarizes previous work on the FMRFa Tv4-neuron specific enhancer.

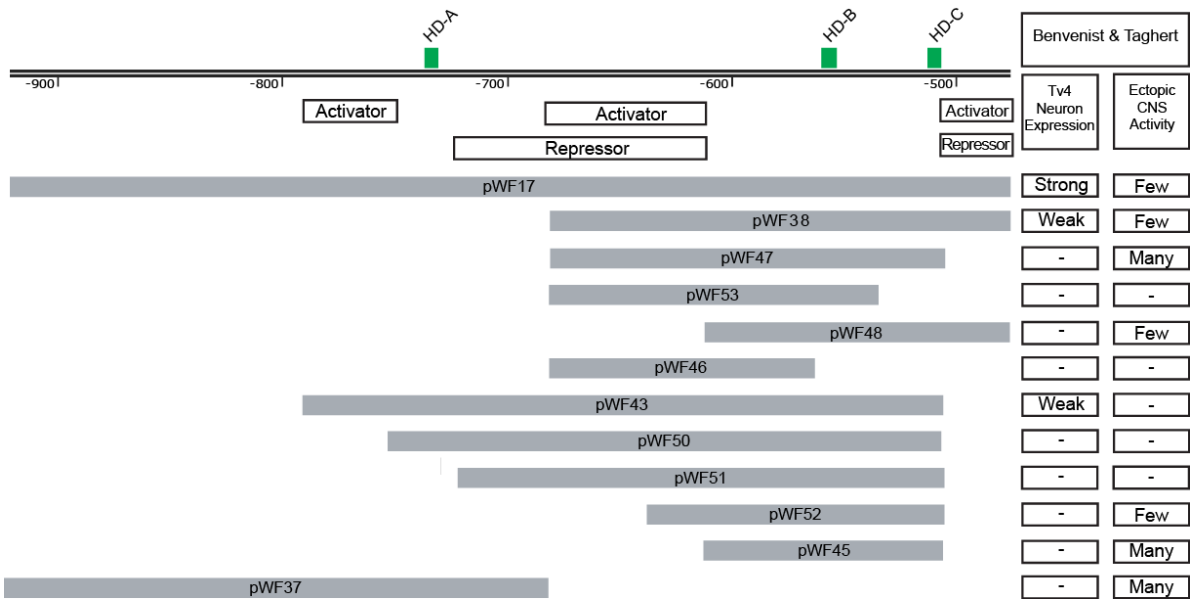


Figure 6. Summary of functional regions within the Tv4-enhancer of *FMRFa* from the Taghert lab.

The full span of the Tv4-enhnacer is shown in the top line, which also denoted the putative homeodomain-binding sites (HD-A/B/C) (green boxes). Below this line are summaries of the regions that were postulated to contain activator or repressor sequences. Note that only HD-C appears to span an activator region. Below this are a series of restriction genomic fragments named pWF #. To the right of each of these, I show the expression level generated by each genomic fragment when placed in a lacZ-reporter transgene '-' refers to an absence of expression in Tv4 neurons.

1.6.3 Smad Binding Sites in the *Drosophila* Genome

The transcriptional activity of the Receptor-Smad Mad (vertebrate Smad1/5/8) and the Co-Smad Medea (vertebrate Smad4) are extremely well conserved. Genomic BMP-response elements (BMP-REs) can activate or repress transcription upon Smad binding. This difference in *cis*-regulatory activity to the same *trans*-acting Smads is controlled by both the sequence of the Smad binding sites themselves, the presence of co-activators and repressors in the cell, as well as the flanking binding sites of the BMP-response element to which additional transcription factors bind (Ross and Hill, 2008b).

1.6.4 The BMP Silencer Element and the Activation Element

Currently, two *BMP-RE* sequence types have been defined to be common in the regulatory regions of numerous BMP-regulated genes in *Drosophila*. These are the Silencer Element (SE) and the Activation Element (AE).

*BMP-Silencer Element (BMP-SE) GRCGNCNNNNNGTC**T** (R= G or A)*
*BMP-Activation Element (BMP-AE) GGCGCCANNNNGNC**V** (V = A,C, or G)*

The sequence requirement and function of the BMP-SE and BMP-AE have been best described in the *Drosophila* wing imaginal disc and early embryonic dorso-ventral gradient, which are long standing models of morphogen gradients, especially with respect to BMP-signalling (Gao et al., 2005; Pyrowolakis et al., 2004; Weiss et al., 2010).

1.6.4.1 BMP-SE And BMP-AE Motif Use in the Wing Imaginal Disc

Please See **Figure 7** for a summary of BMP AE and BMP SE elements in the *Drosophila* imaginal disc.

During wing imaginal disc development, the BMP ligands Dpp and Gbb are expressed in a gradient of BMP activity across the disc with the highest ligand expression in the middle of the disc. BMP activity represses *brk* (Affolter and Basler, 2007; Muller et al., 2003). This generates an inverse gradient of BMP activity and Brk expression across the tissue (i.e. when BMP signalling is high, Brk activity is low). In regions of high BMP signalling, Smads and the transcription factor Schnurri (Shn) bind to BMP-SE sequences in the *brk* locus to repress *brk* expression. In the BMP-SE element, pMad binds to GRCGNC and Medea to GTCT, with Shn acting as a co-repressor that requires the terminal thymidine (GTC**T**) (Marty et al., 2000; Pyrowolakis et al., 2004). In regions of low BMP signalling, Brk is not repressed and acts as a repressor to a large number of BMP-regulated genes by binding to GGCGYY (where Y=C or T) containing sequences in these gene's CRMs (Sivasankaran et al., 2000 ; Zhang et al., 2001). Compare the "BMP Off" scenario in **Figure 7** with the "BMP On" scenario.

Genetic analysis has shown that *brk* mutants largely phenocopy *brk;dpp* and *brk;Mad* double mutants with respect to BMP-dependent gene expression (Marty et al., 2000; Rogulja and Irvine, 2005; Schwank et al., 2012). This means that BMP-directed silencing of the *Brk* repressor appears to be the primary mechanism that activates BMP-dependent genes in the wing disc**.** Smads do not activate most BMP effector genes in the wing disc, such as the well-characterized gene *bifid* (Campbell and Tomlinson, 1999; Jazwinska et al., 1999; Winter and Campbell, 2004). However, genes like *spalt major* do show elevated expression when BMP-signalling is active, although the primary mechanism of its expression is *brinker* de-repression (Marty et al., 2000).

The suppressive Smad, *dad*, is activated by BMP signalling as part of a feed-back loop. The BMP-AE was identified in an intronic region of the *dad* gene (Weiss et al., 2010). The *dad* BMP-AE responds to both Smad and Brk binding that competitively mediate activation and

suppression, respectively (**Figure 7** lower part of panels). The BMP-AE is bound by pMad at GGCGCC and by Medea at GNCV. Unlike the BMP-SE, Shn is not recruited to the BMP-AE since no terminal thymidine is present (Weiss et al., 2010). This enables the element to work as an activator rather than a suppressor, even in the presence of Shn. Brk, however, can repress the BMP-AE by directly binding to the GGCGCC sequence (Marques, 2005; Saller and Bienz, 2001; Weiss et al., 2010). Dad expression therefore follows the relative Brk and pMad activity across the wing imaginal disc, where Dad is expressed most highly in the high pMad activity regions (Marques, 2005; Rushlow et al., 2001; Weiss et al., 2010).

1.6.5 Smad Interaction with Other Binding Partners at the BMP-RE

Almost universally, Smad proteins interact with other DNA-binding cofactors to bind DNA with high affinity and selectivity (reviewed extensively (Marques, 2005)and (Massague et al., 2005)). For example, two Ultrabithorax (Ubx) molecules bind DNA at two flanking sites around Mad and Medea site to repress *spalt* (*sal*) in the fly haltere (Walsh and Carroll, 2007). In mice, a Smad1+Smad4 complex physically interacts with β-catenin at adjacent binding sites to cooperatively regulate *c-Myc* transcription (Hu and Rosenblum, 2005), while FoxO-type transcription factors are recruited to forkhead-binding element sequences closely adjacent to Smad binding sites in several vertebrate genes where they interact with Smad2/3+Smad4 complexes (Gomis et al., 2006; Seoane et al., 2004). Careful inspection of the Smad binding sites in these cases suggest that a higher degree of degeneracy is tolerated at the Receptor-Smad binding site than is suggested by the *Drosophila* studies on the BMP-AE and SE. However, this may be complicated by the fact that there are three orthologues of *Mad* (Smad1/5/8) and two of *Smox* (Smad2/3) that respond to BMP and TGF-β, respectively. Interestingly, similar binding site resolution studies between Mad+Medea complexes with potentially physically-interacting

transcription factors, or even adjacent transcription factor motifs, have not been performed to nearly the same extent in *Drosophila* as similar studies in vertebrates.

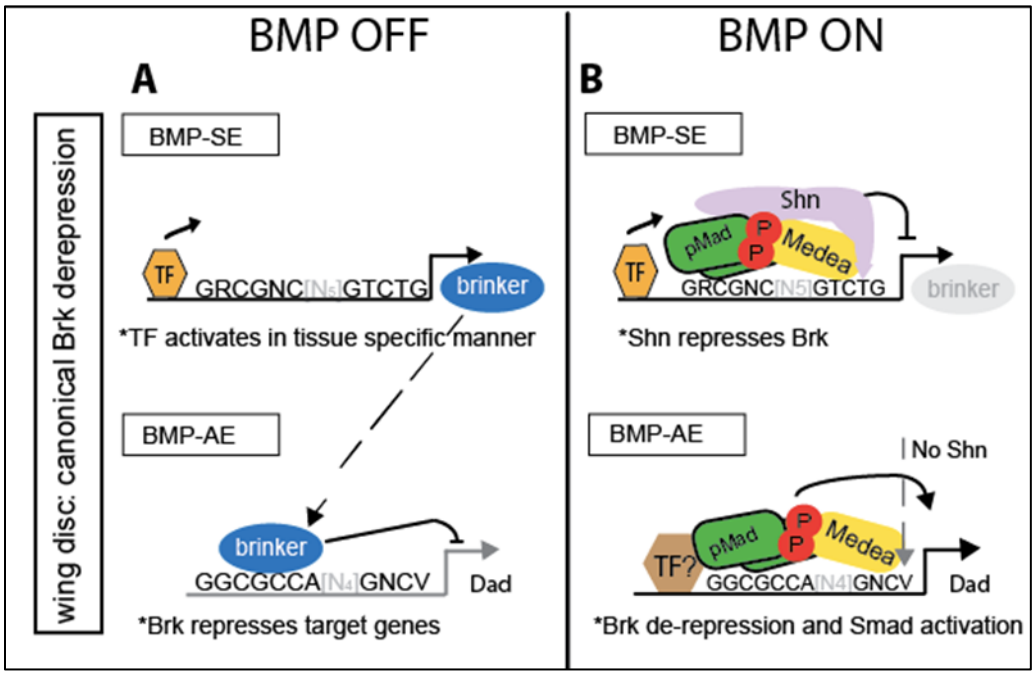


Figure 7. Summary of BMP Silencing Element (SE) and BMP Activating Element (AE) activity in the wing disc.

(**A**) Conditions at the BMP-SE and AE when no BMP ligand is present. Brk is transcribed because of other transcription factors present in the cell. Brk actively suppresses genes at Smad binding sites. (**B**) When cells encounter BMP ligand, phosphorylated Mad with Medea complex with Shn to repress Brk at SEs. De-repression of Brk allows many BMP-regulated genes to be expressed. In some cases Mad and Medea act positively at AEs to turn on gene expression in response to active BMP signalling.

1.7 Thesis Objectives

Following initial neurogenesis, newly born neurons extend axons to the targets they innervate. Target derived signals provide extrinsic developmental cues that interact with intrinsic transcription factor codes to activate terminal differentiation genes such as neurotransmitters. Using the well-defined *Drosophila* model of the Tv4-neuron and the Tv4-neuron specific FMRFa enhancer, my thesis work:

1) Describes how a neuron integrates extrinsic and intrinsic developmental cues at an enhancer.

2) Demonstrates how Seven up, a COUP TF-I/II family transcription factor, acts as a postmitotic timer of neuronal terminal differentiation.

3) Characterizes an atypical BMP-response element with sufficient sequence complexity for genome-wide *in silico* screens.

Finally, in Appendix A, I present data from a side-project wherein I generated mutants for the only two LIM-homeodomain transcription factors that had not been mutated or examined in *Drosophila* previously: the LMX-type genes *CG4328* and *CG32105*. I demonstrate that *CG4328* plays a key role in regulating octopaminergic innervation of the oviduct and that loss of *CG4328* function results in loss of egg-laying behaviour.

2 *CIS*-REGULATORY INTEGRATION OF INTRINSIC AND TARGET-DEPENDENT REGULATORS IS REQUIRED FOR TERMINAL DIFFERENTIATION OF TV4 NEURONS

2.1 Synopsis

Terminal differentiation of many developing neurons occurs once they innervate their target cells, where they receive signals secreted by the target cells. These signals are transduced by presynaptic receptors and a transduced signal is transported to the nucleus by retrograde trafficking. Target-derived signalling by multiple different signalling pathways is required to trigger the differentiation of multiple different cell types in vertebrates and invertebrates. However, how these extrinsic target derived signals integrate with an existing intrinsic transcription factor code to activate target dependent genes in neurons has not yet been characterized at a molecular mechanistic level. We have utilized general models and understanding of how transcriptional activators and repressors integrate at *cis*-regulatory regions as a means to gain insight in to how these mechanisms might be employed in our neurons of interest with respect to target derived gene expression.

Previously, Dr. Allan showed that retrograde BMP signalling induces expression of terminal differentiation in *Drosophila* Tv4 neurons, by inducing the neuropeptide FMRF. Further work by him, colleagues in the Stefan Thor lab, and others elucidated a number of genes required for FMRFa initiation and maintenance, most of which were transcription factors or transcriptional cofactors, as well as a Tv4-neuron specific enhancer found near the *FMRFa* gene. In this chapter, using detailed mutagenesis of the Tv4 *FMRFa* enhancer, traditional transcription factor mutants, and cell-type specific gain of function assays in conjunction with biochemical assays, I test the hypothesis that the Tv4 enhancer integrates the extrinsic BMP signalling with the intrinsic Tv4

neuron transcription factor code. I first define the subset of the known Tv4 transcription factor code that acts at the Tv4-enhancer. By point substitution mutation and deletion analyses in the context of reporter transgenes, I isolate two regions necessary to the function of the Tv4 enhancer. I also found these regions are sufficient to drive appropriate reporter expression in time and space, yet each responds to a different subset of the Tv4 transcription factor code. One of these minimal elements represents a Homeodomain transcription factor response element while the other responds to BMP signalling. I go on to show that Seven up plays an unexpected role in regulating the timing of Tv4 enhancer activation in the postmitotic Tv4-neuron.

2.2 Introduction

Nervous system development requires the differentiation of diverse neuronal subtypes under the direction of combinatorially-acting transcription factors (di Sanguinetto et al., 2008; Skeath and Thor, 2003). Additionally, signals derived from the neuron's axonal targets, such as retrograde BMP signaling, are responsible for activating many subtype-specific neuronal genes (da Silva and Wang, 2011; Hippenmeyer et al., 2004; Nishi, 2003). Target-dependent genes can be associated with neurotransmitter or neuropeptide phenotype (Allan et al., 2003; Coulombe and Kos, 1997; Pavelock et al., 2007; Stanke et al., 2006; Veverytsa and Allan, 2011; Xu and Hall, 2007), or activate transcription factors that promote axonal branching or topographic mapping (Haase et al., 2002; Hodge et al., 2007; Patel et al., 2000; Patel et al., 2003). The gene regulatory function of target-derived signaling is unresolved, but is postulated to temporally couple terminal differentiation to target contact and to specify gene expression. However, these genetic and cellular views are not informed by a molecular understanding of target gene regulation.

In *Drosophila*, target-derived BMP signaling is required for subtype-specific neuropeptide gene expression (Allan et al., 2003; Veverytsa and Allan, 2011) and positively regulates motoneuron synaptic morphology, transmission, and plasticity (Aberle et al., 2002; Goold and Davis, 2007; Marques et al., 2002; McCabe et al., 2003). These require that Glass Bottom Boat (Gbb) ligand acts at presynaptic BMP receptors Wishful thinking (BMPR-II), Thickveins and Saxophone (BMPR-I) (Aberle et al., 2002; Allan et al., 2003; Marques et al., 2002; McCabe et al., 2003). BMPR-I phosphorylates Mad which translocates to the nucleus with Medea to bind DNA as the sequence-specific Smad complex (Raftery and Sutherland, 1999; Shi and Massague, 2003). Smad transcriptional activity typically requires cooperation with other sequence-specific transcription factors (Massague et al., 2005; Ross and Hill, 2008a), which facilitates cell subtype-specific gene regulation (Chen et al., 2008; Mullen et al., 2011; Trompouki et al., 2011) (See **Section 1.6.4** and **Figure 7**).

Target-derived BMP-signaling initiates *FMRFa* in Tv4-neurons in late Stage (Stg) 17 embryos and maintains its expression thereafter (Allan et al., 2003; Eade and Allan, 2009). Genetic analysis showed that Tv4-specific *FMRFa* expression requires target-derived BMP signaling and a combinatorial code of transcriptional regulators, including the LIM-homeodomain factor, Apterous (Allan et al., 2005; Allan et al., 2003; Baumgardt et al., 2009; Karlsson et al., 2010; Miguel-Aliaga et al., 2004). All of these inputs are found in other neurons, but their unique overlap in Tv4-neurons provides a logical molecular explanation for the extreme selectivity of FMRFa expression. In this model, we addressed how Smads and local transcriptional regulators integrate to activate the *FMRFa* gene, and whether target-derived signaling is the primary timer for *FMRFa* initiation. We identified necessary and sufficient *cis*-elements within a previously identified Tv4-specific *FMRFa* enhancer (Benveniste and Taghert,

1999; Benveniste et al., 1998), and defined their transcriptional inputs. We also show that downregulation of the nuclear receptor, Seven up, de-represses FMRFa expression at the time of target contact; thus, we find that target-dependent gene repression prior to target contact is a constituent of the target-dependent gene activation mechanism. These findings reveal surprising complexity at the heart of BMP-dependent *FMRFa* regulation, and provide a novel model for understanding target-dependent differentiation.

2.3 Materials and Methods

2.3.1 Fly Genetics and Stocks

Strains used: *Df(3R)DI-KX23* (sqz^{Df}), sqz^{ie} and *UAS-sqz* (Allan et al., 2003); *UAS-ap,* ap^{RK506} (ap^{LacZ}) (Cohen et al., 1992); ap^{P44} and ap^{md544} (ap^{GAL4}) (O'Keefe et al., 1998); dac^{3} (Mardon et al., 1994); *(Df(2L)Exel7066* (*Dac*^{Df}*); UAS-dac* (Shen and Mardon, 1997); $eya^{Cli\text{-}IID}$ (Pignoni et al., 1997); eya^{D1} (Bonini et al., 1998)*;* $dimm^{rev4}$ and $dimm^{P1}$ (Hewes et al., 2003); grh^{IM} (Wieschaus et al., 1984); *Df(2R)Pcl7B* (grh^{Df}) (Duncan, 1982) *, OK6-GAL4*, wit^{A12} and wit^{B11} (Aberle et al., 2002); svp^{1} (Mlodzik et al., 1990); *UAS-*Mad^{1} (Takaesu et al., 2005); *UAS-*$Glued^{A84}$ (Allen et al., 1999)*;* *UAS-*tkv^{Act} and *UAS-*sax^{Act} (Haerry et al., 1998); *UAS-myc::Mad* (Merino et al., 2009); *UAS-nls.EGFP* (Bloomington, IN). Mutants were kept over *CyO,Act-GFP TM3,Ser,Act-GFP* or *CyO,twiGAL4,UAS-2xEGFP* or *TM3,Sb,Ser,twiGAL4,UAS-2xEGFP*. w^{1118} was used as the control genotype. Flies were maintained at 25°C, 70% humidity**.**

2.3.2 EYFP Reporter Transgene Construction

Empty Tv-nEYFP vector was generated from pUASTattB (Bischof et al., 2007) digested with NheI/SpeI and blunted with Klenow fragment. The LoxP and attB sequences from pUASTattB and the multiple cloning site (MCS), HSP70 promoter, Tra nuclear localization

signal and SV40-polyA sequences from pHstinger (Barolo et al., 2000) were joined with EYFP from pDUAL-YFH1c using SOE PCR to produce an EcoRV-loxP-MluI-MCS-hsp70 pro-EYFP-*tra nls*-SpeI-SV40polyA-AvrII-attB-ZraI cassette that was digested with EcoRV/ZraI and ligated in to the blunted pUASTattB backbone. The Tv4-enhancer was PCR-amplified from Oregon-R with XbaI and EcoRI adaptors, restriction digested and ligated into XbaI/EcoRI digested empty Tv-nEYFP. SOE PCR generated nucleotide substitution and deletion mutants were inserted similarly. XbaI and NheI sites were used for the concatemers. A summary of all mutations and concatemerization sequences in Appendix C Table S 2. Fly transformations were performed by Genetic Services Inc. (Cambridge, MA.) All transgenic reporters were integrated into *attP2* (Groth et al., 2004).

2.3.3 Immunochemistry

Standard protocols were used throughout (Eade and Allan, 2009). *Primary antibodies*: Rabbit and Chicken α-FMRFa c-terminal peptide (1:1000) (Baumgardt et al., 2007); Rabbit α-FMRFamide (1:2000; T-4757 Peninsula Labs); Chicken α-ß-Gal (1:1000, ab9361, Abcam); Guinea Pig α-Dimm (1:1000) (Linkoping U., Sweden); Mouse α-Eya (1:100; clone 10H6) and mouse α-Dac (1:50; Dac2-3) (DSHB; Iowa U., Iowa); Rabbit α-pMad (1:100, 41D10, Cell Signaling Technology); Mouse α-Svp (1:50) (Benito-Sipos et al., 2011). *Secondary antibodies*: Donkey α-Mouse, α-Chicken, α-Rabbit, α-Guinea Pig conjugated to DyLight 488, Cy3, or Cy5 (1:100, Jackson ImmunoResearch).

2.3.4 Image Analysis

More than 5 animals were examined for every genotype. Analysis on the 445 bp enhancers was performed on homozygous reporter lines. Concatemerized *cis*-elements were analyzed as heterozygous reporters. Images were acquired with an Olympus FV1000 confocal microscope

with settings that avoided pixel intensity saturation. Fluorescent intensity of individual Tv4-neurons was measured (or from Eya-positive Tv cluster when no Tv4 marker was detectable) in Image J (US National Institutes of Health). Mean pixel intensity for each neuron was measured from summed Z-projection and background fluorescence for subtraction was measured from an adjacent location. Each datum point of resulting nEYFP intensity was used to calculate mean intensity for a genotype or enhancer variant; each datum point is then represented as a percentage of the mean of the control group. Representative images of Tv-neurons being compared in figures were contrast enhanced together in Adobe Photoshop CS5 (Adobe Systems, Mountain View, CA). All statistical analysis and graphing were performed using Prism 5 (GraphPad Software, San Diego, CA). All multiple comparisons were done with One-Way ANOVA and a Tukey *post-hoc* test or Student's two-tailed *t*-test when only two groups. Differences between groups were considered statistically significant when $p<0.05$. Data are presented as mean ± Standard Error of the Mean (SEM).

2.3.5 Recombinant Transcription Factor Expression and EMSA

Recombinant GST-LIMless-Ap (LIM domains removed) and GST-MadN (MH1 domain only), were fused to the GST in pGEX6p1 (GE Health), expressed in Rosetta *E. coli* (EMD Millipore, Billerica, MA), purified using Glutathione-Sepharose beads (GE Health), and dialyzed into 20mM HEPES pH 7.8, 50mM KCl, 1mM DTT, and 10% glycerol. Aliquots were stored at -80^{o}C. Synthetic oligonucleotide probes were labeled with IRDye 700 (IDT Inc, Coralville, IA). Gel shift assay for *HD-RE* and *BMP-RE* Apterous binding were performed by incubation (30min at 21^{0}C) of 1 μg GST-LIMless-Ap with 5 nM probe in a 20 μl reaction volume (20 mM HEPES pH7.8, 50 mM KCl, 10% glycerol, 0.25 mM EDTA, 0.1mg/ml BSA, 1mM DTT). Gel shift assay for Mad binding at the *BMP-RE* was performed by incubating 300 ng of GST-Mad

with 2.5 nM probe in a 20 µl reaction volume (25mM Tris pH7.5, 35mM KCl, 80mM NaCl, 5mM $MgCl_2$, 3.5mM DTT, 0.25% Tween 20, 1 µg poly dIdC, and 1X Protease Inhibitor cocktail (Roche)). For competition assays, a 500-1000 fold excess of unlabelled wild-type or mutant DNA sequences were co-incubated with the labeled probes. DNA-protein complexes were resolved on a 4% non-denaturing polyacrylamide gel, and imaged immediately on a Licor Odyssey Imager system (Lincoln, NE.)

2.4 Results

2.4.1 The Tv4-Enhancer Responds Appropriately to *FMRFa* Transcriptional Regulators

The *Drosophila* ventral nerve cord (VNC) has a bilateral pair of Tv4 neurons in thoracic segments T1-T3. Tv4 neurons express *FMRFa*, a neuropeptide gene that encodes a prepropeptide (FMRFa) that is processed to multiple amidated FMRFamide neuropeptides (FMRFamide). Tv4 axons innervate the ipsi-segmental dorsal neurohaemal organ (**Figure 8A**), where they access the BMP ligand Glass bottom boat (Gbb) that triggers *FMRFa* expression via retrograde BMP signaling. Most efferent VNC neurons have activated BMP signaling, but *FMRFa* is only expressed in the six Tv4 neurons due to BMP intersection with a Tv4-specific combination of transcription factors (TFs); the sequence-specific TFs Apterous (Ap), Squeeze (Sqz), Dimmed (Dimm) and Grainy head (Grh), and the transcriptional cofactors Eyes absent (Eya) and Dachshund (Dac) (Allan et al., 2005; Allan et al., 2003; Baumgardt et al., 2009; Baumgardt et al., 2007; Benveniste et al., 1998; Miguel-Aliaga et al., 2004) (**Figure 8B**).

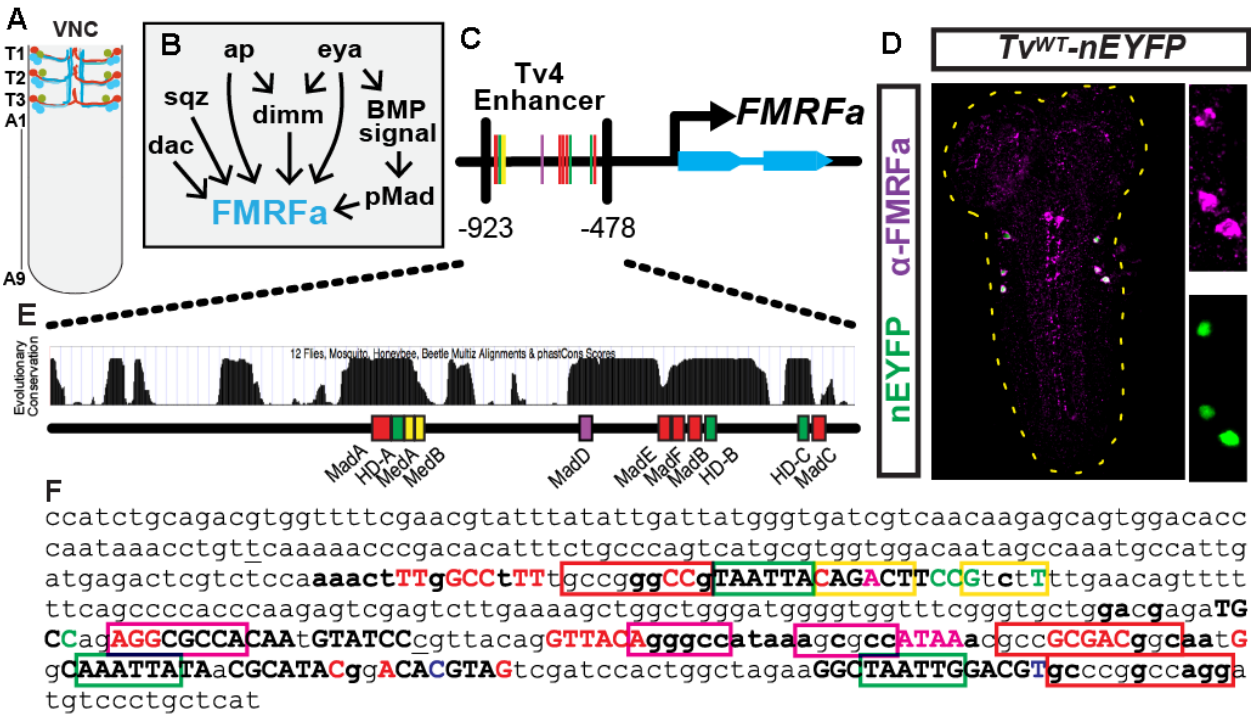


Figure 8. The Tv4-enhancer faithfully reports Tv4-FMRFa expression and contains conserved putative binding sites for FMRFa transcriptional regulators.

(**A**) Tv-neurons in the embryonic/larval VNC; Tv1 neurons (green), Tv2/3 neurons (blue) and Tv4 neurons (red). (**B**) Transcription factors postulated to directly regulate *FMRFa* in Tv4-neurons. (**C, D**) A 445 bp enhancer that drives Tv4-neuron expression. Representative image of Tv^{WT}*-nEYFP* expression only in Tv4 neurons. (**E**) Conservation histogram of the Tv4-enhancer across 12 *Drosophila* species (high peaks=best conserved) from UCSC Browser. Below, we show the location of putative homeodomain, Mad and Medea sites (see F), demonstrating clustering of these sequences within conserved regions. (**F**) Sequence of Tv4-enhancer from Oregon R showing conserved nucleotides (capitalized and/or bold), and putative Homeodomain (green box), Mad (red or magenta box) and Medea (yellow box) binding sites. Conservation of nucleotide identity shown using Relaxed EvoPrint (EvoprinterHD); showing two layers of conservation. First layer; bolded capital letters are nucleotides conserved in *D.simulans, D.sechellia, D.erecta, D.yakuba, D.ananassae, D.pseudoobscura, D.virilis, D.grimshawi or D.willistoni,* and colored bases are sequences present in all but one species. Second layer; small bolded letters are conserved in all species except *D.ananassae* and *D.willistoni*. SNPs compared to the *D. melanogaster* reference genome sequence are underlined.

A 445bp *cis*-regulatory region (the Tv4-enhancer) is sufficient for Tv4-specific expression and is *apterous*-dependent (Benveniste and Taghert, 1999; Benveniste et al., 1998). We amplified this sequence from Oregon R and found it to contain two single nucleotide polymorphisms (SNPs), compared to the reference genome. Recently sequenced wild *Drosophila* species concur with the Oregon R sequence, thus the reference genome contains the atypical SNPs (Langley et al., 2012) (**Figure 8F**). A reporter transgene containing the Tv4-enhancer (*Tv^{WT}-nEYFP*) was integrated into genomic *attP2* using phiC31-integrase that faithfully reports only in Tv4 neurons (**Figure 8C, D**). In early first Instar larvae (L1 larvae), we tested *Tv^{WT}-nEYFP* expression in genetic backgrounds that have previously been demonstrated to reduce FMRFa expression, by quantifying the number (per VNC) and intensity (neurons normalized to the mean fluorescence intensity of the control) of nEYFP in Tv4 neurons (**Figure 9A-C**).

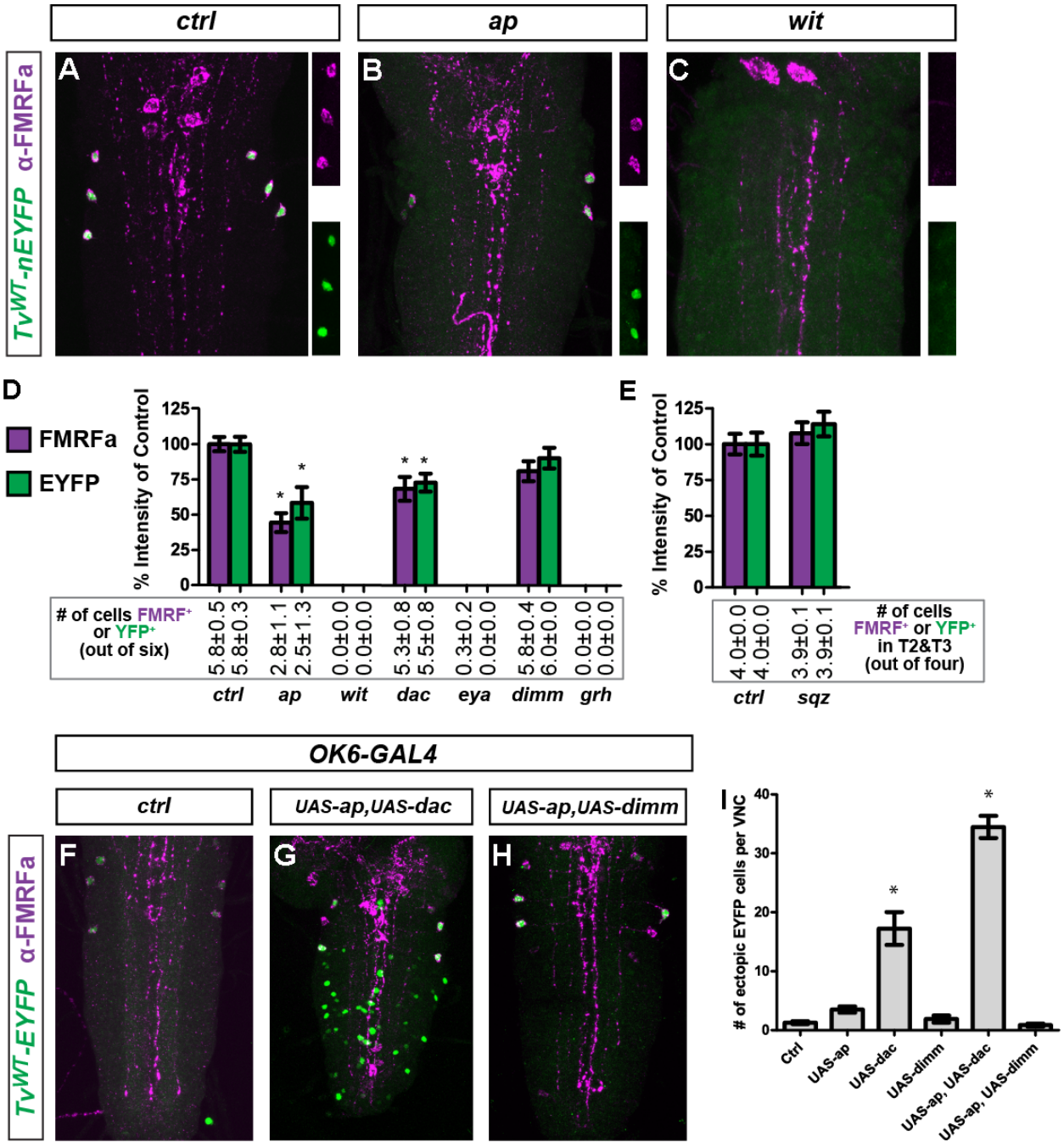


Figure 9. The Tv4-enhancer responds appropriately to known FMRFa transcriptional regulators.

(**A-C**) Representative images of Tv^{WT}-*nEYFP* (green) and FMRFa (magenta) in control (*ctrl*), *ap* and *wit* mutants, showing reductions in expression. **(D)** Graph showing % fluorescence intensity for FMRFa immunoreactivity and TV^{WT}-*nEYFP* reporter expression in mutants for FMRFa regulators, relative to the mean of the control. Numbers below columns represent the number of Tv4-neurons that express detectable FMRFa or TV^{WT}-*nEYFP* expression. (**E**) In *sqz* mutants, we show intensity and FMRFa expression for T2 and T3 segments only. Data shown as mean±SEM; compared by one-way ANOVA with Tukey HSD *post-hoc* test. *=p<0.05 to controls. n=10-20 animals per genotype. (**F-H**) *OK6-GAL4* drove *UAS-ap*, *UAS-dac* and *UAS-dimm*, alone and in combination. Total number of nEYFP expressing non-Tv4 neurons was counted. **(I)** Only *UAS-dac* and *UAS-ap, UAS-dac* induced ectopic nEYFP. Data shown as mean number of ectopic nEYFP cells ± SEM. n=10-20 VNCs per genotype. All data represented as mean±SEM and compared by one-way ANOVA with Tukey HSD *post-hoc test,* *=p<0.001 to controls.

Genotypes: Tv^{WT}*-nEYFP.* ***wit*** *(*Tv^{WT}*-nEYFP,*wit^{A12}*/*Tv^{WT}*-nEYFP,*wit^{B11}*).* ***ap****, (*ap^{GAL4}*/*ap^{P44}*;*Tv^{WT}*-nEYFP).* ***sqz*** *(*Tv^{WT}*-nEYFP,*sqz^{ie}*/*Tv^{WT}*-nEYFP,*sqz^{Df}*).* ***dac*** *(Df(2L)Exel7066/*dac^{3}*;*Tv^{WT}*-nEYFP).* ***eya*** *(*$eya^{Cli\text{-}IID}$*/*eya^{D1}*;*Tv^{WT}*-nEYFP).* ***grh****, (*grh^{IM}*/*grh^{Df}*;*Tv^{WT}*-nEYFP).* ***dimm*** *(*$dimm^{rev4}$*/*$dimm^{P1}$*;*Tv^{WT}*-nEYFP)*

Tv^{WT}-nEYFP exhibited mostly an identical dependence to known transcriptional regulators as the FMRFa neuropeptide (**Figure 9A-E, Figure S 10**). Loss of BMP signaling in *wit* nulls eliminated FMRFa immunoreactivity and *Tv^{WT}-nEYFP* expression (**Figure 9C,D**). In strong *ap* hypomorphs, *Tv^{WT}-nEYFP* was expressed in ~2.5 Tv4 neurons per VNC at 58% of control intensity (**Figure 9B,D**). The cofactor *dac* is only modestly required for FMRFa in embryos, but its overexpression upregulates FMRFa, and it acts combinatorially with *apterous* to trigger ectopic FMRFa in motoneurons (Miguel-Aliaga et al., 2004). Similarly, in *dac* nulls, *Tv^{WT}-nEYFP* is expressed in ~5.5 Tv4 neurons per VNC at 72% of control intensity (**Figure 9D**). Overexpression of Dac in Tv4 neurons (by *ap^{GAL4}*) upregulated *Tv^{WT}-nEYFP* to 144%±10% of control levels ($p<0.01$ Two-tailed t-test, n=48 and n=56 Tv4 neurons for control and overexpression, respectively). Also, ectopic *Tv^{WT}-nEYFP* expression was activated in motoneurons by *OK6-GAL4*-driven misexpression of *UAS-dac* alone, or *UAS-dac* and *UAS-ap* together (**Figure 9F-I, Figure S 10**). The cofactor *eya* is essential for FMRFa (Miguel-Aliaga et al., 2004), and *Tv^{WT}-nEYFP* was entirely eliminated in strong *eya* hypomorphs in late Stg 17 embryos (**Figure 9D, Figure S 10)**. In *grh* nulls, Tv4 neurons do not appear to be generated (Baumgardt et al., 2009), and predictably, *Tv^{WT}-nEYFP* expression was entirely eliminated (**Figure 9D, Figure S 10)**.

In contrast to previous evidence that FMRFa expression is *sqz* and *dimm*-dependent (Hewes et al., 2003), our data suggest that this regulation is not at the transcriptional level. *Tv^{WT}-nEYFP* was not downregulated in either *sqz* nulls (**Figure 9D, Figure S 11**), or in strong *dimm* hypomorphs. Regarding *sqz*, we confirmed that T1 segment Tv4 neurons are often not generated, and that supernumerary Nplp1-expressing Tv1 neuron are generated **(Figure S 11**), as reported (Allan et al., 2005; Baumgardt et al., 2009; Baumgardt et al., 2007). Thus, we quantified *Tv^{WT}-*

nEYFP number and intensity in Tv4 neurons in segments T2 and T3, but found no significant effect on FMRFa or Tv^{WT}*-nEYFP* expression in *sqz* mutants (**Figure 9E)**. Regarding *dimm*, in *dimm* mutants we observed 47%±6% FMRFamide immunoreactivity staining intensity compared to controls ($p<0.001$ Two-tailed t-test, n=48 and n=36 Tv4 neurons for $dimm^{Rev4}/dimm^{P1}$ *and* $dimm^{Rev4}/+$ controls, respectively). In contrast, we observed no significant reduction in FMRFa immunoreactivity or Tv^{WT}*-nEYFP* fluorescence intensity (**Figure 9D)**. Similarly, co-misexpression of *UAS-dimm* with *UAS-ap* in motoneurons failed to trigger ectopic Tv^{WT}*-nEYFP* expression (**Figure 9G**). We previously reported that *dimm* regulated FMRFamide processing from FMRFa, but not *FMRFa* transcript levels in adult Tv4 neurons (Eade et al., 2012). This corresponds to data showing that *dimm* is required for upregulating convertase and amidase enzymes, as well as the secretory machinery (Allan et al., 2005; Hamanaka et al., 2010; Park et al., 2008), which together generate and store mature amidated neuropeptide. Finally, Dimm binds typical E-box sequences (Park et al., 2008), yet the E-box-like sequences in the Tv4-enhancer are poorly conserved and 5' of the HD-RE, a region that is not required for wild type Tv4-enhancer activity (see below).

2.4.2 The Tv4-Enhancer Has Conserved Putative Homeodomain and Smad Binding Motifs

Our data identify Smads and Apterous as primary sequence-specific regulators of the Tv4-enhancer; thus, we wished to identify possible binding motifs in the Tv4-enhancer. First, we identified conserved sequences within the Tv4-enhancer, across twelve sequenced *Drosophila* species (Clark et al., 2007; Stark et al., 2007), using phastCONS in the UCSC Genome Browser and EvoprinterHD (Yavatkar et al., 2008) (**Figure 8E-F, Figure S 12 & Figure S 13**). Amongst these conserved sequences, we located putative Apterous and Smad-binding motifs, based on

published sequence preferences (Benveniste et al., 1998; Noyes et al., 2008; Pyrowolakis et al., 2004; Weiss et al., 2010). All three previously-identified Apterous-binding sequences (Benveniste et al., 1998) are conserved, and denoted HD-A,B,C (**Figure 8E,F, Figure S 12 & Figure S 13**). Intriguingly, each HD motif is immediately adjacent to GC-rich sequences with similarity to Mad-binding sequences (Kim et al., 1997; Xu et al., 1998) that we denote Mad-A,B,C (**Figure 8E,F, Figure S 12 & Figure S 13**). These sequences show phylogenetic degeneracy, but a GC-rich sequence is retained in all species (**Figure S 12 & Figure S 13**). We also found other Mad-like motifs, Mad-D,E,F. Of these, only Mad-D is highly conserved and precisely matches a well-characterized *Drosophila* Mad-binding sequence [GGCGCCA] (Weiss et al., 2010) (**Figure 8E,F, Figure S 12 & Figure S 13**). *Drosophila* Smads often act at bipartite Mad and Medea binding motifs, such as the bipartite *Drosophila* BMP-activator element [GGCGCCA(n4)GNCV] or BMP-silencer element [GRCGNC(n5)GTCT] (Pyrowolakis et al., 2004; Weiss et al., 2010). The Tv4-enhancer has two putative Medea-binding sequences (GNCV or GTCT) near Mad-A, denoted Med-A,B. Only Med-A is well conserved, and Med-A and Med-B are separated from Mad-A by 6bp and 15bp, respectively (**Figure 8E,F, Figure S 12 & Figure S 13**). Mad-A and Med-A represent a potential bipartite motif [GGGCCG(n6)CAGAC], although the Med-A motif is in reverse complement orientation to the ideal BMP-Silencer Element, a change found to prevent Smad complex-binding (Gao and Laughon, 2007).

2.4.3 Two Highly Conserved *Cis*-Elements Are Necessary for Tv4-Enhancer Activity *In Vivo*

Homeodomain TFs act cooperatively or collaboratively with Smads at coupled motifs at other *cis*-elements (Brugger et al., 2004; Grocott et al., 2007; Suszko et al., 2008; Walsh and Carroll, 2007; Zhou et al., 2008). Such a model could account for restriction of FMRFa

expression in Tv4 neurons, as Ap and BMP signaling only coincide in Tv4 neurons in the VNC. Thus, we proceeded to test which of the three, coupled HD/Mad regions in the Tv4-enhancer may provide a substrate for Ap and Smad cooperativity.

To identify essential HD and Smad-binding sequences, we performed substitution and deletion mutagenesis studies of the Tv4-enhancer, placing each mutant Tv4-enhancer reporter transgene into *attP2*, to allow for quantitative comparison of reporter expression. We quantified the number of Tv4 neurons with nEYFP (in early L1 larvae), as well as fluorescence intensity normalized to the mean of the control (**Figure 10A-C).** A reporter-only empty vector control was used as the relative zero, Tv4 neurons with nEYFP reporter intensity above the upper 99% confidence interval for the empty vector control (9.7% of Tv^{WT}-*nEYFP*) were counted as expressing nEYFP. Most mutants partially reduced reporter expression, but only two profoundly reduced expression (**Figure 10C**). Mutant HD-A ($Tv^{\Delta HD\text{-}A}$-*nEYFP*) severely reduced expression to 0.6 Tv4-neurons per VNC at 3.9% of control intensity. Mutant Mad-D ($Tv^{\Delta Mad\text{-}D}$-*nEYFP*) essentially eliminated expression. HD-A and Mad-D motifs are contained within ~30bp highly conserved regions that are separated by ~100bp of poorly conserved sequence. To assess the relative importance of these two *cis*-elements, we deleted the conserved islands containing either the HD-A or the Mad-D *cis*-elements. Both eliminated reporter expression. Thus, both *cis*-elements are required non-redundantly (**Figure 10B4,B5**). Deletion of regions outside these *cis*-elements had partial or no effect on reporter expression. Intriguingly, deletion of the sequence between the conserved HD-A and Mad-D *cis*-elements also dramatically reduced reporter expression (**Figure 10B7**). To determine whether this region acts as a spacer or has informational content, we mutated it in two ways; changing every second base to the complement sequence to maintain local GC/AT content, and also a non-canonical nucleotide

transversion (Evans et al., 2012). In both cases, reporter expression was dramatically reduced (**Figure 10B8,B9**). Thus, this intervening sequence contains unknown informational content that we believe is import for HD-A and Mad-D region activity. We did not observe any ectopic reporter expression, suggesting that no prominent repressor activity exists amongst any of these sequences.

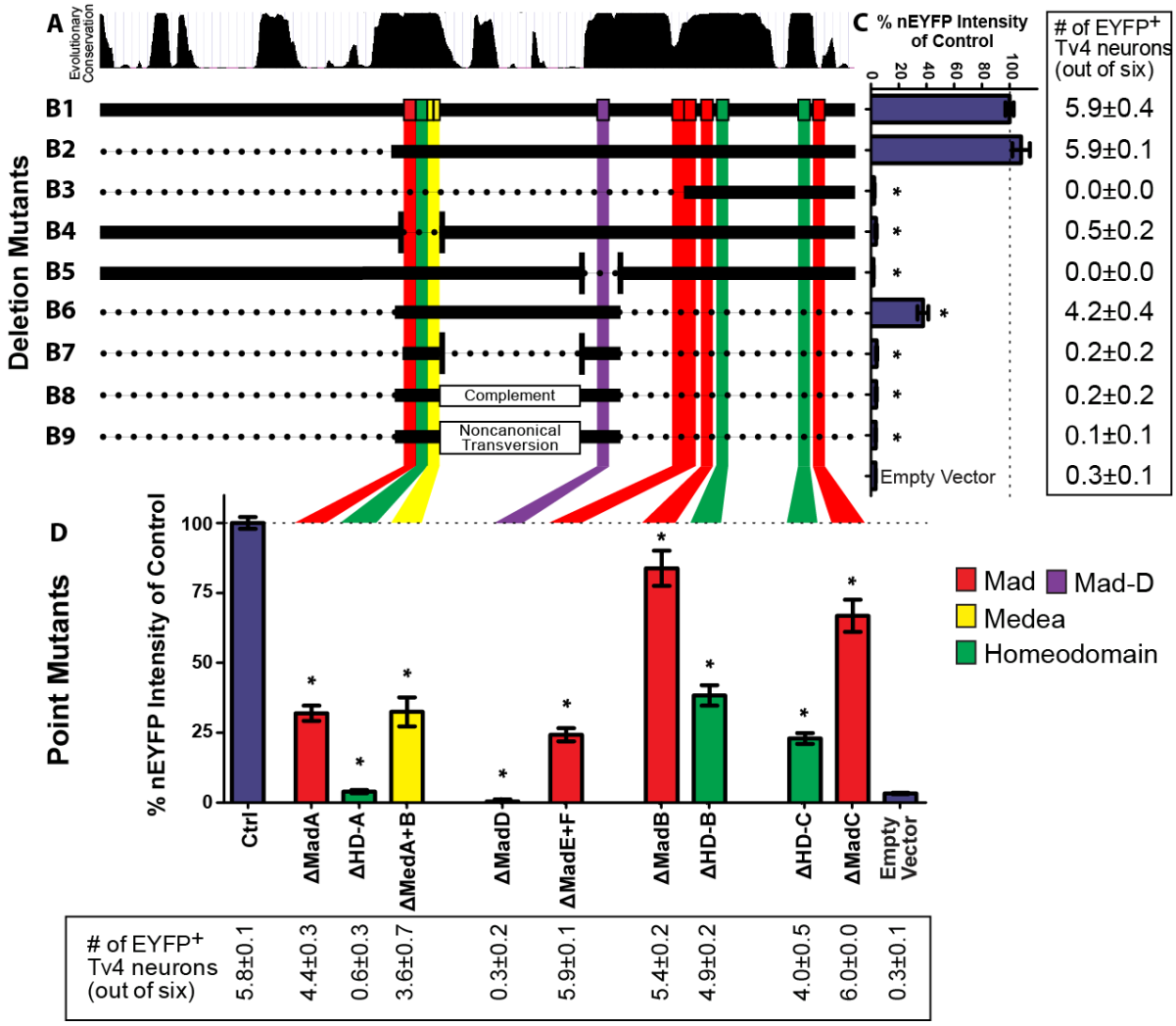


Figure 10. Mutant sequence analysis of the Tv4-enhancer.

(**A**) UCSC Browser conservation track for the Tv4-enhancer. (**B**) Location of putative HD (green), Mad (red/magenta) and Med motifs (yellow), shown as coloured vertical bars. We generated deletion transgenes (B1-B9; dotted region is deleted). **B1** is wild type, denoted by the complete horizontal black line. **B2-B9** show the region present in the reporter (horizontal black bar), and the deleted regions (dotted regions). **B4,5,7** show an internal deletion between fusion points that are denoted by vertical bars. **B8,9** describes the type of sequence conversion made within the region shown. (**C**) Graph shows reporter expression of control or deleted enhancers, and an empty reporter vector. This shows % nEYFP intensity relative to mean of B1 control, as well as the number of Tv4-neurons (out of six) expressing nEYFP above the 99% confidence interval of the empty vector control. Deletion of HD-A (B3,B4) or Mad-D (B3,B5) severely reduced expression. The region spanning HD-A to Mad-D motifs express at a moderate level, but only if the intervening sequence is unaltered (B7-B9). (**D**) Graph shows expression of point mutants at Mad, HD and Med motifs. Only mutations that alter HD-A or Mad-D caused severe reduction of reporter expression. n=10-20 animals per genotype. Data shown as mean±SEM and compared by one-way ANOVA with Tukey HSD *post-hoc* test, *=$p<0.05$ to controls.

2.4.4 The HD-RE (HD-A-Containing) and BMP-RE (Mad-D-Containing) *Cis*-Elements Each Encode Sufficient Information for Tv4-Neuron Spatial Expression

What *cis*-regulatory information is encoded within the HD-A and Mad-D elements? We generated reporter constructs with concatemeric direct repeats of the conserved regions surrounding HD-A (25bp spanning Mad-A, HD-A, Med-A) and Mad-D (39bp). As expected from our mutant analysis, monomers of each *cis*-element did not generate reporter expression (not shown). However, concatemeric repeats of 2x or more generated reporter expression in Tv4 neurons with robust, consistent expression being observed by *6xHD-A-nEYFP* and *4xMad-D-nEYFP* reporters. Their expression was highly restricted to Tv4-neurons. Ectopic expression was only observed in small numbers of cells in the brain (*6xHD-A-nEYFP* only) and L3 eye imaginal disc (*4xMad-D-nEYFP* only) (**Figure S 14**).

Thus, remarkably, both *cis*-elements have sufficient information for highly Tv4-specific expression in the VNC (**Figure 11B,D).** We further found that these two *cis*-elements are uniquely sufficient for Tv4 expression, as concatemerization of any other Tv4-enhancer sequences failed to generate Tv4-neuron expression, but instead either generated no expression or weak ectopic expression (**Figure 11C,E,F**).

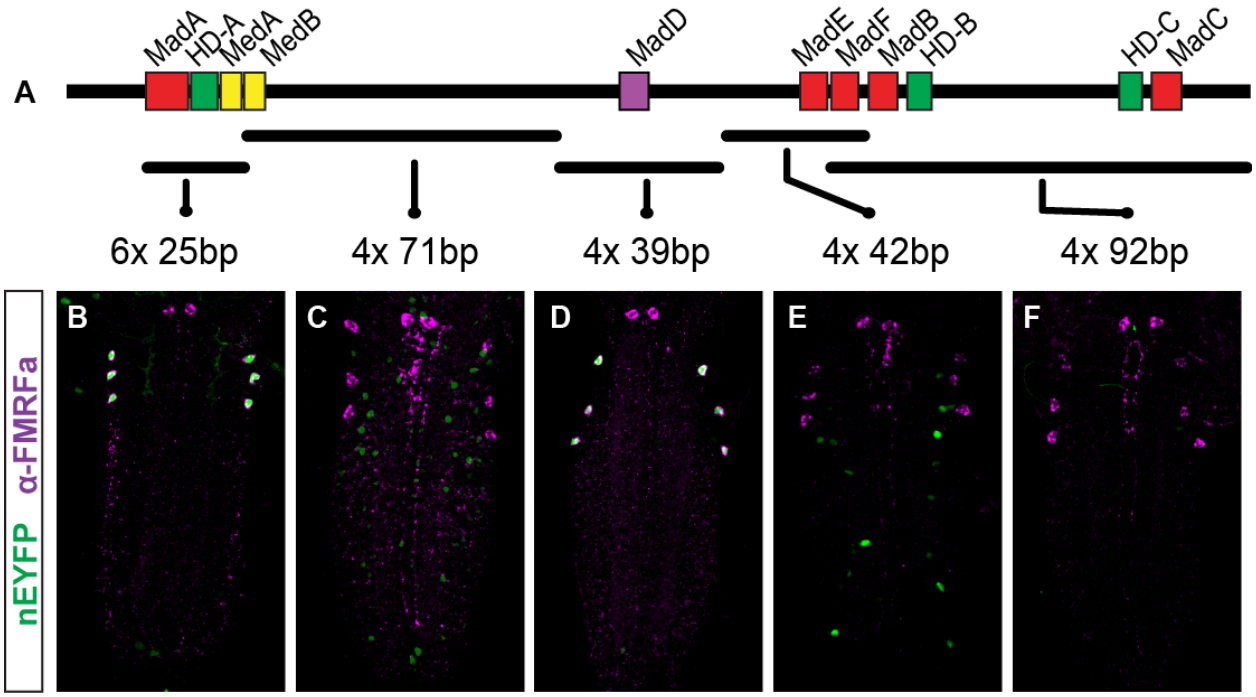


Figure 11. The 6xHD-A and 4xMad-D cis-elements encode sufficient information for Tv4-specific expression.

(**A**) Relative position of putative HD, Mad and Medea motifs. We concatemerized fragments from the Tv4-enhancer (shown in A). Horizontal line indicates region concatemerized. The number of concatemeric repeats is shown, as is nucleotide number within each repeated sequence. (**B-F**) Reporter expression for each concatemer in early L1 larval VNC. Only the 25bp HD-A-containing conserved region (**B**) and the 39bp Mad-D containing conserved region (**D**) generate reporter expression in Tv4-neurons.

These faithful Tv4-specific reporters, that include the only required sequences for Tv4-enhancer activity, allowed us to determine which transcriptional regulators act through each cis-element. First, we found a profound difference in BMP-dependence between the two cis-elements. 6xHD-A-nEYFP was entirely unaffected in wit mutants, even though this region is the only one with a putative bipartite Mad and Med motif. Further, no change in 6xHD-A-nEYFP expression was observed after blocking retrograde BMP signaling, using ap^{GAL4} to drive UAS-$Glued^{\Delta 84}$, a truncated allele of $p150^{Glued}$ that blocks dynein-dependent retrograde transport, nuclear pMad accumulation and FMRFa expression (Allan et al., 2003) (**Figure 12A**). Thus, this *cis*-element does not respond to target-derived signaling yet encodes Tv4-specific expression.

In contrast, *4xMad-D-nEYFP* was severely reduced in *wit* nulls and by expression of *UAS-Glued$^{\Delta 84}$* (**Figure 12C**). Overexpression of Mad^1 (*UAS-Mad^1*) also eliminated *4xMad-D-nEYFP* expression (**Figure 12C**). Mad^1 cannot bind DNA but is phosphorylated, couples to Medea and accumulates in the nucleus normally (Takaesu et al., 2005). We tested the sequence-specificity of Mad binding to Mad-D by electrophoretic mobility shift assay (EMSA) (**Figure 12G**). Purified GST-Mad band shifted a Mad-D region DNA probe in a GGCGCC sequence-specific manner (**Figure 12G**). Henceforth, we termed this *cis*-element the BMP-Response Element (BMP-RE).

In *ap* mutants, *6xHD-A-nEYFP* was significantly down-regulated to 27% of controls (**Figure 12A**). This was expected due to the perfect Ap-binding sequence in HD-A, previous biochemical evidence for Ap binding (Benveniste et al., 1998), and the importance of the HD-A to Tv4-enhancer activity. EMSA analysis further supported this; GST-Ap band-shifted an HD-A sequence DNA probe in a TAATTA sequence specific manner (**Figure 12E**). Unexpectedly, *4xMad-D-nEYFP* was also reduced to 23% of control intensity in *ap* mutants (**Figure 12D**), in

spite of the lack of a putative Ap binding site. GST-Ap failed to band shift the BMP-RE sequence, thus the regulation by Apterous appears to be indirect **(Figure 12F)**. We henceforth term the HD-A *cis*-element the Homeodomain-Response Element (HD-RE). Finally, we examined HD-RE and BMP-RE responsiveness to the other regulators of Tv^{WT}*-nEYFP*. Both *cis*-elements are eliminated in *eya* mutants (**Figure 12A,E**) and upregulated by ~400% by *dac* overexpression in Tv4 neurons (**Figure 12B,D**). Thus, both co-regulators coordinately regulate trans-activation from both *cis*-elements.

Together, these data reveal that the native Tv4-enhancer has two separable, necessary *cis*-elements; the Ap-recruiting HD-RE and a Mad-recruiting BMP-RE that must synergize for enhancer activity, despite both *cis*-elements encoding sufficient information for Tv4-specific expression. We specifically choose the term 'synergize' here rather than 'additive' because monomers of either response element, or larger enhancers that have either response element deleted, essentially do not express any appreciable reporter activity. Instead each minimal element produces a combined effect greater than the sum of their separate activities when placed together.

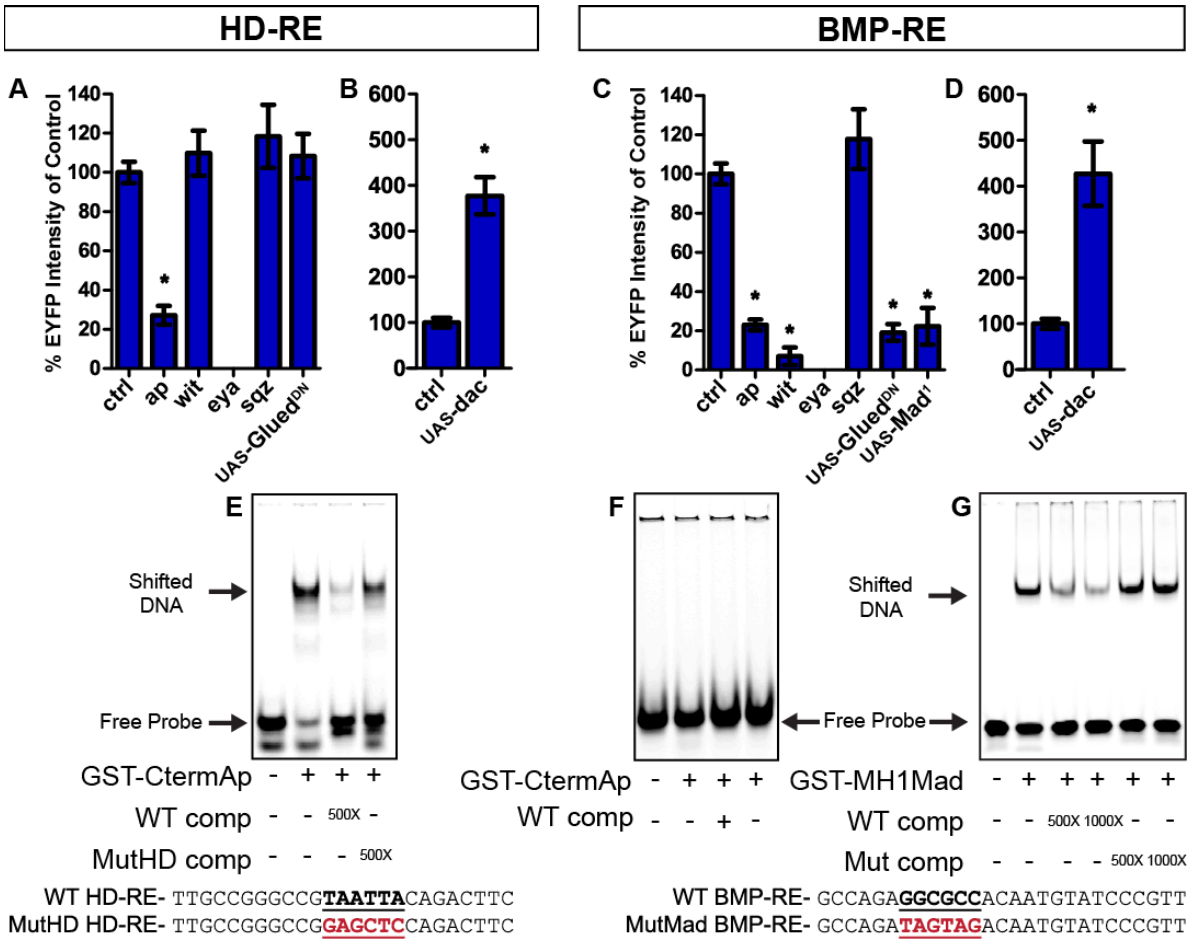


Figure 12. Genetic and biochemical analysis of HD-RE and BMP-RE.

(**A-D**) Expression of HD-RE (*6xHD-A-nEYFP*) and BMP-RE (*4xMad-D-nEYFP*) reporters in mutant and gain of function genotypes, shown as the % intensity of the mean of the control. (**E-G**) EMSA shows sequence-specific binding of Apterous to the HD-A motif in HD-RE and of Mad to the Mad-D motif in BMP-RE. (**A**) HD-RE expression is reduced in *ap* and *eya* mutants, but not in *wit* or *sqz* mutants or upon *UAS-Glued*^{DN} expression. (**B**) *UAS-Dac* expression increased *HD-RE* expression in Tv4-neurons. (**C**) Expression of BMP-RE is reduced in *ap, wit* and *eya* mutants, and by expression of *UAS-Mad*^{1} or *UAS-Glued*^{DN}. (**D**) *UAS-Dac* expression increased *BMP-RE* expression in Tv4-neurons. (**E**) HD-RE labeled probe (sequence shown below; WT HD-RE) was shifted in the presence of GST-LimlessAp and out-competed by an excess of wild type unlabeled probe (WT comp), but not by unlabeled mutant HD-A probe (Mut comp). (**F**) BMP-RE is not shifted in the presence of GST-LimlessAp. (**G**) BMP-RE is shifted by GST-MadN binding, and this is out-competed by increasing amounts of unlabeled wild type probe, but not by unlabeled mutant Mad-D probe. Data shown as mean±SEM and compared by one-way ANOVA with Tukey HSD *post-hoc* test,*=p<0.05 to controls, n=10-20 animals per genotype. ***Genotypes: ctrl*** *(Tv*$^{[cis\text{-}element]}$*/+).* ***wit*** *(Tv*$^{[cis\text{-}element]}$*,wit*^{A12}*/wit*^{B11}*).* ***ap****, (ap*^{GAL4}*/ap*^{P44}*; Tv*$^{[cis\text{-}element]}$*/+).* ***sqz*** *(Tv*$^{[cis\text{-}element]}$*,sqz*^{ie}*/sqz*^{ie}*).* ***eya*** *(eya*$^{Cli\text{-}IID}$*/eya*^{D1}*;Tv*$^{[cis\text{-}element]}$*/+).* ***uas-dac*** *(OK6-GAL4/+;Tv*$^{[cis\text{-}element]}$*/+* vs**.** *OK6-GAL4/+;Tv*$^{[cis\text{-}element]}$*/UAS-dac).* ***UAS-Mad***^{1} (*ap*^{GAL4}*/+;Tv*$^{[cis\text{-}element]}$*/+* vs**.** *ap*^{GAL4}*/UAS-Mad*^{1}*;Tv*$^{[cis\text{-}element]}$*/UAS-Mad*^{1}). ***UAS-Glued***^{DN} (*ap*^{GAL4}*/+;Tv*$^{[cis\text{-}element]}$*/+* vs**.** *ap*^{GAL4}*/ UAS-Glued*^{DN}*;Tv*$^{[cis\text{-}element]}$*/+*).

2.4.5 BMP-Signaling and Seven Up Coordinate the Timing of FMRFa Initiation

We postulated that the BMP-RE must provide the temporal switch for FMRFa initiation. In such a model, insufficiency of an HD-RE monomer would require target contact and BMP-RE activity for timely gene activation. This predicts that *6xHD-A-nEYFP* initiates reporter expression prior to target contact since all known HD-RE regulators are present in Tv4-neurons by late Stg 14. Unexpectedly, however, we found that HD-RE initiates reporter activity at the same time as BMP-RE and FMRFa (**Figure 13A-D**). This paradoxical observation is not explained by a second retrograde signal, as *UAS-Glued*$^{\Delta 84}$ had no effect on the HD-RE reporter (**Figure 12A**). Thus, we predicted that HD-RE responds to a novel timer that runs coincidentally to target contact.

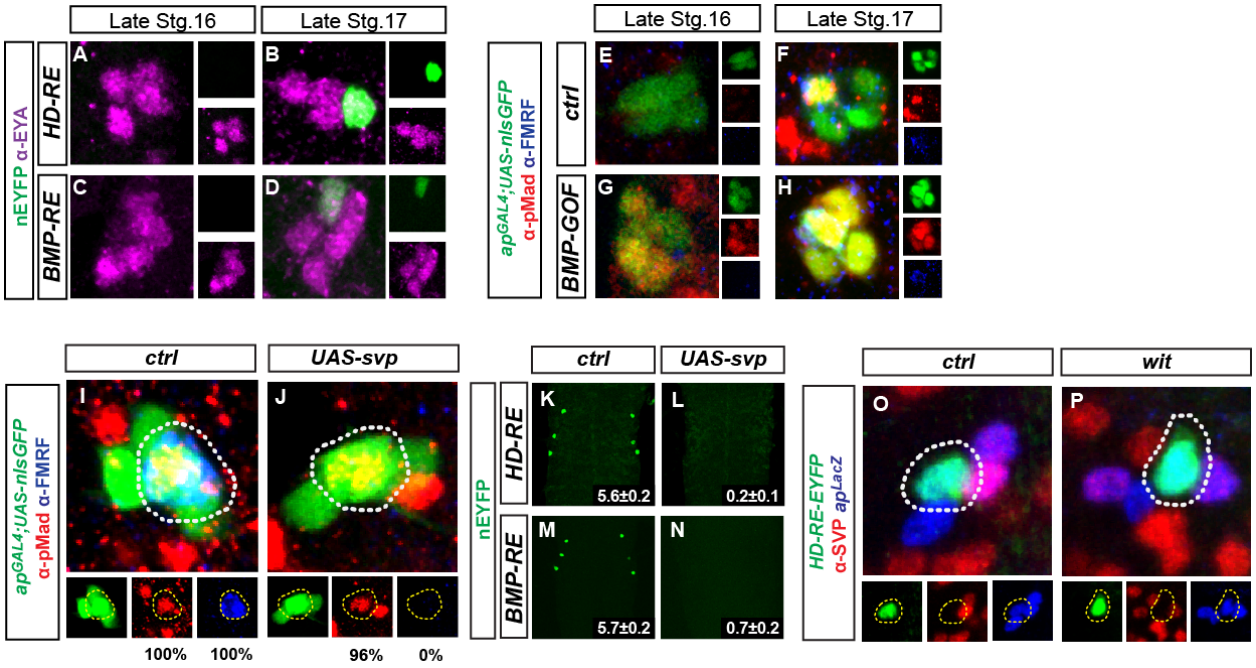


Figure 13. Svp represses FMRFa via the HD-RE and BMP-RE.

(**A-D**) HD-RE and BMP-RE reporters are not expressed at late Stg 16 but are expressed by late Stg 17 (after Svp loss). (**E-H**) Precocious activation of BMP (shown by pMad immunoreactivity, red) is achieved at Stg 16 in Tv-neurons (green) by expression of *UAS-tkv^{Act},UAS-sax^{Act}* and *UAS-myc::Mad*. However, in these animals, FMRFa expression (blue) is initiated at the same time as controls; late Stg 17. (**I,J**) Persistent *UAS-svp* expression eliminates FMRFs immunoreactivity but does not prevent pMad accumulation in Tv4 nuclei. The fraction of Tv clusters with pMad or FMRFa expressing neurons is shown (n=30 for controls; n=48 for *UAS-svp*). (**K-N**) HD-RE and BMP-RE reporter expression is eliminated by persistent *UAS-svp*. Mean number of nEYFP-positive Tv4-neurons per animal ± SEM is shown (n=10-16 animals per group, $p<0.001$ two-tailed *t*-test to controls). (**O,P**) Svp expression is not detectable at early L1 in control or *wit* mutant animals. Thus, BMP signaling is not required to downregulate Svp. The *HD-RE-nEYFP* is used to identify Tv4 in these *wit* mutants. Svp is detected in Tv2 and Tv3 neurons. ***Genotypes***: **BMP gain of function** (*ap^{GAL4}/+* vs. *ap^{GAL4}/UAS-tkv^{Act},UAS-sax^{Act};+/UAS-myc::Mad/+*). ***Svp* gain of function** (*ap^{GAL4}/+* vs. *ap^{GAL4}/+;+/UAS-svp)* or (*ap^{GAL4}/+; Tv$^{[cis-element]}$/+* vs. *ap^{GAL4}/+; Tv$^{[cis-element]}$/UAS-svp).* **BMP loss of function** (*ap^{lacZ}/+;wit^{A12}/ Tv$^{6xHD-A-nEYFP}$* + vs. *ap^{lacZ}/+;wit^{A12}/ wit^{A12},Tv$^{6xHD-A-nEYFP}$*)

The existence of a second timer would be further supported if precocious BMP activity in Tv4-neurons could not precociously initiate *FMRFa* expression. We tested this using ap^{GAL4} to drive excess Mad (*UAS-myc::Mad*) constitutively phosphorylated by co-expressing $UAS\text{-}tkv^{Act}$ and $UAS\text{-}sax^{Act}$ (Allan et al., 2003). This generated high pMad immunoreactivity in Tv-neurons prior to target contact by Stg 16, and highest in Tv4 neurons (Allan et al., 2003) (**Figure 13E-H**), yet precocious FMRFa immunoreactivity was not observed at late Stg 16 (**Figure 13G;** n=42 Tv clusters) or mid Stg 17 (n=48 Tv clusters). Not until the normal time of FMRFa activation at late Stg 17 is FMRFa detectable, even when BMP signaling is precociously and highly activated (**Figure 13F,H**). Thus, initiation of FMRFa in Tv4 neurons does not simply await target contact, as is generally assumed for target-dependent gene expression, but instead competence to respond to the target-derived signal appears to be gated via loss of a repressor.

Towards identifying a second putative timer, we considered the nuclear receptor, *seven up (svp)*, which is downregulated in Tv4 neurons during early to mid Stg 17, immediately prior to FMRFa initiation. Overexpression of Svp blocks *dimm*, *Nplp1* and *FMRFa* in Tv1 and/or Tv4-neurons (Benito-Sipos et al., 2011). Together with *svp* null phenotypes, *svp* was postulated to regulate differential generation of Tv-neurons. However, the coincidental timing of Svp downregulation and FMRFa initiation led us to test if Svp acted after lineage-determination in Tv4 neurons to repress FMRFa until mid Stg 17.

An essential biphasic role for *svp* in the NB5-6T neuroblast lineage (that gives rise to Tv-neurons) precludes a loss of function analysis (Benito-Sipos et al., 2011). We verified key findings of that study. First, we confirmed that late Stg17 svp^{1} null mutants lose the Eya-positive Tv-cluster in 52% of hemisegments (n=54). Second, in 70% of Tv clusters that do form, we observed supernumerary eya expressing cells with an average of 5.1±1.5 neurons per Tv cluster

(n=26 clusters). Previous analysis concluded that Tv-neurons showed identity confusion and an absence of Tv4 neurons in svp^{1} (Benito-Sipos et al., 2011). We furthered these studies by showing that pMad immunoreactivity was absent in any Tv-neuron in svp^{1} mutants. Could this lack of BMP signaling lead to a loss of FMRFa in *svp* nulls? We tested this using the BMP-unresponsive *6xHD-A-nEYFP* reporter, but found that it was not expressed in svp^{1} nulls. Thus, as previously postulated, the Tv4 neuron is not generated in svp^{1} nulls.

We maintained Svp expression by driving *UAS-svp* from ap^{GAL4}, and found that this eliminated FMRFa, HD-RE and BMP-RE reporters (**Figure 13I-N**). However, pMad accumulation in Tv4 nuclei was not blocked; thus, unique Tv4-neuronal identity and retrograde BMP signaling were unaffected by persistent Svp (**Figure 13I,J**). This shows that Svp overexpression does not block the generation of Tv4-neurons, but it acts to block terminal differentiation of neuropeptidergic fate. Next, we examined whether BMP signaling downregulates Svp. Using *6xHD-A-nEYFP* to identify Tv4-neurons in *wit* mutants, we found that Svp expression was not maintained beyond target contact even when BMP signaling was totally abrogated (**Figure 13O,P**).

Thus, we conclude that FMRFa initiation is temporally regulated by two timers. First, Svp acts as an intrinsic repressor that permits HD-RE and BMP-RE activity via timed de-repression and also coordinates neuropeptide expression with the capacity to process the neuropeptide (via *dimm* de-repression). Second, target-activated BMP signaling directly activates FMRFa via direct Mad activity at the BMP-RE *cis*-element. Third, although these two events are temporally coincidental, there is no evidence for a cross-regulatory genetic hand-over from one timer to the other.

2.5 Discussion

We uncovered the *cis*-regulatory logic and underlying enhancer architecture through which intrinsic and extrinsic regulatory inputs of *FMRFa* in Tv4 neurons are detected and interpreted. Two levels of integration are apparent. First, appropriate spatiotemporal information is independently encoded by two separable conserved *cis*-elements. Each element integrates distinct inputs of the overall FMRFa combinatorial code; HD-RE mediates Apterous binding whereas the BMP-RE mediates Smad binding. A second level of integration is found in the synergism of HD-RE and BMP-RE. Although each *cis*-element contains sufficient spatiotemporal information when concatemerized, they are insufficient as monomers within the Tv4-enhancer. Thus, the critical BMP/Apterous inputs are detected at separate *cis*-elements that synergistically drive Tv4-specific gene expression. Both *cis*-elements are responsive to the non-DNA binding transcriptional cofactors, Dac and Eya; thus, they likely act as co-activators at both *cis*-elements.

Smads bind cognate DNA sequences with low affinity (Shi et al., 1999; Zawel et al., 1998) and typically require interaction with other transcription factors for transcriptional activity (Blitz and Cho, 2009 ; Ross and Hill, 2008a). Functionally, this provides a mechanism for cell-specific and BMP-dependent gene expression (Blitz and Cho, 2009 ; Ross and Hill, 2008a). Cooperative or collaborative interactions between homeodomain transcription factors and BMP (or TGFβ)-activated Smads have been commonly observed (Brugger et al., 2004; Grocott et al., 2007; Lamba et al., 2008; Li et al., 2006; Liang et al., 2012; Marty et al., 2001; Suszko et al., 2008; Walsh and Carroll, 2007; Xu et al., 2005; Zhou et al., 2008). Although untested in neurons, lessons from such studies predicted cooperativity between Apterous and Smads at the Tv4-enhancer, as their expression/activity only intersects in Tv4 neurons, and they genetically interact to control FMRFa. However, our studies show that Apterous and Smads bind at distinct

cis-elements. This leaves open the question of what interacts with BMP signaling, and also Ap, at their respective *cis*-elements that contributes to the exquisite cell-specificity of these short DNA sequences. Presumably, unidentified necessary sequence-specific transcription factors act at the conserved sequences flanking those that bind Ap or Mad. Ongoing studies aim to identify these transcription factors and test their activities. See Section 4.3.2 for a discussion on how new regulators can be identified and screened for.

It is unclear why two separable spatially- and temporally-encoding *cis*-elements are utilized, when either one could conceivably function alone. One clue may derive from the ectopic, but non-overlapping, expression that concatemers of each *cis*-element generates. Such non-overlapping ectopic expression may suggest that both *cis*-elements are attenuated in the Tv4-enhancer to drive expression only in cells where both *cis*-elements are activated. Indeed, attenuation of *cis*-element activity to restrict target gene expression has been demonstrated, via reduced transcription factor affinity or by inclusion of repressive elements (Jiang and Levine, 1993; Lin et al., 2006). However, without evidence for instructional attenuation within HD-RE and BMP-RE, it is difficult to assess whether ectopic expression of the concatemers derives from an alleviation of attenuation, or is an artefact of the concatemerization. A second clue derives from studies showing that robust and specific expression can be achieved by the accumulation of multiple low activity *cis*-elements. Examination of Ultrabithorax-mediated repression of *spalt* in the developing haltere showed that multiple Ultrabithorax monomer binding sites are required together for spatially-restricted repression in this tissue (Galant et al., 2002). More recently, the accumulation of multiple weak Ultrabithorax-Extradenticle binding sites were shown to be required together to confer robust and specific expression of *shaven baby* in *Drosophila* epidermal tricomes (Crocker et al., 2015). Thus, the use of two separable individually

insufficient *cis*-elements that recruit distinct inputs, yet each provide correct spatiotemporal expression, may be a solution to ensuring robust and exquisitely specific expression in only 6 neurons of the nervous system. All sequenced *Drosophila* species to date appear to have very well conserved HD-RE and BMP-RE elements with similar spacing.

Our finding that precocious activation of BMP-signaling in Tv4 neurons failed to activate precocious FMRFa activation was unexpected, but it provides an explanation for the delayed activation of the BMP-unresponsive HD-RE. It is generally assumed, although without experimental evidence, that target-dependent genes are primed (or pre-specified) for expression prior to target contact, and merely await the target-derived signal for activation. This is a reasonable hypothesis given that the majority of target-dependent genes function in target communication (neurotransmitter or neuropeptide phenotype), mature electrophysiological properties (ion channels), or intra-target branching (via activation of ETS-type transcription factors). It seems reasonable that these are not activated until target contact, and a target-derived signal appears a robust mechanism. However, we show that such genes may be repressed prior to target contact, rather than solely requiring target-induced activation. We find that downregulation of nuclear receptor Seven up prior to target contact de-represses the Tv4-enhancer, making it competent for target-dependent activation. A cross-regulatory mechanistic hand-off from loss of an intrinsic repressor to gain of an extrinsic activator may have been an attractive model; indeed, BMP signaling regulates COUP-TF-I and -II (Seven up ortholog) in vertebrates (Satoh et al., 2009). However, we find that these timers are not cross-regulatory.

Why would target-dependent genes be repressed prior to target contact? Perhaps this ensures fidelity in gene activation; although, this would require demonstration of precocious activation in the absence of both timers. This is not currently possible with regard to the Seven

up repressor because null *seven up* mutants have lineage defects that prevent Tv4 neuronal specification. However, we propose an additional role for such a binary timer switch. We postulate that downregulation of the intrinsic repressor may be tied to completion of neuronal specification, and functions to block terminal differentiation until the specification process is completed. Functionally, this provides the capability for extending the delay in terminal differentiation even beyond the time of target contact, if such a delay would be beneficial in a given cellular context. The *seven up* (*COUP-TF I/II*) gene plays a conserved role in lineage specification during nervous system development (Jacob et al., 2008), and regulates two transitions of lineage fating in the NB5-6T lineage that generates the Tv4 neuron (Benito-Sipos et al., 2011). *Seven up* is required for the switching of temporal transcription factors Hunchback to Kruppel during neuroblast lineage progression in the *Drosophila* nervous system (Kanai et al., 2005; Mettler et al., 2006), including in the NB5-6T lineage (Benito-Sipos et al., 2011), as well as late postembryonic neuroblast lineage fate switching (Maurange et al., 2008). However, perhaps more directly relevant to our study is the role of *seven up* in differentially specifying Tv1,2,3,4 neuronal fates during lineage progression of NB5-6T (Benito-Sipos et al., 2011). It is initially expressed by all newborn Tv1-4 neurons, but its progressive timed downregulation in the Tv1 and Tv4 neurons is required for their differential specification. The few Tv-neuron clusters that form in *seven up* nulls generate an excess of Tv1 neurons and exhibit lineage confusion between Tv1-3 neurons that also results in loss of the Tv4 neuronal fate. Thus, we would postulate that the progressive loss of Seven up in Tv4 provides a timer to signal the end of its role in the specification process in this lineage, and, by virtue of it additionally repressing FMRFa activation, this downregulation is further utilized to de-repress the competence of the FMRFa gene to respond to target-derived BMP signaling and permit terminal differentiation.

Thus, we propose that Seven up plays a novel switching function to de-repress terminal differentiation in addition to its switching roles in lineage specification. This is supported by the additional role for Seven up in repressing expression of the pro-neurosecretory *dimmed* transcription factor in Tv4 neurons (Benito-Sipos et al., 2011). In this way, downregulation of Seven up in Tv4 temporally permits differentiation of the neurosecretory phenotype in Tv4 neurons as well as activation of the neuropeptide itself.

Detailed functional analysis of the *cis*-regulatory architecture of other target- and BMP-dependent neuronal genes will be required to determine whether the principles learned here for target-dependent gene expression is unique to the *FMRFa* gene, or generalizable to most target/BMP-dependent genes.

3 FUNCTIONAL ATTENUATION OF A BMP-RESPONSE ELEMENT IS REQUIRED FOR NEURON SUBTYPE-SPECIFIC BMP-DEPENDENT GENE REGULATION IN DROSOPHILA NEURONS.

3.1 Synopsis

The terminal differentiation of neurons can require target-derived retrograde bone morphogenetic protein (BMP) signaling. In *Drosophila*, retrograde BMP signaling is required for terminal differentiation of *Drosophila* Tv4 neurons to express the FMRFa neuropeptide. Previously, we identified a 39bp *cis*-element that mediates BMP-dependent *FMRFa* gene transcription. Here, we examine the mechanism of gene activation from this BMP response element (BMP-RE). *Drosophila* BMP-dependent gene activation typically involves de-repression via BMP-silencing of the default repressor, *brinker*, or coordinated *brk* de-repression with direct activation via binding of the BMP transcriptional effectors Smads, Mad and Medea. We find that the 39bp BMP-RE of *FMRFa* contains a highly conserved sequence motif GGCGCCA(N_4)GTAT that comprises canonical Brinker- and Mad-binding sequences (GGCGCC) and an appropriately-spaced, putative Medea site (GTAT). This motif differs from the previously identified BMP-silencer element (GRCGNC(N_5)GTCT) and BMP-activator element (GGCGCCA(N_4)GNCV) sequences, primarily in the critical C of the Medea binding site (GTAT vs. GCTC/GNCV). Such a C>A switch is predicted to prevent Smad recruitment, thereby making Brinker de-repression the predicted model for *FMRFa* activation. Our genetic analyses herein shows that this BMP-RE is not activated by *brinker* de-repression, but instead requires pMad and Medea for activation. Our biochemical analyses show that pMad and Medea indeed bind this sequence, but that this binding is greatly attenuated by the C>A switch. To test the functional relevance of this attenuated binding *in vivo*, we altered the Medea binding site sequence; to a non-Medea binding form (ACTA), and to the GTCT and GNCV consensus

sequences. We found that all conversions eliminated activity of the *cis*-element in the Tv4 neuron. However, conversion to more optimal GTCT or GNCV sequence generated ectopic expression in other pMad+ neurons. Thus, we have identified a novel BMP activation *cis*-element, that we term BMP-AE2, whose attenuated Smad recruitment (through a single nucleotide switch) is functionally required to prevent ectopic BMP-dependent gene activation and generate highly cell subtype-specific BMP-dependent gene expression.

3.2 Introduction

Neurons receive retrograde target-derived signals that are essential for neural survival, differentiation, morphology, and synaptic function and plasticity (da Silva and Wang, 2011; Deppmann and Ginty, 2006; Hippenmeyer et al., 2004; Marques, 2005; Salinas, 2005). Target-derived BMP signaling is essential for *Drosophila* and vertebrate neural differentiation, synaptic function and plasticity (Allan et al., 2003; Angley et al., 2003; Berke et al., 2013; Chou et al., 2013; DuVal et al., 2014; Eade and Allan, 2009; Guha et al., 2004; Henriquez et al., 2011; Hodge et al., 2007; Kelly et al., 2013; Lopez-Coviella et al., 2005; Majdazari et al., 2013; McCabe et al., 2003; Miguel-Aliaga et al., 2008; Pavelock et al., 2007; Veverytsa and Allan, 2011; Xiao et al., 2013). Retrograde BMP-signaling in *Drosophila* neurons is triggered by the ligand Glass bottom boat (Gbb) (Allan et al., 2003; McCabe et al., 2003) acting at a presynaptic BMP-Receptor (BMP-R) complex of the BMP-RII receptor Wishful thinking (Wit) and the BMP-RI receptors, Thickveins and Saxophone. Kinase activity of BMP-RI receptors phosphorylates the receptor-Smad, Mad (Mad; vertebrate Smad1/5/8). Two phospho-Mad (pMad) proteins then bind the co-Smad, Medea (vertebrate Smad4), to form the Smad complex (Gao and Laughon, 2006; Gao et al., 2005) (**Chapter 1 Figure 2**). The Smad complex

translocates to the nucleus to regulate transcription by binding DNA in a sequence-specific manner at BMP-response elements (BMP-REs). Smad-dependent transcription is the primary mediator of BMP function in fly neurons (Berke et al., 2013).

In *Drosophila*, *cis*-regulatory mechanisms of Smad activity have not been examined in the nervous system but extensive studies in other *Drosophila* tissues and organisms have shown Smad *cis*-regulatory activities to be extremely well-conserved (Ross and Hill, 2008a; Schmierer and Hill, 2007). Studies largely in the *Drosophila* wing imaginal disc identified two BMP-RE sequence types in the *cis*-regulatory regions of numerous fly BMP-regulated genes (Pyrowolakis et al., 2004; Weiss et al., 2010); the BMP-Silencer Element (BMP-SE; GRCGNC(N_5)GTCT; R= G or A) and the BMP-Activation Element (BMP-AE; GGCGCCA(N_4)GNCV; V = A, C, G)(Affolter and Basler, 2007) (**Chapter 1 Figure 2 & Figure 7**). Within these sequences, pMad binds GRCGNC or GGCGCC, and Medea to GTCT or GNCV (Marty et al., 2000; Pyrowolakis et al., 2004). The silencer and activator activities of these *cis*-elements is shaped by differential recruitment of the conserved, co-repressor/activator, Schnurri, (Shn) (Blitz and Cho, 2009; Feng and Derynck, 2005; Pyrowolakis et al., 2004; Ross and Hill, 2008a), and the Dipteran-specific repressor Brinker (Brk), which binds sequences matching the consensus GGCGYY (Y=C,T) (Sivasankaran et al., 2000; Zhang et al., 2001), to repress BMP-activated genes in *Drosophila* (Weiss et al., 2010; Zhang et al., 2001). Notably, the Brinker-binding sequence matches many Mad-binding sequences, such as in the BMP-AE (GGCGCC), and in the case of the *BMP-AE* of *dad*, Brinker and pMad compete for binding to this sequence in order to repress or activate expression, respectively. In fact, in the wing imaginal disc, BMP-dependent de-repression of Brinker activity is the primary activator mechanism for BMP-dependent gene expression.

High-level BMP-signaling represses *brinker* gene expression via multiple *cis*-regulatory BMP-SE's. Recruitment of the co-repressor Shn to these BMP-SE's is mediated by the terminal thymine of the Medea-binding site, and is required for *brinker* repression. In regions of low BMP activity, Brk is relieved of this BMP-dependent repression, and it directly represses a battery of BMP-regulated genes by binding *cis*-regulatory GGCGYY sequences (Sivasankaran et al., 2000; Zhang et al., 2001). The central role of *brinker (brk)* as the mediator of BMP-gene regulation in the wing disc is evident in the phenocopy of mutants in the wing BMP ligand *decapentaplegic* (*dpp*) mutants, with that of *brk;dpp* and *brk;Mad* double mutants (Marty et al., 2000; Rogulja and Irvine, 2005; Schwank et al., 2012). Perhaps counter-intuitively, coordinated direct activation by Smads is either not involved (as for *bifid* (Campbell and Tomlinson, 1999; Jazwinska et al., 1999; Winter and Campbell, 2004)), or it merely acts to elevate the magnitude of activated gene expression (as for *spalt major* (Marty et al., 2000)).

The BMP-AE motif was identified in the *dad* locus (*daughters against dpp;* vertebrate Smad6/7) (Weiss et al., 2010), where it mediates graded activation/repression of *dad* by Smad/Brinker binding, respectively. The BMP-AE is bound by pMad or Brinker at *GGCGCC* and by Medea at *GNCV*. Shn is not recruited, due to the lack of a terminal thymine, which allows the BMP-AE to function as an activator (Weiss et al., 2010). Brk and pMad compete for binding to *GGCGCC*, making *dad* expression responsive to relative Brk and pMad activity across their inverse gradients (Kirkpatrick et al., 2001; Rushlow et al., 2001; Weiss et al., 2010). Thus, the precise sequence of the BMP-RE specifies Smad transcriptional activity via differential recruitment of Schnurri and Brinker; making the BMP-RE sequence predictive of function (Pyrowolakis et al., 2004; Weiss et al., 2010).

BMP signaling activates *FMRFa* only in the six Tv4 neurons, out of >400 BMP-activated neurons in the *Drosophila* embryonic and larval nervous system, due to genetic interaction with a Tv4-specific TF code of numerous cell-specific TFs (Allan et al., 2005; Allan et al., 2003; Miguel-Aliaga et al., 2004). In Chapter 2, we described the *cis*-regulatory logic of this integration at a 445bp Tv4-neuron-specific *FMRFa* enhancer. This led to the identification of two *cis*-elements that directly recruit inputs from the LIM-homeodomain TF, Apterous, at a short homeodomain response element (HD-RE), and from BMP signaling at a short 39bp *cis*-element that binds pMad. A canonical pMad-binding site of sequence GGCGCC mediates pMad binding, and specific substitution mutation of this sequence results in an abrogation of Tv4-enhancer activity (**Chapter 2 Figure 10**), thereby phenocopying *wit* mutants (**Chapter 2 Figure 9**). However, the BMP-RE does not have a canonical Medea-binding sequence (**Chapter 2 Figure 8**), leaving unresolved the nature of Smad activity at the BMP-RE in the FMRFa enhancer. We wished to examine how Smads mediate such highly cell-specific and BMP-dependent enhancer activity from this short BMP-RE. To this end, we examined the binding of pMad and Medea to this BMP-RE and explored the roles of Smad, Schnurri and Brinker activity and the specific sequence of the BMP-RE to determine how the FMRFa BMP-RE generates BMP-dependent gene activation.

3.3 Materials & Methods

3.3.1 Fly Genetics

Strains used: Med^{C246} (McCabe et al., 2004), Med^{13} (Hudson et al., 1998), Med^{DF} (Df(3R)ED6361) (Ryder et al., 2007), wit^{A12} and wit^{B11}(Aberle et al., 2002), brk^{XA} (Campbell and Tomlinson, 1999), shn^{1} (Grieder et al., 1995). Mutants were kept over *CyO,Act-GFP*

TM3,Ser,Act-GFP or *CyO,twiGAL4,UAS-2xEGFP* or *TM3,Sb,Ser,twiGAL4,UAS-2xEGFP*. w^{1118} was used as the control genotype. Flies were maintained at 25°C, 70% humidity**.**

3.3.2 EGFP Reporter Transgene Construction

Empty pThunderbird EGFP vector was generated from Tv-nEYFP (see **Chapter 2**) and from sequence within pHstinger (Barolo et al., 2000). Tv-nEYFP was digested with AscI and SpeI. The multiple cloning site (MCS), HSP70 promoter, EGFP. Tra nuclear localization signal and SV40-polyA sequences from pHstinger (Barolo et al., 2000) were liberated with AscI and SpeI were ligated in to the cut Tv-nEYFP backbone. The Tv4-enhancer was PCR-amplified from Oregon-R with XbaI and EcoRI adaptors, restriction digested and ligated into XbaI/EcoRI digested empty Tv-nEYFP. SOE PCR generated nucleotide substitution and deletion mutants were inserted similarly. Summary of all mutations and concatemerization sequences in Appendix C. Fly transformations were performed by Genetic Services Inc. (Cambridge, MA) and Rainbow Transgenic Flies, Inc. (Camarillo, CA) All transgenic reporters were integrated into *attP2* (Groth et al., 2004).

3.3.3 Immunochemistry

Standard protocols were used throughout (Eade and Allan, 2009). *Primary antibodies*: Rabbit α-FMRFa c-terminal peptide (1:1000) (Baumgardt et al., 2007); Rabbit α-pMad (1:100, 41D10, Cell Signaling Technology); Mouse α-Eya (1:100; clone 10H6); mouse α-Dac (1:50; Dac2-3) (DSHB; Iowa U., Iowa). *Secondary antibodies*: Donkey α-Mouse, α-Rabbit, conjugated Cy5 (1:100, Jackson ImmunoResearch).

3.3.4 Gel Shift Assay

Mad, Medea, Tkv^{QD}, encoding expression plasmids and pAc5.1/V5-His plasmid were transfected to 3 x 10^6 S2 cells in a six well plate. After 48 hours, cells were harvested and lysed for gel shift assay. Oligonucleotides (BMP-AE2) derived from TV neuron enhancer region, AE-like AE2 and SE-like AE2 were synthesized and labeled with IRDye 700 by Integrated DNA Technologies (IDT). DNA and protein binding was performed by incubating 20 μg of lysate protein with 1 μl of 50nM IRDye 700-labeled probe in a 20 μl reaction buffer containing 25mM Tris pH7.5, 35mM KCl, 80mM NaCl, 3.5mM DTT, 5mM $MgCl_2$, 0.25% Tween 20, 1 μg poly dIdC, 10% glycerol and 1x protease inhibitor cocktail for 30 min at room temperature. If super shifts were performed, then 1 μg of antibody or mouse IgG was added and incubated for another 30 min at room temperature. For competition assay, a 1-1000 fold stoichiometric excess of unlabeled DNA fragments or mutant DNA fragment were incubated with the labeled probes. Competitor sequences are as follows, mutations are underlined.

Wild-type competitor BMP-AE2 (GCCAGAGGCGCCACAATGTATCCCGTT).

AE-like AE2 (GCCAGAGGCGCCACAATGTCGCCCGTT).

SE-like AE2 (GCCAGAGGCGCCACAATGTCTCCCGTT).

Mad site mutated BMP-AE2 (GCCAGATAGTAGACAATGTATCCCGTT).

Medea site mutated BMP-AE2 (GCCAGAGGCGCCACAATACTACCCGTT).

The DNA-protein complexes were resolved on a 4% non-denaturing polyacrylamide gel for 1.5 hours at 70 V in 1 × TGE buffer. After electrophoresis, the gel was imaged immediately by Licor Odyssey imager system.

3.3.4.1 Image and Statistical Analysis

More than five animals were examined for every genotype. Analysis on the 445 bp EYFP enhancers was performed on homozygous reporter lines; all other reporter lines were analyzed as heterozygotes. Images were acquired with an Olympus FV1000 confocal microscope with settings that avoided pixel intensity saturation. Representative images of Tv-neurons being compared in figures were contrast enhanced together in Adobe Photoshop CS5 (Adobe Systems, Mountain View, CA). All statistical analysis and graphing were performed using Prism 5 (GraphPad Software, San Diego, CA). All multiple comparisons were done with One-Way ANOVA and a Tukey *post-hoc* test or Student's two-tailed *t*-test when only two groups were present. Differences between groups were considered statistically significant when $p<0.05$. Data presented as mean ± SEM.

3.4 Results

Loss of function mutants in *gbb* and *wit* abrogated FMRFa expression at the transcriptional level (Allan et al., 2003; Eade and Allan, 2009). Also, we found that overexpressing Mad^{l} (a dominant *Mad* allele with a mutation that selectively disrupts DNA-binding) abrogates FMRFa transcription (Eade and Allan, 2009), consistent with the *Mad* null phenotype (Marques et al., 2003). Further, we showed that recombinant Mad (N-terminal DNA-binding MH1 domain) sequence-specifically binds to the GGCGCC sequence of the *FMRFa* BMP-RE (Section 2.4.4). Here, we wished to determine the nature of Smad complex binding at the BMP-RE.

3.4.1 A Putative pMad/Medea Motif in the FMRFa BMP-RE.

First, we examined the BMP-RE sequence for possible Medea-binding sites. No canonical Medea-binding sequences (GTCT or GNCV) exist in the *FMRFa* BMP-RE. However, the sequence GTAT in the *FMRFa* BMP-RE is positioned with the same spacing as is GTCT or

GNCV, relative to the pMad binding site, in the BMP-SE and BMP-AE motifs, respectively; compare GGCGCCA(N_4)GTAT to GRCGCC(N_5)GTCT and GGCGCCA(N_4)GNCV (**Figure 14C**) (Pyrowolakis et al., 2004; Weiss et al., 2010). This GTAT sequence is not simply an anomaly in laboratory strains of *D. melanogaster*; the *FMRFa* BMP-RE motif is extremely well conserved throughout *Drosophila* species and different *D. melanogaster* strains, in both putative Mad and Medea (GTAT) sequences, and in their relative spacing (**Figure 14B**). Thus, we hereafter refer to the *FMRFa* BMP-RE motif as a BMP-AE2 motif, to reflect the subtle sequence modification of the Medea-binding site to a GTAT in this activating BMP-RE.

Intriguingly, previous studies would suggest that the *FMRFa* BMP-RE motif appears to be unsuitable for pMad/Medea binding. These studies show that the C at position 3 of the Medea motif is critical for Smad binding and enhancer activity. For example, the GTCT of BMP-SE has an absolute requirement for a C at position 3; indeed, an A at that position (to make the *FMRFa* GTAT motif) uncoupled Mad, Medea and Schnurri binding to the motif in *vitro,* and abrogated enhancer activity *in vivo* (Pyrowolakis et al., 2004). Also counter-intuitive to the fact that the *FMRFa* BMP-RE is an activator sequence is the T in position 4. This plays a critical binary role with regards Shn co-repressor recruitment to the GTCT of BMP-SE, and its non-recruitment to the GNCV of the BMP-AE is required for its activator function (Weiss et al., 2010). The apparent lack of an appropriate Smad-binding motif is particularly intriguing in light of the perfect consensus Brinker binding site of the BMP-RE, GGCGCC, that we have to this point attributed to Mad binding (Sivasankaran et al., 2000; Zhang et al., 2001). Thus, looking at the FMRFa BMP-RE sequence in the light of the literature, its sequence would predict that the sequence mediates activation via *brinker*-dependent de-repression.

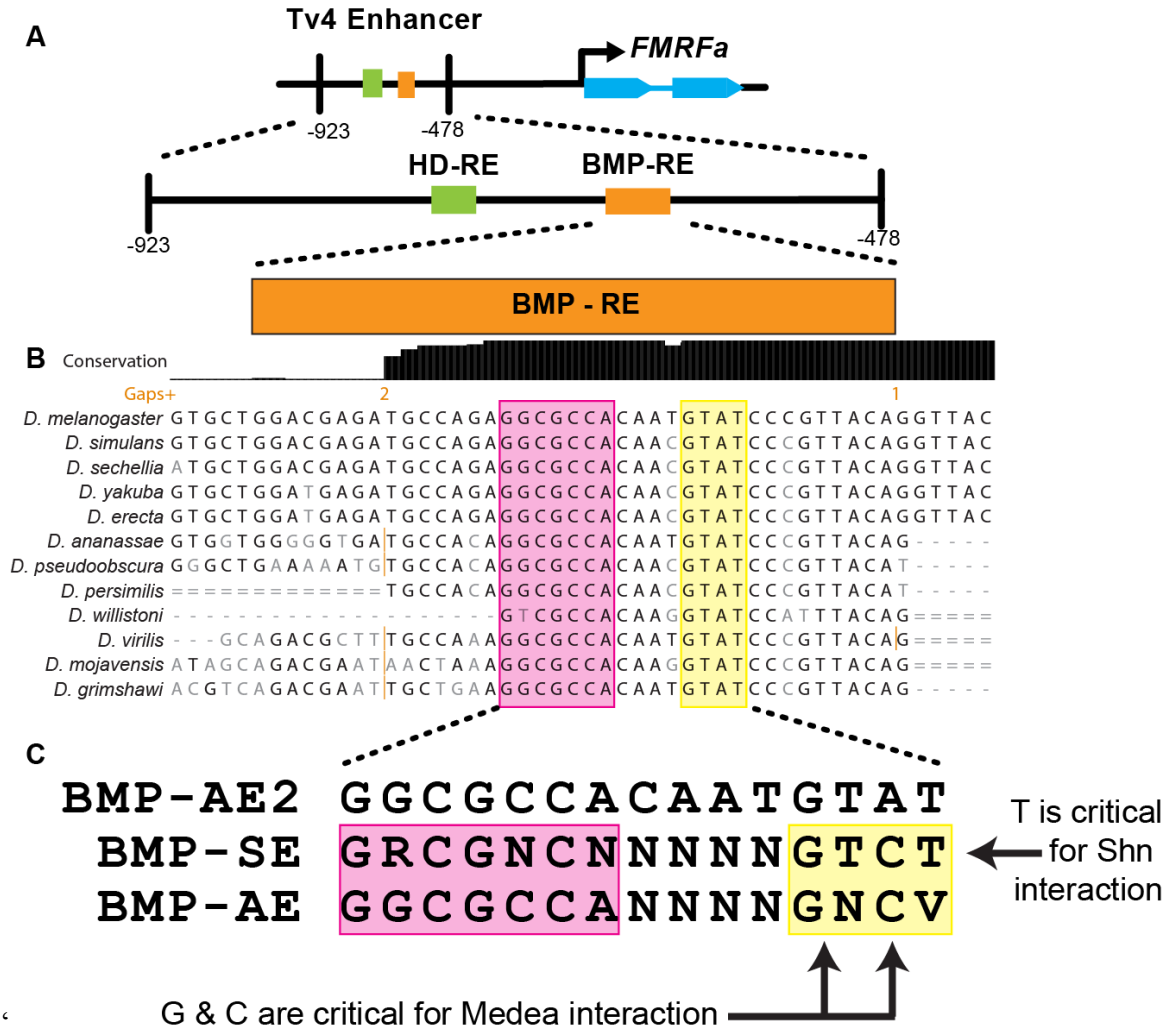


Figure 14. A novel Smad recruitment sequence (BMP-RE) in the Tv4 neuron specific enhancer of the FMRFa gene.

The Tv4 enhancer depicted in (**A**) contains two *cis*-elements critical for FMRFa activation, the homeodomain response element (HD-RE which recruits Apterous), and the BMP response element (BMP-RE) which mediates the BMP-dependence of the Tv4-enhancer and binds recombinant Mad sequence-specifically to its GGCGCC motif (Chapter 2). (**B**) Output from the UCSC Browser shows sequence conservation through the entire BMP-RE across all sequenced *Drosophila* species. Highlighted in magenta is the extremely well conserved putative Mad/Brinker binding site and in yellow is an absolutely conserved putative Medea-binding site. (**C**) Comparison of the putative Smad recruitment site of the FMRFa BMP-RE (BMP-AE2) with the previously identified BMP-SE and BMP-AE. The BMP-AE2 and BMP-AE Mad/Brinker binding sites are exact matches. The spacing between Mad and Medea binding sites is absolutely conserved between all three BMP-REs. In contrast, the Medea site is 'perfectly wrong' in two regards: First, the C>A conversion is predicted to uncouple Smad binding to this element, and a C in this position is critical for BMP-SE and BMP-AE function. Second, the terminal thymine should recruit the co-repressor, Schnurri, in spite of the fact that the BMP-AE2 is an activator element.

3.4.2 *Brinker* and *Schnurri* Are Not FMRFa BMP-RE Regulators

These factors led us to two rationalized mechanisms of *FMRFa* BMP-RE function (Affolter and Basler, 2007). First, the BMP-RE is incapable of recruiting pMad/Medea for direct FMRFa activation; instead, the GGCGCC sequence acts as a Brinker recruitment site. Second, if the pMad/Medea complex is indeed recruited, the additional recruitment of Shn, acting as a co-activator, would be required for FMRFa expression. We tested these models. First, we examined FMRFa and Tv4^{WT}-EYFP expression in *brinker* hemizygotes (brk^{XA}/Y). Even by late L3 larvae, we failed to detect any enhanced anti-FMRFa or $Tv4^{WT}$-*EYFP* reporter activity (**Figure 15 A&B**). However, this may not be unexpected if BMP signaling had entirely eliminated *brk* expression in Tv4 neurons. Thus, we directly tested whether BMP-signaling activates *FMRFa* expression via BMP-dependent repression of *brk,* in *brk/Y ;* wit^{A12}/wit^{B11} double mutants. Unexpectedly, we found that FMRFa expression was not activated in this genotype (**Figure 15 A-E**). Thus, BMP signaling does not activate FMRFa indirectly via de-repression of *brk*. Next, we tested a co-activator role for *shn* in FMRFa activation. We found that FMRFa expression was activated normally in shn^{1} nulls at late Stg 17 embryos (**Figure 15 F&G**). Thus, FMRFa does not require *shn* for activation. We conclude that neither Brk nor Shn play a role in the BMP-dependence of FMRFa, making this gene's BMP-dependent gene regulation distinct from mechanisms found outside the central nervous system.

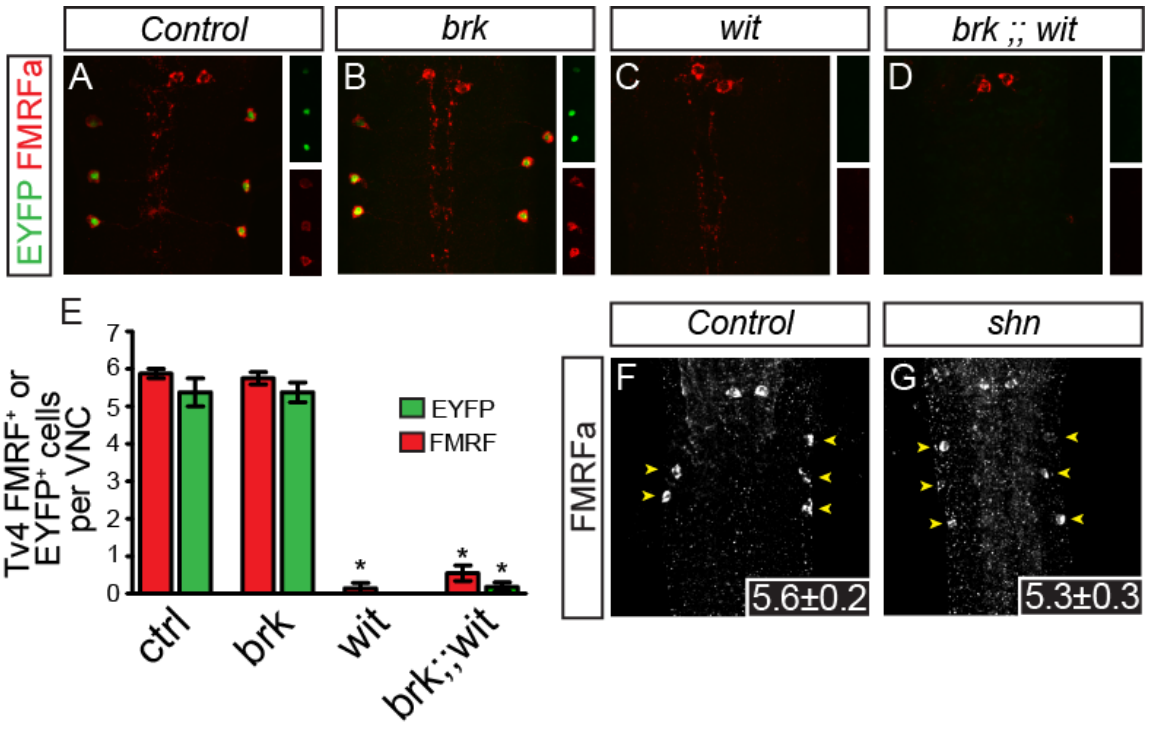


Figure 15. *brinker* and *schnurri* are not required for BMP-dependent FMRFa expression. ***(A,B)*** Expression of BMP-dependent FMRFa peptide and the Tv4-reporter are not affected in *brinker* mutants. **(C,D)** Loss of FMRFa peptide and Tv4-reporter expression in BMP signaling-deficient *wit* mutant is not rescued by loss of *brinker* function in the double mutant. **(E)** Quantification of data in A-D (n=7-12 per group, * $p<0.01$ One-way ANOVA with Tukey HSD post-hoc). **(F,G)** No change in the number of FMRFa expressing neurons is observed between control and *shn* mutants in late stage embryos. Arrow heads point to Tv neurons. (Average number of FMRF^{+} Tv neurons per VNC n=5 per group). ***Genotypes***: *Tv^{WT}-nEYFP. brk (brk^{XA}/Y;;Tv^{WT}-nEYFP/Tv^{WT}-nEYFP).* ***wit*** *(Tv^{WT}-nEYFP,wit^{A12}/Tv^{WT}-nEYFP,wit^{B11}).* ***brk;;wit*** *((brk^{XA}/Y;; Tv^{WT}-nEYFP,wit^{A12}/Tv^{WT}-nEYFP,wit^{B11}).* ***shn*** *(w;shn^{1}/+ vs. shn^{1}/shn^{1}).*

3.4.3 Medea Is Required for FMRFa Expression

These data led us to hypothesize that the *FMRFa* BMP-RE directly recruits Mad and possibly Medea for gene transactivation. Thus, we wished to resolve the nature of the Smad complex recruited by this non-canonical BMP-RE sequence. First, we tested whether Medea is even required for *FMRFa* expression, which surprisingly had not been performed previously. In two null *Medea* backgrounds, immunoreactivity to the FMRFa prepropeptide was entirely lost (**Figure 16A-C**; n=12-15 VNCs for each genotype). We also examined EYFP reporter expression driven from the full length Tv4-enhancer (Tv^{WT}-*EYFP*), and the concatenated BMP-RE and HD-RE Tv4-enhancer fragments (see **Chapter 2**), in *Medea* nulls. As expected, Tv^{WT}-*EYFP* and *BMP-RE-EYFP* expression were eliminated, while the BMP-insensitive *HD-RE-EYFP* exhibited expression comparable to controls (**Figure 16**; n=9-15 VNCs for each genotype). We conclude that Medea is required for BMP-dependent activity of the *FMRFa* BMP-RE.

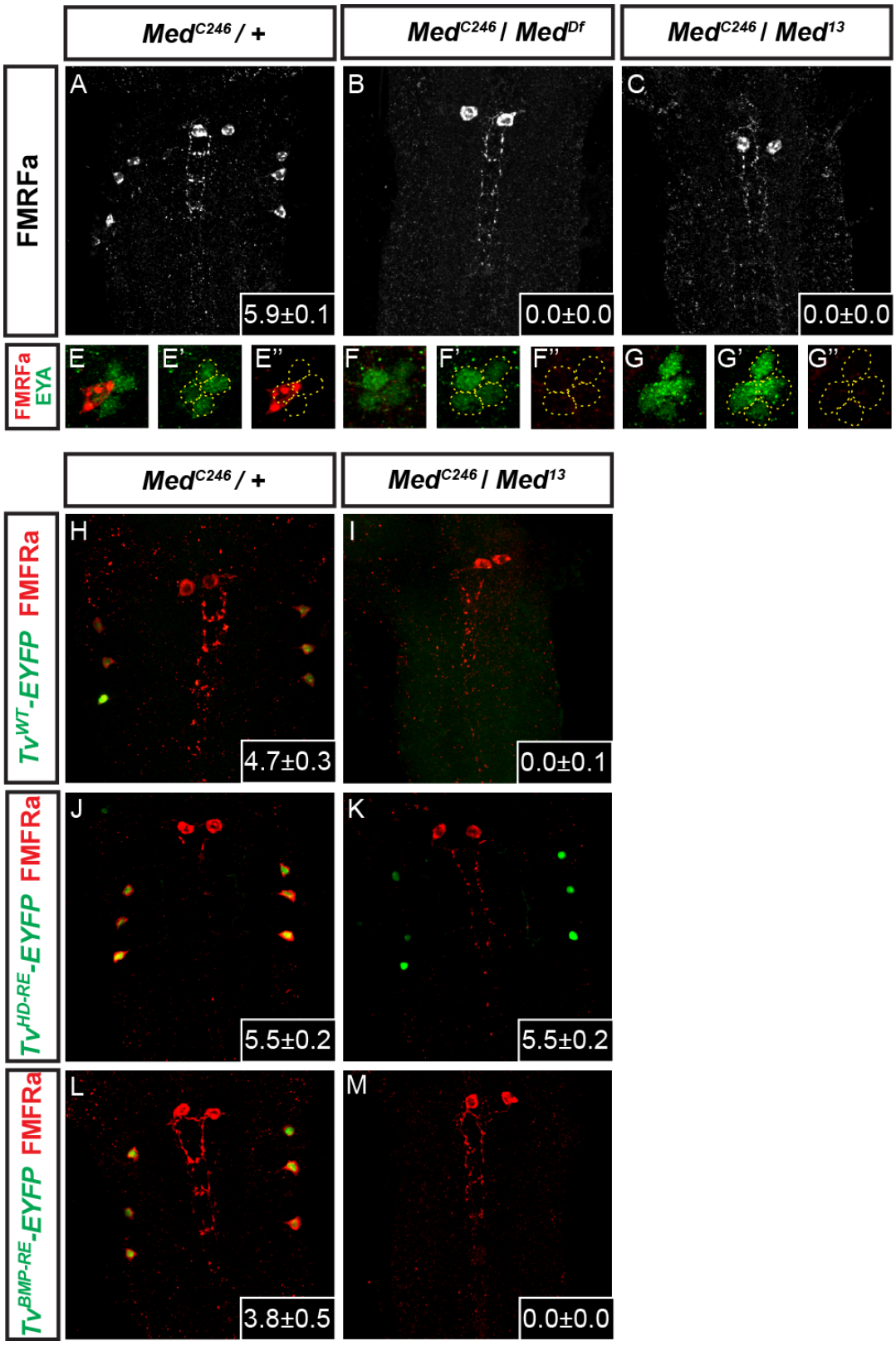


Figure 16. *Medea* is required for the BMP-dependence of FMRFa and its BMP-RE.

(**A-C**) FMRFa immunoreactivity in Tv4-neurons is lost in Medea mutants (**B,C**) compared to controls (**A**). The number of FMRFa-expressing Tv4 neurons are shown in the inset. Med^{C246} and Med^{13} are genetic nulls and Med^{Df} (*Df(3R)6361*) is a molecularly defined deletion of *Medea*. (**E-G**) The small panels beneath each large panel are fluorophore splits from the same genotype but with the VNC's imaged from the side for easier visualization of the Tv-cluster. Anti-Eya (Eyes absent) identifies the Tv neurons with high specificity (circled), to show that they are not absent in Medea mutants. (**H-M**) Panels showing the expression of the Tv4-EYFP reporter (**H,I**), the HD-RE-EYFP reporter **(J,K**) and the BMP-RE-EYFP reporter (**L,M**), in control (**H,J,L**) and in *Med* mutants (**I,K,M**) (n=10-20 animals per group). These data show that the Tv4- and BMP-RE reporters require *Medea* for expression, as predicted by their requirement for BMP signaling. Moreover, they show that the HD-RE reporter is insensitive to *Medea*, as predicted by the lack of BMP-dependence of this reporter.

3.4.4 The GTAT Sequence of the FMRFa BMP-RE Mediates Attenuated Smad Binding *In Vitro*

Based on previous data, the GTAT sequence of the *FMRFa* BMP-RE should uncouple a pMad/Medea/Schnurri complex (Pyrowolakis et al., 2004), and hence should not function as a BMP-RE. As this contrasts with the BMP-dependent activation generated by this BMP-RE previously demonstrated (See **Chapter 2**); we therefore sought to characterize the ability of the FMRFa BMP-AE2 sequence to recruit Smad proteins, in comparison to the well-characterized BMP-AE and BMP-SE motifs. To explore how differences in sequences between these BMP-RE's affects Smad recruitment, we transfected S2 cells with activated BMP signaling components and used total cell lysates to perform Electrophoretic mobility shift assay (EMSA) experiments on infrared dye labeled oligonucleotides containing different BMP-RE sequences. To activate BMP-signaling in S2 cells, we transfected S2 cells with constitutively-activated Thickveins type I receptor (Tkv^{QD}), and full length epitope tagged FLAG::Mad and Myc::Medea (driven from the pAc5.1 vector). **Figure 17** (Lane 8) shows that co-transfection of all three plasmids resulted in a band shift of the *FMRFa* BMP-AE2 probe by the S2 cell lysate. Lanes 2-7 show that if any one of the components of activated BMP signaling isn't co-transfected, band-shift of the FMRFa BMP-AE2 sequence does not occur (**Figure 17**). Addition of mouse anti-Myc or mouse anti-FLAG IgG to the lysate resulted in a supershifted band, whereas the control mouse IgG did not have any effect on retardation of the labeled FMRFa BMP-AE2 oligonucleotide. Thus, our data show that Medea is required for pMad recruitment to the FMRFa BMP-RE *in vitro* which parallels a requirement for *Medea* in *FMRFa* expression *in vivo*. While the supershift data shows that Mad and Medea proteins are associated with the shifted probe. Further experimentation is required, however, to demonstrate that both Mad and Medea are present at the same time in the same complex.

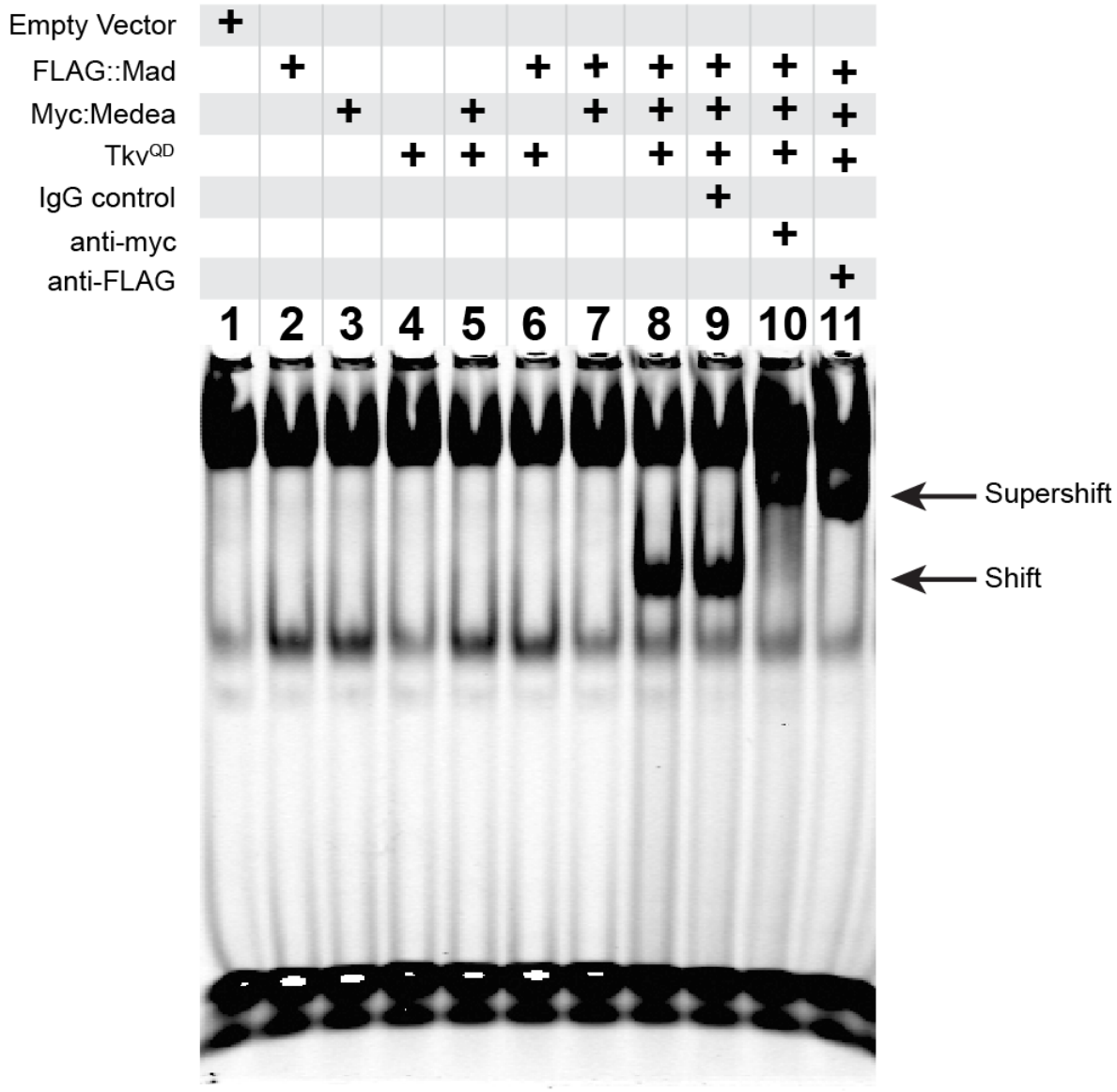


Figure 17. Mad and Medea are required to bind the FMRFa BMP-RE.

EMSA using the FMRFa BMP-AE2 DNA oligonucleotides labeled with IRDye700 subjected to binding with extracts from S2 cells transfected with FLAG::Mad, myc::Medea, and/or activated BMP-receptor Tkv^{QD}. Lanes 2-6 show no specific band shift that is different from empty-vector only transfected cells (Lane 1). Co-transfection with FLAG::Mad, Myc::Medea, and Tkv^{QD} generates a band shift (Lane 8). Addition of antibodies against Myc-tag (Lane 10) or FLAG-tag (Lane 11) causes a super shifted band compared to IgG control. These data show that Mad, Medea, and activated Tkv receptor are needed to generate a band-shift of the FMRFa BMP-AE2. Supershifts indicate that both the FLAG-tagged Mad and the Myc-tagged Medea interact with the FMRFa BMP-AE2 complex that retards labeled oligonucleotide migration.

Next, we tested the requirement for the putative pMad and Medea binding sequences in the FMRFa BMP-AE2. We collected lysates from BMP-activated S2 cells and mixed them with IRDye-labeled wild type FMRFa BMP-AE2 probe. This resulted in a robust and specific band shift **(Figure 18** Lane 2). We then tested the ability of 'cold' unlabeled *FMRFa* BMP-RE to compete with this 'hot' labeled probe binding, at 10x, 100x and 1000x stoichiometric excess. As expected, a cold wild type sequence competed successfully to greatly reduce the bandshift at 10x and eliminate it at higher concentrations. In contrast, a cold probe containing mutated pMad-binding site (GGCGCC > TAGTAG) only partially reduced the bandshift of the hot probe at 1000x (**Figure 18** Lane 8). Thus, this sequence is required for pMad/Medea recruitment. A cold probe containing a mutant Medea-binding sequence (GTAT > ACTA) was a strong competitor for the hot probe, eliminating the band at 100x and greater. However, comparison of lanes 2 and 9 (10x competition of the wild type and Medea-mutant probes) shows that the Medea mutant is a weaker competitor than the wild type probe. Finally, a cold probe that was double mutant for both Mad and Medea sites was a very weak competitor, only showing a reduced hot probe bandshift at 100x concentration. These data indicate that the GGCGCC site is critical for pMad/Medea recruitment but that the GTAT site only contributes very weakly to this.

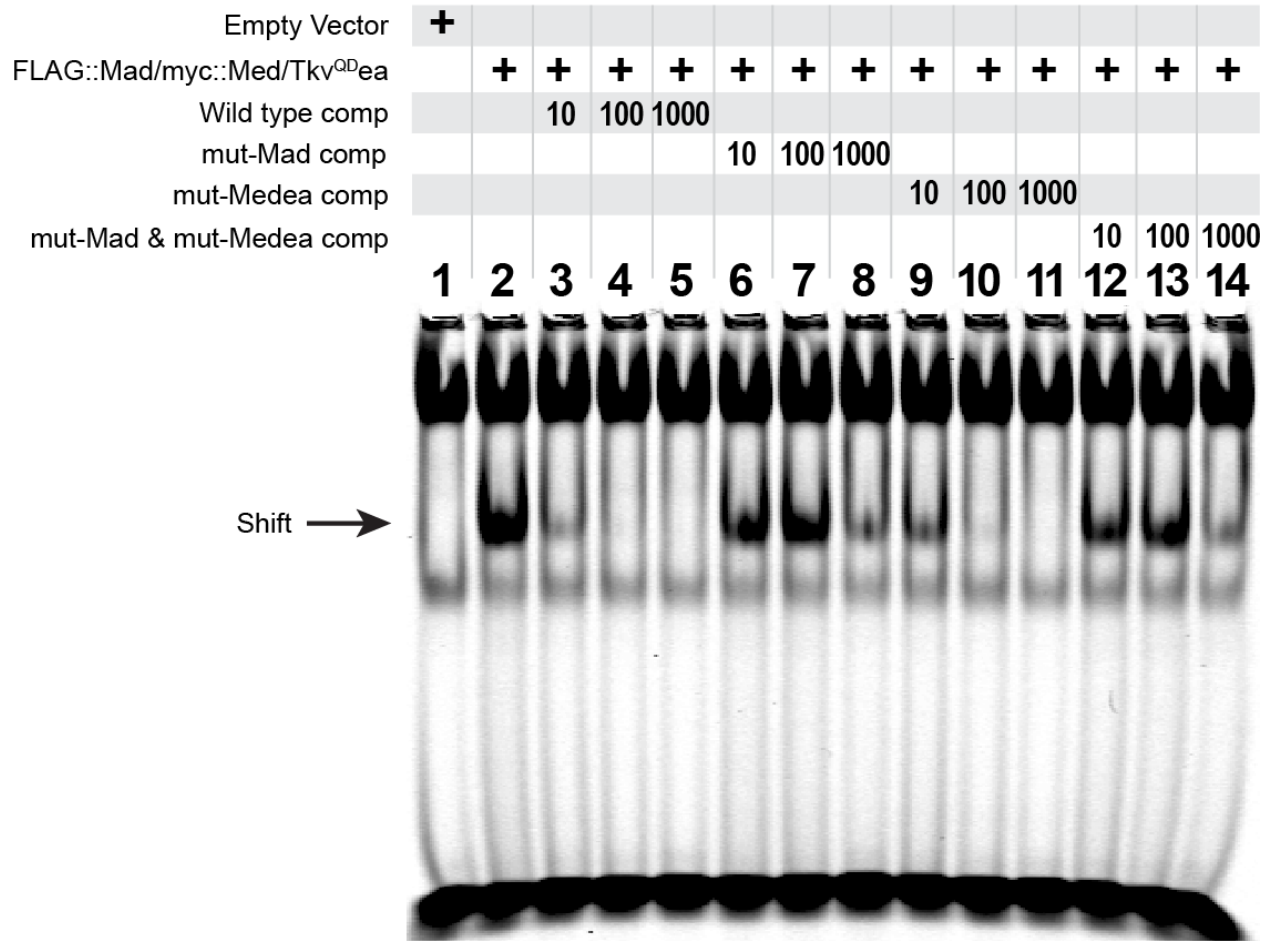


Figure 18. Mad and Medea sites are necessary for efficient Smad recruitment of the FMRFa BMP-AE2.

EMSA using FMRFa BMP-AE2 DNA oligonucleotides labeled with IRDye700 subjected to binding with extracts from S2 cells transfected with FLAG::Mad, myc::Medea, and activated BMP-receptor Tkv^{QD}. Numbers indicate the fold stoichiometric excess of cold competitor. (**Lanes 2-5**) Wild type FMRFa BMP-AE2 is efficiently out competed by wild-type cold competitor at 100-fold or more stoichiometric excess. (**Lanes 6-8**) By comparison, cold FMRFa BMP-AE2 with a mutated Mad site does not efficiently out compete at even 1000-fold excess. (**Lanes 9-11**) Mutated Medea site cold competitor shows a mild reduction in its ability to out compete the labeled probe. (**Lanes 12-14**) Mutation of both the Mad and Medea site strongly decrease the ability of the cold-probe to compete. These data show that the Mad site is more important than the Medea site in recruiting Smads to the FMRFa BMP-RE complex. However, both play an important role.

Next, we wished to test the relative Smad recruiting activity provided by the BMP-AE2 Medea motif in comparison to the Medea-binding sequences CTCG from the BMP-AE (specifically the *dad* AE (Weiss et al., 2010)) and GTCT from the BMP-SE. To this end, we generated a series of hot and cold probes comprising the FMRFa BMP-AE2 sequence either wild type for the GTAT sequence or altered at this sequence to CTCG or GTCT to be AE-like or SE-like, respectively. In **Figure 19**, we show how cold versions of these probes (at 1x, 5x and 10x stoichiometric ratios in relation to hot probe) compete with hot wild type FMRFa BMP-AE2 sequences. As seen in **Figure 18**, the wild type cold BMP-AE2 sequence almost completely outcompeted its hot counterpart robustly by 10x. In contrast, both the AE-like and the SE-like sequence mutants were stronger competitors, completely outcompeting the AE2 hot probe by 5x and greatly reducing the bandshift, even at equimolar concentrations with the hot probe. These data strongly suggest that the GTAT sequence displays attenuated Smad recruitment activity in relation to these other sequence motifs.

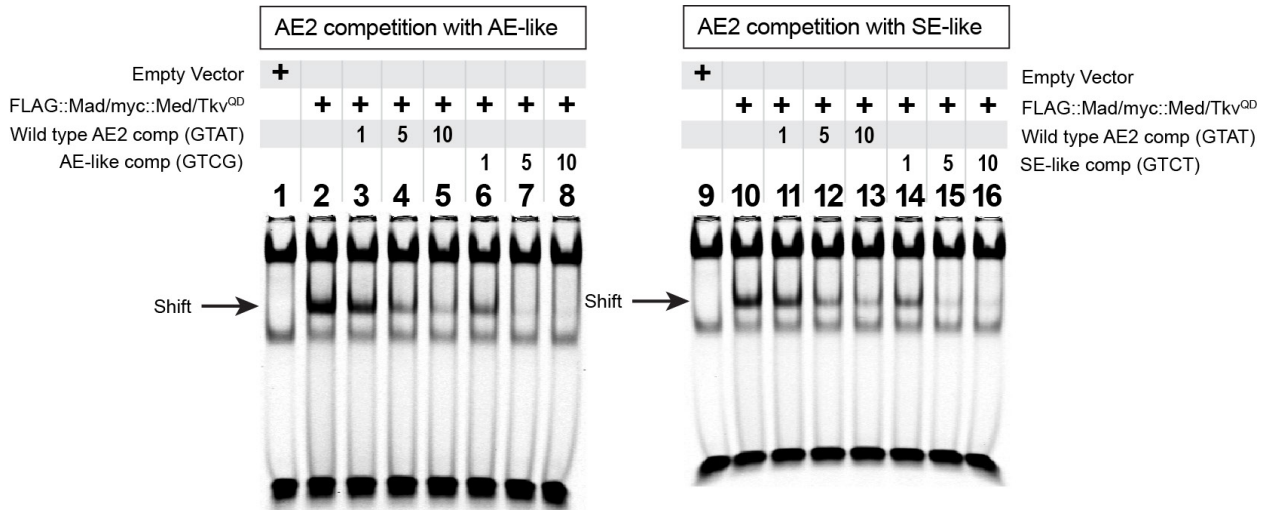


Figure 19. BMP-AE-like and BMP-SE-like sequences recruit Smads more efficiently than the FMRFa BMP-AE2.

EMSA using FMRFa BMP-RE DNA oligonucleotides labeled with IRDye700 subjected to binding with extracts from S2 cells transfected with FLAG::Mad, myc::Medea, and activated BMP-receptor Tkv^{QD}. Numbers indicate the fold stoichiometric excess of cold competitor. Images are from experiment with the same lysate run at the same time on two separate gels. **(Lanes 2-5 & 10-12**) FMRFa BMP-RE is strongly outcompeted by cold wild-type probe at 10-fold stoichiometric excess. (**Lanes 6-8**) Cold competitor with the FMRFa-AE2 Medea site converted to be AE-like strongly competes by 5-fold excess and totally outcompetes by 10-fold excess. (**Lanes 14-16**) Cold competitor with an SE-like conversion of the Medea site shows weak competition at equimolar amounts and strong competition at 5-fold excess. These data show that both the AE-like and the SE-like conversion of the FMRFa BMP-AE2 Medea site increases the ability of the BMP-RE to recruit Smads.

In a reciprocal experiment, we examined the ability of the wild-type AE2 GTAT sequence to compete with hot probe containing the BMP-AE2 sequence with the GTAT sequence converted to be AE-like GTCG or the SE-like GTCT sequence (**Figure 20**). We predicted that the wild type GTAT sequence would be less capable of competing with the hot GTCG and GTCT mutants. In confirmation, we found that the cold wild type AE2 sequence was less efficient at competing with mutated (>AE-like or >SE-like) hot probes, compared to cold AE (GTCG) or cold SE-like (GTCT) sequences.

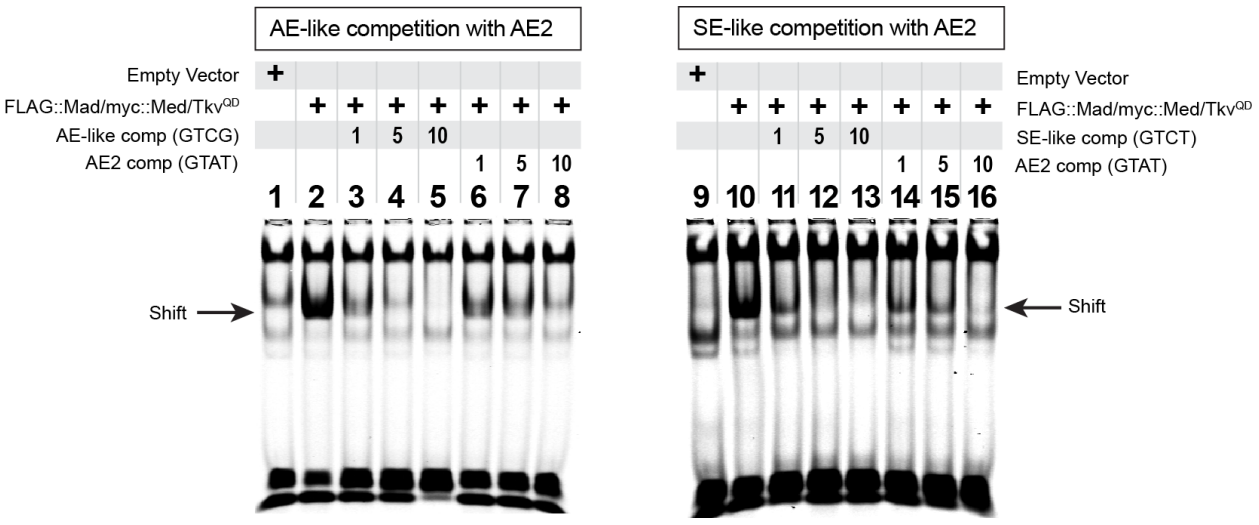


Figure 20. BMP-AE and BMP-SE sequences recruit Smads more efficiently than the FMRFa BMP-AE2.

EMSA using FMRFa BMP-RE DNA oligonucleotides labeled with IRDye700 subjected to binding with extracts from S2 cells transfected with FLAG::Mad, myc::Medea, and activated BMP-receptor Tkv^{QD}. Numbers indicate the fold stoichiometric excess of cold competitor. Images are from experiment with the same lysate run at the same time on two separate gels. (**Lanes 2-5**) FMRFa BMP-RE with an AE-like Medea site is strongly outcompeted by cold probe of the same sequence at 5-fold or more stoichiometric ratios. (**Lanes 6-8**) Cold competitor with the wild type AE2-like Medea site fails to strongly compete at 5-fold excess but does by 10-fold excess. (**Lanes 10-13**) FMRFa BMP-RE with an AE-like Medea site is strongly outcompeted by cold probe of the same sequence at 5-fold or more stoichiometric ratios. (**Lanes 14-16**) Cold competitor with the wild type AE2-like Medea site fails to strongly compete at 5-fold excess but does out compete by 10-fold excess. These data show that both the AE-like and the SE-like conversion of the FMRFa BMP-RE Medea site increase the ability of the BMP-RE to recruit Smads.

3.4.5 Functional Consequences of BMP-AE2 Medea-Binding Site Mutation *In Vivo*

From our studies to this point, we concluded that the AE2-like GTAT sequence attenuates Smad-recruitment in comparison to that observed with the AE-like and SE-like sequences. This raises the question of why this motif is attenuated in this subtle manner. Thus, we tested if an attenuated Medea-binding sequence in the *FMRFa* BMP-RE is functionally necessary for normal BMP-RE activity. We first examined whether the putative Medea-binding sequence was even necessary for activity in Tv4-neurons. To do this, we mutated the GTAT sequence to a sequence that should abrogate all Medea binding, by conversion to ACTA. This was done within the context of the otherwise wild-type full-length 455bp Tv4 enhancer. Notably, this completely abrogated reporter expression in Tv4-neurons and no ectopic expression in other neurons was detected (**Figure 21A&B**). Thus, Medea-binding to this sequence, albeit biochemically attenuated, is essential for activity of the Tv4-enhancer.

Next, we wished to determine if attenuation of the BMP-RE of FMRFa is functionally required for Tv4-specific enhancer activity. We postulated that conversion of the Medea site to a more optimal Medea site, such as by conversion to a characteristic BMP-AE sequence (GNCV) or a BMP-SE sequence (GTCT), would result in ectopic BMP-dependent enhancer activity in the case of the BMP-AE conversion, and a BMP-dependent repression of enhancer activity in the case of the BMP-SE conversion. To test this, within the context of an otherwise wild type Tv4-enhnacer, we substituted the GTAT sequence for GACG (the BMP-AE sequence of the *dad* enhancer) and for GTCT (the canonical BMP-SE sequence of the brk enhancer). Notably, both conversions caused ectopic enhancer activity in other pMad+ neurons of the VNC (**Figure 21F&G**), and eliminated Tv4-enhancer activity in the Tv4 neurons (**Figure 21C&D**). We conclude that the GTAT Medea-binding sequence of the *FMRFa* BMP-AE is necessary for

activation in Tv4 neurons, and that its attenuation of Smad binding at the BMP-AE2 is functionally required to prevent ectopic BMP-dependent *FMRFa* activation in other neurons, in order to generate Tv4-neuron-specific expression.

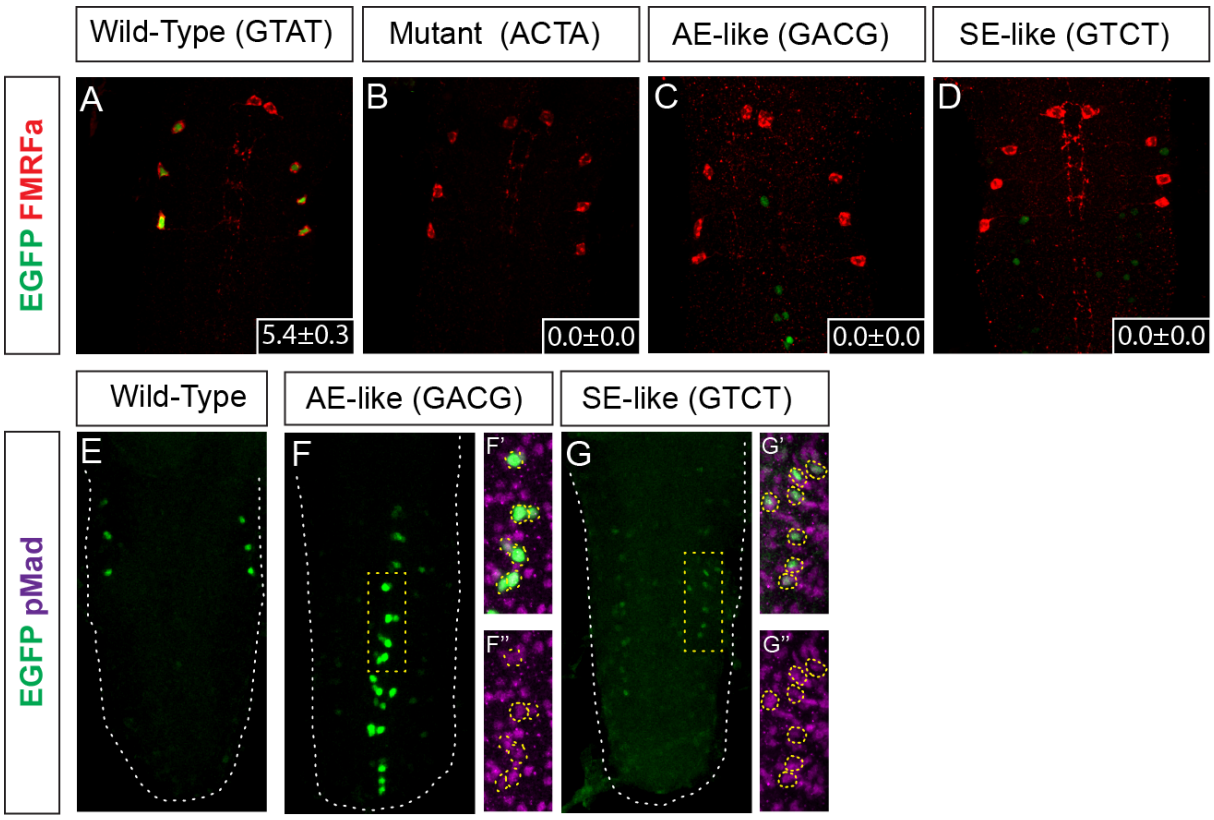


Figure 21. The FMRFa BMP-AE2 Medea site is necessary for the function of the Tv4-neuron specific enhancer.

Conversion of the GTAT site in the wild-type 445bp Tv4 neuron specific FMRFa enhancer (**A**) to a mutant version (ACTA) results in a complete loss of reporter gene expression in the Tv4 neurons **(B**). Conversion to an AE-like (GACG) (**E**) or an SE-like (GTCT) sequence (**D**) also results in a total loss of reporter expression in the Tv4 neurons. The AE-like and SE-like conversions, however, produce expression in ectopic cells outside of the Tv4 neurons (**F&G**). These ectopic cells co-localize with anti-pMad staining. Dotted yellow box indicates region of inset. Numbers in bottom corners of images indicate the mean number of EGFP positive Tv4 neurons per VNC ± SEM out of the possible six Tv neurons (n=5-9 animals per group).

3.4.6 Bioinformatic Identification of AE2-Like Motifs in the *Drosophila* Genome.

We examined whether the FMRFa BMP-AE2 represented the only instance of this atypical BMP response element in the *Drosophila* genome. Using genome-wide sequence detection software (MERMER, http://www.insilicolabs.com/experiment/index.php) we located all genomic instances of the BMP-AE2 motif, using a stringent sequence pattern GGCGCCANNNNGTAT. We found 55 instances located within 20 kb of 78 genes' transcriptional start sites. To find potential non-coding regulatory regions we removed sequences that were in protein coding, 5', and 3' untranslated regions. We then filtered for conservation using phastCONS in UCSC Browser, which compares 12 sequenced *Drosophila* species and determines statistically significantly enriched evolutionary conservation between genomes. Specifically, the "AND" logic function in the UCSC genome browser was used to determine overlap between the BMP-AE2-like sequences with the "Most Conserved" phastCONS output. 27 well conserved BMP-AE2-like sequences were found near 40 genes. To further refine putative BMP-AE2 like sequences that may play a regulatory role in the CNS, we then determined if the nearby genes were known to be expressed in the *Drosophila* CNS, utilizing *Drosophila* modENCODE RNA-seq data from dissected tissues (Boley et al., 2014) and the BDGP *in situ* database (Weiszmann et al., 2009). This filtering retained 19 BMP-AE2-like sequences near 20 CNS-expressed genes. These genes are listed in **Table 1**. Two of these genes, *sog* and *dad,* are involved in BMP signalling. *Dad* is positively regulated by BMP signalling and the location of the putative AE2 element is 5' of the transcriptional start site (the canonical *dad* AE is intronic)(Weiss et al., 2010), while *sog* is an extracellular inhibitor of BMP signalling that interferes with BMP ligand-receptor interaction (Sawala et al., 2012; Yu et al., 1996). *NetB* and *comm* are known to play a role in axon guidance (Mitchell et al., 1996; Tear et al., 1993; Therianos et al., 1995). Neuronal differentiation in a sex specific manner is regulated by *fru*

(Billeter et al., 2006; Castellanos et al., 2013) while *HGTX (Nkx6)* is expressed in subsets of neuronal populations, including BMP-activated motoneurons (Seibert et al., 2009; Uhler et al., 2002). Collectively, the high frequency of nervous system development-related genes, in addition to genes with known BMP pathway roles, gives us some confidence that the BMP-AE2 may be a nervous system mediator of BMP-dependent gene regulation.

Table 1. List of genes near BMP-AE2 sequences that are highly conserved and not in protein coding or untranslated regions of known gene transcripts.

Location (Revision 5 of *D. melanogaster* Genome)	**Nearby CNS Enriched Gene**	**Additional comment on known function or expression**
chr3L:13563966-13563968	*bru-3*	Very strong CNS expression according to *in situ* stages 13-16
chr2R:5793168-5793170	*Fmrf*	Positive control
chr2L:12796327-12796329	*kek1*	Strong CNS expression EGFR antagonist
chr3R:14182794-14182796	*CG15803*	
chr2R:19405205-19405207	*chrw*	Strong CNS expression based on RNA-seq. Rab family GTPase
chr2L:15189634-15189636	*CG4168*	
chr3L:11341529-11341531	*CG6163*	
chr2R:15087211-15087213	*Happyhour*	Protein Kinase, Alcohol hypersensitive behavioural mutant
chr3L:15769864-15769866	*comm*	Axon guidance
chrX:14322048-14322050	*dpr8*	a.k.a. defective proboscis extension response 8
chr3R:14324289-14324291	*fru*	Involved in sexual dimorphic neuron development
chr2L:9319176-9319178	*Ggamma30A*	Involved in multiple aspects of eye and nervous system development
chr3L:14582846-14582848	*HGTX*	Plays key roles in motoneuron development
chr3R:24524509-24524511	*htt*	Huntington ortholog
chrX:14590776-14590778	*NetB*	Role in Axon guidance
chr3R:12878838-12878840	*dad*	Uncharacterized potential BMP-RE 5' of Dad transcriptional start site
chrX:15510844-15510846	*sog*	BMP ligand antagonist
chr2R:20581655-20581657	*ST6Gal/CG33988*	ST6Gal mutants have behavioural defects CG33988 is strongly CNS enriched
chr3R:5662649-5662651	*Teh1*	

3.5 Discussion

Our previous work identified a 39bp BMP-RE within the Tv4-enhancer of *FMRFa* that generated Tv4-specific and BMP-dependent reporter activity, upon 2-4x concatemerization. Within this *cis*-element, we showed that Mad binds with sequence-specificity to the highly conserved GGCGCC sequence. We explored the nature of this *cis*-element's BMP-dependence. We started our analysis by identifying sequences 3' of this Mad site with exact spacing and similar sequence to the Medea binding site of other well characterized BMP-REs. These include the BMP-AE and the BMP-SE. The BMP-AE2 identified here only differs from either BMP-RE in a single, but critical, nucleotide in the Medea-binding sequence. This C>A nucleotide switch predicted that it should not directly recruit Smads but may recruit Brinker and Schnurri. This suggested a de-repression model for BMP-activation via this *cis*-element. In stark contrast, however, our genetic analysis clearly indicates that neither *brinker* nor *schnurri* are involved in activating this BMP-RE, and that *Medea* is absolutely required. These data led us to examine the putative attenuated Medea-binding site more closely. A series of EMSA studies demonstrated that this Medea site could indeed recruit a Smad complex, but with greatly attenuated binding, with respect to the Medea sequences of the BMP-AE or BMP-SE. This was followed by mutant analysis of the Medea site and testing their consequences through *in vivo* transgenic reporter analyses, which demonstrated that the GTAT Medea sequence is indeed essential for Tv4-enhancer activity. While our data strongly suggest that Medea most likely binds to this attenuated motif, we have not formally excluded the possibility that, while Medea is essential for Smad recruitment to BMP-AE2, it is not in fact required to bind the GTAT sequence. In this light, it is intriguing that we show that conversion of the Medea motif into a more optimal binding motif actually abrogates enhancer activity. Why might weaker binding, but not a total abrogation of potential binding, be absolutely required for activity? This raises the possibility

that an unknown TF acts at the Medea site, and attenuated Medea-binding is required for this to occur. Future studies aim to address whether Medea binding is required, by generating *UAS-Medea* transgenes with point mutations specifically in the DNA-binding residues of Medea. We anticipate that such a mutant would not eliminate FMRFa expression if DNA-binding is not required, but would eliminate FMRFa expression if DNA-binding is required.

Having shown that the BMP-AE2 sequence indeed acts as a *bona fide* pMad/Medea binding site *in vitro*, we wished to test if attenuation of this binding was functionally consequential *in vivo*. We 'hyperactivated' the BMP-AE2 motif by converting it into the more optimal Smad recruiter sequences of the BMP-AE and BMP-SE. The generation of ectopic reporter activity in ectopic pMad+ neurons of the VNC verified that the attenuated Smad recruitment does serve to restrict BMP-dependent activity of this *cis*-element to the Tv4 neurons. Thus, we propose that the BMP-AE2 sequence identified here represents a novel BMP-RE type. It is now apparent that a strong global BMP-dependent activator (AE), a strong global BMP-dependent repressor (SE) and a weak cell-specific BMP-dependent activator (AE2) can be generated from subtle nucleotide substitutions within the 4-nucleotide Medea-recruitment site of these response elements. The extreme sequence similarity yet highly diverse functional output of these three identified BMP-REs offers a highly efficient means of generating multiple different responses from very subtle sequence alterations.

BMP signaling in non-neural tissues controls gene expression through Smad transcription factors and the co-regulators Brinker and Schnurri. Perhaps the best-defined model for Smad transcription regulation comes from work on BMP-dependent patterning and growth of the fly wing imaginal disc (Affolter and Basler, 2007). Secreted BMP ligand (Decapentaplegic; Dpp) forms an extracellular morphogen gradient across the tissue (Teleman and Cohen, 2000). This

gradient patterns the wing disc by establishing an inverse gradient of the repressor *brk* (Affolter and Basler, 2007; Muller et al., 2003). The BMP-SE and BMP-AE *cis*-elements are required to 'read' these gradients into precise expression domains of BMP-responsive genes.

In these cases, the *BMP-RE* sequence is highly constrained, beyond the simple needs of pMad and Medea binding, by the additional needs of Brinker or Schnurri recruitment. Thus, examination of the FMRFa BMP-AE2 would have predicted roles for Brinker and Schnurri. However, neither are involved in BMP-activation of FMRFa. With the benefit of hindsight, this is not an unexpected result. The FMRFa gene is activated by a singular target-dependent input, thus the needs of generating a position-appropriate response to a morphogenic gradient of BMP activity may not be required of this BMP-RE. We postulate that this dispenses with the need of computing the relative amount of activating and repressive input, thus the use of *brinker* and *schnurri* are no longer required. This concept would raise similar questions for other neurons, with one caveat; perhaps the ability to compute activating and repressing inputs would provide a means for homeostatic regulation of BMP-RE output. Further investigation of the regulatory mechanisms at the BMP-REs of other neuronal genes will resolve these questions.

Smads can act cooperatively with local TFs to bind *cis*-regulatory sequences and synergistically regulate gene expression (reviewed in (Feng and Derynck, 2005; Ross and Hill, 2008a)). Inherently, interaction with different TFs in each cell subtype can promote diversity in target gene selection. However, TFs that cooperate with Smads in neurons have not been identified and remain essentially unstudied in the literature with specific respect to *Drosophila*. Indeed, even our detailed study of the *FMRFa* BMP-AE2 showed that none of the known sequence-specific transcription factors in the Tv4 neuron serves this function at the FMRFa BMP-AE2 (Chapter 2). Regardless, the BMP-AE2 sequence of FMRFa has conserved flanking

sequences that are highly likely to be the binding sites for these TFs in the Tv4 neuron. We have initiated studies to identify these TFs and to further analyze how they are recruited to the FMRFa BMP-RE and interact with the bound Smads. Of particular interest will be to examine whether cooperative interactions between the Smads and (a subset of) these bound flanking TFs permit the binding of Medea to the attenuated GTAT sequence of the BMP-AE2. In this context, it was intriguing to observe that conversion of the BMP-AE2 into BMP-AE and BMP-SE type motifs abrogated Tv4-enhancer activity in Tv4 neurons. We had postulated that such conversion would have resulted in enhanced expression in Tv4 neurons (as it does in other neurons). Perhaps enhanced Smad recruitment prevents the binding and/or cooperative interaction of these other TFs. However, this awaits future examination, as we anticipate that identification of the TFs that act at the sequences flanking the BMP-AE2 will be required in order to determine why enhancement of Smad recruitment to the FMRFa BMP-RE eliminates enhancer activity.

Does the FMRFa BMP-RE represent a solution to BMP-regulated gene expression in the nervous system? We identified only 27 unique highly conserved sequences in the *Drosophila melanogaster* genome, *in silico,* that conform to the sequence and spacing of the Mad and Medea sites in the FMRFa BMP-AE2. Of these, 19 were found to be near genes enriched in the CNS and a number of these appear to be involved in BMP pathway itself or in cellular processes impacted in motoneurons in *wit* mutants (See Table 1). Future studies will examine whether these represent *bona fide* BMP-response elements in neurons and whether these genes participate in BMP-dependent neuronal processes.

4 CONCLUSION

4.1 Overview of Research Undertaken

This thesis aimed to characterize the integration of cell intrinsic transcription factors with target derived signals that regulates the activation of genes associated with the terminal differentiation of neurons. Specifically I characterized the way in which the Tv4-specific CRM of the *FMRFa* gene integrates retrograde BMP-signalling with the collection of transcription factors present from earlier stages of development. In so doing, I have:

(1) ***Shown that target dependent signalling directly integrates at an enhancer with intrinsic transcriptional codes to drive terminal differentiation of a neuron.*** Target-activated BMP signaling acts via binding of its Smad transcriptional effectors to the Tv4-enhancer, where it acts to combinatorially specify gene expression with intrinsic transcriptional regulators.

(2) ***Demonstrated that highly specific gene expression can be generated from obligatory synergism between two low-activity heterotypic cis-elements that encode the same output from distinct inputs.*** Two *cis*-elements are required for activation but have very different sequences and mediate distinct inputs. However, upon examination of concatemers, they drive the same highly restricted spatial and temporal patterns.

(3) ***Found that target-dependent gene regulation employs an intrinsic repressor and an extrinsic activator that together generate a high fidelity activation time for target-dependent gene expression.*** Svp acts as a repressor of target-dependent gene activation that is epistatic to BMP signalling. Its timed downregulation coincides with target contact to repress target-dependent gene activation prior to target contact. This is a novel concept for the field of target-dependent gene regulation, and the first proposed role for a COUP-TF II transcription factor family member in the timing of target-dependent gene expression in postmitotic neurons.

(4) ***Identified and characterized a novel type of BMP activation element that defies the canonical model.*** The *FMRFa* BMP-RE is highly similar in sequence and relative

spacing to the canonical AE-like or SE-like elements. However, although its putative Medea binding site is critical for its activity, a single nucleotide change in a critical position greatly attenuates its binding to Smads. We postulate that this attenuation is critical to prevent hyperactivation of the element upon Smad binding, and thereby support cell-subtype-specific activity.

4.2 Placing This Thesis in the Context of the Field

4.2.1 A First Model of Target Derived Signalling Integration with Intrinsic Cues to Drive Neuronal Terminal Differentiation

The necessity of target-derived signalling for terminal differentiation of neurons is well characterized, yet no molecular level mechanism has previously been described to account for how target-derived signalling is interpreted by the cells. The work of this thesis represents a rigorous description of target-dependent gene regulation in neurons at the gene regulatory level.

For my studies, I took advantage of the Tv4 neuron and the target-derived BMP-dependence of the neuropeptide gene *FMRFa*. A large number of regulators of FMRFa expression, most of which were transcription factors or transcriptional cofactors, were previously identified through genetic means and a 445 bp enhancer had been defined (Allan et al., 2005; Allan et al., 2003; Benveniste and Taghert, 1999; Benveniste et al., 1998; Massague et al., 2005). I tested the hypothesis that some, or all, of the factors may act directly at this enhancer to regulate FMRFa expression. I have shown the *cis*-regulatory logic and underlying enhancer structure through which these intrinsic and extrinsic regulatory inputs of FMRFa in Tv4 neurons are detected and integrated. This enhancer consists of two essential, separable, conserved *cis*-elements. Each *cis*-element integrates parts of the overall Tv4 combinatorial code. A homeodomain response element (HD-RE) mediates Apterous binding and a BMP-response element mediates Smad binding (BMP-RE). Both *cis*-elements contain very different sequences

and different direct inputs, yet remarkably contain the same spatial and temporal information; however, detection of this necessitated the concatemerization of each *cis*-element. These findings are the first to demonstrate how target-derived signalling is interpreted by neurons to activate terminal differentiation gene expression.

4.2.1.1 Comparison to What Was Known Previously About the Tv4-Enhancer

Previous work in the Taghert lab had identified the Tv4 neuron specific enhancer. Initial deletional analysis by the Taghert lab largely agrees with the deletion series and point mutation constructs that I made and analyzed (**Figure 22**), except for two key points. Their constructs pWF43 and pWF50, which encompassed both the HD-RE and BMP-RE should have been expected to express reporter activity in the Tv4 neuron, but were found to be weak or inactive, respectively. Also, while the Taghert lab observed ectopic expression with numerous deletions and assigned these missing sequences as repressive sequences, we did not detect any repressive sequences in our analysis, as no point mutation or deletion mutant showed consistent ectopic reporter repression or increased reporter activity. These effects may be artifactual due to genomic positional effects of the p-element-based transgenic reporters and/or the strongly amplified LacZ chromogenic staining the Taghert lab used. In contrast, our approach used site-directed transgenesis and a fluorescent protein reporter. See **Figure 22** for comparison of our new model of the Tv4-enhancer to the previous model.

The Taghert and Thomas labs also examined the relative role of the putative Ap-binding site in the Tv4-enhancer by mutational analysis (Benveniste et al., 1998). They compared wild type reporters with reporters mutated at HD-C or HD-A/B/C together. This analysis found that while the HD-C mutant only modestly downregulated reporter expression, mutation of all three HD sites abrogated expression totally. They concluded that each Ap site plays a role in FMRFa

expression but their relative contribution was unclear. Our deletion analysis reported in Chapter 2 resolves this relative contribution of each HD site, and shows that HD-A is essential while HD-B/HD-C provide booster function to ensure appropriate high-level expression.

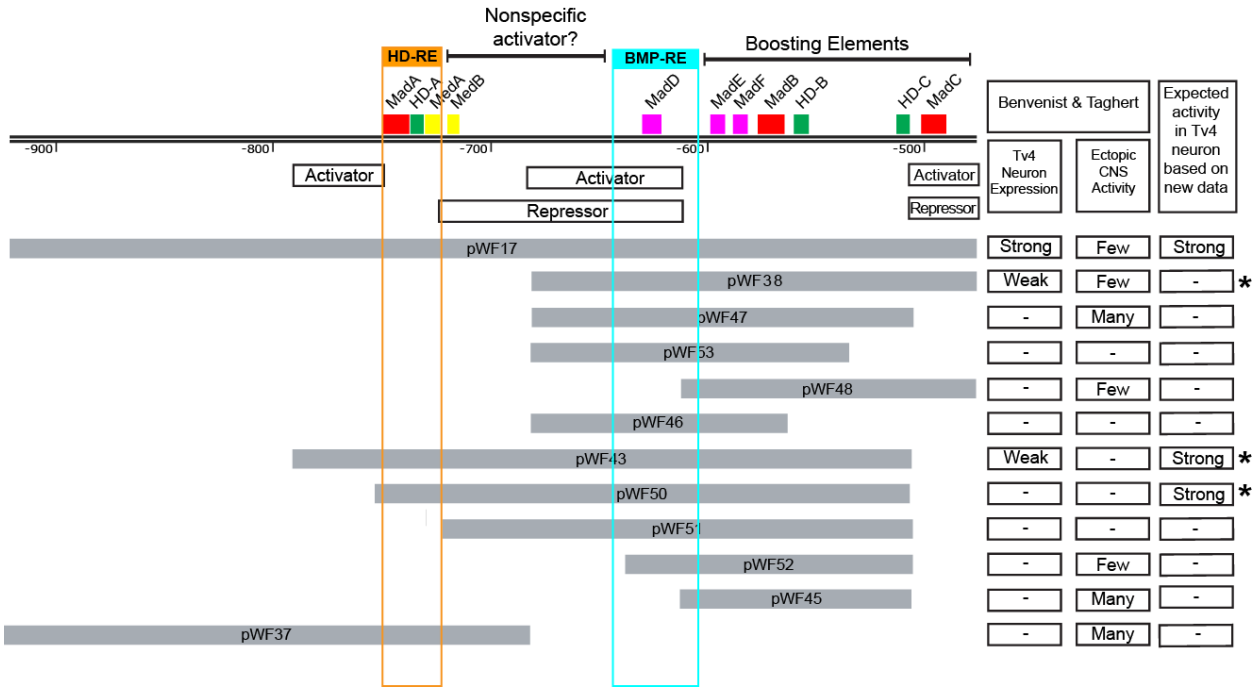


Figure 22. Comparison of previous Tv4 enhancer model with our current model.

Conflicting data between Taghert lab data and our model is marked with an asterisk.

4.2.1.2 Commentary on Requirement of the BMP and Homeodomain Response Elements

My mutational analysis of the FMRFa enhancer reveals that multiple heterotypic weak response *cis*-elements can contribute to robust expression. Two critical *cis*-elements, the HD-RE and BMP-RE are necessary for enhancer function. However, it also appears than there is a generic activator (the intervening sequence between the HD-RE and BMP-RE) as well as the other putative Mad and Apterous binding elements. My deletional and concatemerization analysis demonstrated these 'boosting' elements cannot by themselves drive expression in the Tv4 neurons. The HD-RE and BMP-RE must both be present in the enhancer for enhancer activity; they play linchpin roles. I propose that this system may exist to provide a double-assurance mechanism of transactivation of *FMRFa*. They integrate two sets of inputs that individually may give ectopic expression if allowed to reach *trans*-activation threshold independently. Apterous and BMP signalling occur in many other cell types in the CNS, but they uniquely overlap in the Tv4 neuron. Both essential *cis*-elements must be engaged by their cognate transcription factors before strong transcription of the FMRFa gene is achieved, preventing ectopic expression of FMRFa in time or space. This principle of multiple weak *cis*-elements summating or synergizing multiple transcriptional inputs has been observed for other enhancers, including those that respond to signalling pathways; however, the minimal elements in these cases tend to be multiple "low quality" versions of similar response elements rather than different sequences that interpret totally different components of a transcriptional code (Ramos and Barolo, 2013).

Our finding that concatemerization of the *FMRFa* minimal *cis*-elements generates the same spatiotemporal output as the full Tv4-enhancer could be viewed from positive and negative perspectives. First, the positive perspective; concatemerization allowed us to examine the

information content of the minimal *cis*-elements, as the monomers failed to report any activity. From this perspective, I would recommend concatemerization when exploring other short *cis*-elements. The primary caveat here would be whether the concatemer generated considerable ectopic expression making interpretation of relevant regulatory inputs suspect. This was fortunately not the case for the short *cis*-elements used herein. However, another important consideration would be viewed as the negative perspective. To what extent is the expression observed in the concatemer artifactual, perhaps representing aberrant cooperativity between repeated TFs bound at the repeated TF binding sites? It could be argued that concatemerization of the minimal elements aberrantly stabilises trans-activating complexes or even forces inappropriate cooperativity to a degree that appreciable reporter expression is detected, but is in fact artifactual. For example, it has been demonstrated, with respect to gene regulation in response to morphogen gradients, that cooperative interactions are pervasive and have been shown to lower threshold responses to morphogens, such as Bicoid and Dorsal (Burz et al., 1998; Jiang and Levine, 1993). Even with all these potential problems taken into consideration, it should be noted that 'workhorse' reporter constructs, such as the *UAS* cassette in *Drosophila* (5X UAS response elements to GAL4 protein) (Brand and Perrimon, 1993) or the TOPFLASH reporter (6X TCF/LEF binding sites) (Molenaar et al., 1996) are used widely and regularly by whole communities of researchers. These are faithful reporters of their specific inputs, indicating that concatemerization of cis-elements does not inherently generate ectopic activity.

One potential way to circumvent such issues would be to use better stabilized and stronger reporters, such as by using modified 5' and 3' UTRs in the reporter construct, to increase reporter transcript stability and translation. For example, the addition of the virally derived syn21 5' UTR in otherwise conventional reporter constructs similar to the YFP and GFP

reporter vectors used in this thesis can increase reporter activity 7-fold alone (Pfeiffer et al., 2012). Also, GAL4/UAS can be used to amplify reporter expression, and enzymatic reporters can be used to enhance the level of 'report'.

4.2.1.3 Cell Identity and Minimal Elements

The core concept of cell specialization being driven by combinatorial codes of transcription factors that overlap uniquely in time and space has been well demonstrated (Allan and Thor, 2015; Istrail and Davidson, 2005; Lelli et al., 2012). At the intersection of these overlaps, common factors give rise to unique instances of transcription factor coding in the cell that drive unique gene expression patterns. This transcription factor-centric way of thinking about gene regulation should also include the cognate activities of *cis*-regulatory regions. In the context of *Drosophila* Tv4 neurons, complex regulatory networks of transcription and transcriptional co-factors working over time and space have been extensively characterized. Despite nearly a dozen transcription factors having been identified as necessary for *FMRFa* expression, once the Tv4 neuron is properly specified, it is only the activities of Apterous and the pMad-Medea complex binding the enhancer that are required to turn on *FMRFa* at the right time and place (with the caveat that we do not know all the regulators of *FMRFa*). Further, these activities can be parsed out into Apterous binding the HD-RE and the pMad-Medea complex binding the BMP-RE. From the transcription factor-centric perspective, Apterous and pMad expression only overlap in the Tv4 neuron in the VNC, unlike many of the other *FMRFa*-regulating factors that show overlapping expression patterns elsewhere in the CNS. Collectively, this makes for a neat and tidy model suggestive of *FMRFa* requiring a Boolean logic AND gate coincidence of Ap+pMad to transactivate. However, concatemerized FMRFa HD-RE expresses specifically in the Tv4 neuron. It will express even when one part of the elegant Boolean Tv4-specific overlap model is

removed in BMP signalling mutants. One could make the conjecture that this is due to an additional factor that is specific to the Tv4 neuron that is activating the HD-RE. In Section 4.3.2, I propose an RNAi screen that likely would identify this cofactor.

These observations raise the question of what defines cell identity. If pMad and Apterous no longer uniquely overlap in the Tv4 neuron in a BMP signalling mutant, and *FMRFa* is not activated in this mutant, is the Tv4 neuron in the mutant really the same cell as the Tv4 neuron in a wild type animal? Given that construction of categories of objects or phenomenon is a fundamental activity of science used to help understand the physical world, and FMRFa expression within the Tv cluster is a key discriminating characteristic that lets us categorize the Tv4 neuron as being different from its sibling Tv1-3 neurons, the simplistic answer is "no, the cell is no longer really a Tv4 neuron". Since this marker of cell identity is integral to a known function of the cell (to release the neurotransmitter FMRFamide which regulates muscle contractility (Ormerod et al., 2015) and modulate a number of different behaviours including an escape response (Klose et al., 2010; Lenz et al., 2015)), this loss of a marker is tied to a physiologically relevant activity as well, strengthening the argument that the Tv4 cell identity has been lost in a BMP signalling mutant. However, the BMP signalling-independent expression of the concatemerized HD-RE suggests that there is still some level of cellular identity being conferred to the Tv4 cell in BMP signalling mutants. All known direct regulators of *FMRFa* in the Tv4 cell continue to be expressed in a BMP signalling mutant, and morphologically the cell appears to be unchanged since the axon still projects to the right location. Cell identity could then be argued to lie on a continuum of practically definable characteristics that are contingent on the scale and breadth of the number of molecular markers one is able to observe. This sliding scale of "it-ness" of a cell type is philosophically interesting, but to some extent practically

useless to the experimentalist. Arguing that no useful distinction can be made between two extremes because it is hard to place a definable point on the spectrum where the extremes meet does not productively suggest new experiments or insights as to how new conceptual models can be constructed. Potentially the best approach for scientists therefore may be to discuss cell identity with respect to positive evidence of characteristics (e.g. the presence rather than the absence of a molecular marker, physiological parameter, or a type of morphology). We can more easily talk about what something *is* rather than what it is not. This perspective was integral to my thesis work. The extensive genetic analysis I performed was almost entirely dependent on the loss of a positive marker of identity (i.e. FMRFa or a reporter construct expression) as being indicative of regulatory roles of genetic inputs or enhancer elements in loss and gain of function scenarios.

4.2.1.4 Does the Tv4 Enhancer Resemble the Enhanceosome or the Information Display Model?

The Tv4 FMRFa enhancer shows aspects of both the Enhanceosome model and the Information Display model of enhancer organization and function (Kulkarni and Arnosti, 2003; Merika et al., 1998). With respect to the Enhanceosome model, the whole Tv4 enhancer's activity is contingent on the presence of a single attenuated Mad and Medea binding site. Disruption of either the Mad or Medea site leads to a total loss of enhancer activity, an effect that is equivalent to removing activation of the BMP pathway, such as in *wit, Mad* or *Medea* mutants. Further, deletion of either the HD-RE or the BMP-RE results in dramatic loss of Tv4 enhancer activity as well. Spacing between the HD-RE and BMP-RE may also be important because direct fusion of the HD-RE and the BMP-RE together (with no intervening spacing) produces little to no transgene expression. It is notable, in this regard, that the HD-RE ad BMP-RE retain

relatively conserved spacing of about 100bp in all 12 sequenced *Drosophila* species and that the relative 5'->3' positioning of the HD-RE and BMP-RE are well conserved, even though the actual sequence in the intervening region is poorly conserved (See Figure S 12 and Figure S 13). Therefore it could be argued that the 'all or none' nature of an Enhanceosome, wherein the spacing and presence of all *cis*-elements is necessary for gene activation, is present in the Tv4-enhancer. One could further test this model by inserting extra repeats of the intervening sequence between the HD-RE and BMP-RE and/or deleting portions of the intervening sequence to alter the distance between the elements. If adding extra 'spacer' sequence between the HD-RE and BMP-RE results in loss of function, then these results would support an Enhanceosome model of the FMRFa Tv4 specific enhancer.

Mutation of putative Mad and Apterous binding sites outside of the HD-RE and BMP-RE also generate some degree of loss of enhancer function. This summation of these distributed weaker sites in addition to the HD-RE and BMP-RE to give strong reporter expression is reminiscent of the distributed nature of enhancers proposed in the information display model. Perhaps this suggests that the Tv4 FMRFa enhancer is a mixed model. It is Enhanceosome-like with respect to the HD-RE and BMP-RE minimal *cis*-elements but also displays Information Display-like behaviour across the multiple elements that help to boost expression levels but are less critical to the overall activity of the enhancer. Given that similar hybrid models can be proposed for other enhancers in the literature, this suggests that the Enhanceosome and the Information Display models are really two extremes of a spectrum. Their utility as descriptors and conceptual models is not diminished, but their "pure" versions may in time come to be shown to be rare as more enhancers are examined in detail.

4.2.2 The Timing of Neuron Differentiation by Target Derived Signal Activation and Intrinsic Timer De-Repression

The field has tacitly assumed that pre-specified cells simply 'wait' for target-derived signals to activate terminal differentiation genes. This view was essentially without direct experimental evidence. Work from the Allan lab showed that removal of active BMP-signalling after initial normal differentiation of the Tv4 neuron resulted in loss of FMRFa expression and that active BMP signalling is required throughout the life to maintain *FMRFa* expression (Eade and Allan, 2009). This could be taken to suggest that BMP-signalling acts like a simple on-switch; its activity is required to transcribe FMRFa. Our lab has attempted to find environmental conditions that modulate FMRFa expression, such as starvation, the circadian cycle or age; however, no evidence has been found for a regulatory 'gain' role for BMP signaling in controlling the level of FMRFa expression. My work in this thesis has shown that the BMP-dependence of FMRFa is due to the requirement of Smad acting at *cis*-regulatory elements to combinatorially specify FMRFa expression in the Tv4 neuron. Thus, it was surprising that precocious activation of BMP signalling prior to target contact in Tv4 neurons failed to activate FMRFa expression, in spite of the fact that all known intrinsic transcriptional activators of FMRFa are present in the Tv4 neuron from stage 15. This seemingly paradoxical result was reconciled by showing that Svp acts as a repressor of *FMRFa* transcription (acting at the level of both HD-RE and BMP-RE), which is expressed up to the time of target contact. Thus, the timed loss of Svp expression immediately prior to the time of target contact, makes the Tv4 cell competent to activate FMRFa in response to the target-dependent BMP signal.

Svp acts early in the NB5-6T lineage as well as the specification of the Tv cluster towards the end of the lineage. Although COUP-I and –II factors have been long known to mediate timing events during development to switch neural progenitor fate, our observations here

indicate that this family of transcription factors may also serve as a timing switch to make neurons competent to terminally differentiate. From high throughput expression studies, it is known that Svp expression is maintained in many parts of the *Drosophila* CNS throughout life (Brown et al., 2014). Our data raise the possibility that *svp* expression may be dynamic in these other regions of the postmitotic CNS and perhaps functions to gate the expression of other terminal differentiation genes.

Constant genetic inputs, including target derived signalling, are required to maintain the activation of terminal differentiation genes throughout life, leading to the implication that constant input is required to retain cell identity (Eade and Allan, 2009; Eade et al., 2012). Our data regarding *svp* may additionally suggest that a reciprocal mechanism involving the presence of active suppressive mechanisms may exist in postmitotic neurons that may be important to prevent precocious differentiation of cells and to refine cell identity over developmental timelines. Moreover, such repressive mechanisms could conceivably be re-engaged later in development (or adulthood) as an alternate means to de-differentiate a neuron under certain conditions, even in the presence of other activating inputs. It is intriguing to note that both target-derived BMP signaling and Svp activity are ligand-activated. Svp is an orphan nuclear receptor, raising the possibility that its activity may be regulated by an unknown cell-autonomous or homeostatic mechanism. In this regard, it is intriguing to note that the Allan lab has observed cytoplasmic-localized Svp in Tv4 neurons at numerous stages in the late larva and pupa (unpublished observations).

Downregulation of early-expressed transcription factors for appropriate terminal differentiation may not be restricted to just *svp,* even in the highly focused model of the Tv4 neuron, let alone developing neurons in general. Many of the factors expressed early in Tv

neuron development, that are necessary for Tv4 neuron specification (e.g. *col*, *nab*, and *grh)*, rapidly downregulate prior to terminal differentiation of the Tv4 neuron. These factors stay down-regulated for the life of the animal and are not required for the maintenance of *FMRFa* expression in the adult (Eade et al., 2012). If one maintained the expression of *col*, *nab*, or *grh* using the Gal4-UAS system, would FMRFa expression activate at the right time, or would these factors be suppressive as well?

More generally, factors that suppress differentiation of stem or progenitor cells are well described in multiple tissue types and model systems involving what we have been calling "intrinsic" transcription factors (e.g. GATA factors) as well as downstream signalling molecules from "extrinsic" pathways as diverse as the TGFβ and Wnt pathways (Boland et al., 2004; Watanabe et al., 2005; Ying et al., 2003). Usually, however, these instances involve cell division of the less differentiated progenitors into more differentiated daughter cells. Our example of Svp's activity in the Tv4 neuron then may be a similar process, but simply played out in a post-mitotic cell.

4.2.3 The FMRFa BMP-Response Element Represents a New Type of BMP Response Element

The sequence of the FMRFa BMP-RE is convention breaking, thus we name it the BMP-AE2 type element. The sequence of the FMRFa BMP-AE2 would suggest that it should not bind Medea and not *trans*-activate promoter activity at all, as the C>A conversion within the BMP-AE2 Medea-motif was previously shown by electrophoretic mobility shift assay of BMP-activated cell lysates to uncouple Smad and Schnurri binding from the BMP-SE (Pyrowolakis et al., 2004; Weiss et al., 2010). Further, the terminal thymidine of the BMP-AE2 suggests that the FMRFa BMP-AE2 should recruit Shn and act as a Silencing element instead. Finally, the lack of

a canonical Smad-recruitment site, but the presence of a canonical` Brinker consensus site would further suggest that FMRFa activation would likely occur via *brinker*-dependent de-repression. However, the FMRFa BMP-AE2 quite obviously acts as a transcriptional activator and neither *brinker* nor *schnurri* play any genetic role in its activation. Instead, both genetic and biochemical evidence demonstrate that Medea and Mad do interact with this sequence; however, Medea interacts in a highly attenuated manner that results in much reduced Smad binding, relative to the other BMP-RE types, BMP-AE and BMP-SE (See Chapter 3).

Many of the BMP response elements that have previously been identified and verified in *Drosophila* have been functionally examined in the context of the wing imaginal disc or in early embryo dorso-ventral patterning. Since Tv4 neurons sample BMP ligand at a single point, the synapse, rather than existing as cells in an epithelial sheet in a BMP morphogen gradient, I would suggest that the dynamics with which the FMRFa BMP-AE2 responds to BMP signalling may be different. There may be no need for Brinker de-repression to refine the BMP response.

Discovery of other putative BMP-AE2-like BMP-REs will help verify if the FMRFa BMP-RE is a unique BMP-RE, or *the* founding member of a novel family of BMP response elements (See Section 4.3.3 for proposed experiments). Given that FMRFa and the Tv4 FMRFa enhancer were not affected by *brk* or *shn* loss of function, further investigation into the presence of, and function of, components of the BMP pathway, such as Brk, Shn, and Dad in the fly CNS may help shed light on the sequence flexibility of BMP-RE sequences in the CNS. This is an area that is currently under active investigation in the Allan laboratory.

In summary, when we started the work, it was deemed possible that FMRFa expression was activated in a BMP-dependent manner by an indirect mechanism. For example, neurotrophin 3

dependent differentiation of spinal motoneurons occurs through the activation of ETS transcription factors which then turn on other genes necessary for differentiation (de Nooij et al., 2013). Our identification of a unique BMP response element, the third one to be well-defined in *Drosophila*, demonstrates that there may be higher flexibility in the sequences that are BMP responsive throughout the genome than previously appreciated.

4.3 Limitations And Future Directions

4.3.1 Limitations of Enhancer Analysis and Lessons Learned from Our Approach

4.3.1.1 Sufficiency and Necessity are Key

Enhancer analysis is typically performed with the first objective of defining the necessary and sufficient regions of non-coding DNA that regulate a gene's expression. After these regions have been defined, further work on transcription factor binding, chromatin modification, and enhancer architecture can be performed. Initial identification of necessary and sufficient elements can be technically challenging and highly labour intensive. Designing better experiments to optimize the identification of important regulatory regions would represent an improvement in the current art of enhancer bashing and an increase in the quality of life for graduate students. From what I have learned in the dissection of the *FMRFa* enhancer, I will discuss my views on optimal approaches to enhancer bashing in the following sections.

4.3.1.2 Large Regions Followed by Identification of Single Sites

Enhancers described in the literature vary greatly in size, often ranging from hundreds to thousands of base pairs. Many studies will identify regions in the hundreds of base pair range and then isolate putative binding sites for interrogation by mutagenesis. Isolation of a binding site that shows strong loss or gain of function of reporter activity is usually used as an indication

of a TF's binding activity at that site. Biochemical evidence in the form of ChIP or bandshift assays is usually used as secondary verification, although the increasingly common use of ChIP-seq type technologies is making this method a primary tool for TF binding and enhancer detection. However, often residual enhancer activity is left. Does this mean that there are other binding sites, potentially more cryptic, to the same transcription factor? This can be tested by follow-up experiments by placing the mutated enhancer in a mutant background; if there is further loss of function, potentially more than one binding site for the same factor may be necessary for enhancer activity. As further detailed analysis of enhancer is performed, it is becoming apparent that many, if not most, are composed of both essential sequences and booster or 'shadow' *cis*-elements that provide some level of residual activity. Comprehensive analysis of all of these sites may be necessary to parse the necessary and sufficient elements.

4.3.1.2.1 Divide and Conquer

Many large enhancers are divided into multiple shorter minimal elements. Defining the boundaries and function of minimal elements can be both intellectually and technically difficult. In this thesis we took a candidate based approach, taking advantage of the broad deletion analysis performed by the Taghert lab (Benveniste and Taghert, 1999; Benveniste et al., 1998). Additionally, sequence-specific regulators with known transcription factor binding sites had been identified, and we matched them up with evolutionarily conserved regions to help potentially predict functionality. My initial goal was to determine where (and ultimately how) the known inputs were interpreted at the level of the enhancer, thus this candidate approach appeared the most direct approach. Our sequence analysis pointed to the existence of a high likelihood Ap/Smad integration point at the HD-RE, given the tight juxtaposition of putative Ap, Mad and Medea sequences. The only apparent anomaly at this *cis*-element was the inappropriate

orientation of the Medea site in HD-RE (it is in reverse complement orientation compared to other BMP-REs) with regards the Mad site, an orientation shown to abrogate Smad complex formation (Gao and Laughon, 2007). Otherwise, this site seemed the only location with the requisite Ap, Mad and Medea-binding sites, and conceptually promised a simple cooperative model for Ap/Smad enhancer binding and synergistic promoter trans-activation. We also found that the three Apterous sites previously identified by the Taghert and Thomas groups were indeed highly conserved, and intriguingly were adjacent to potential (although in the case of HD-B/C, low quality) Mad-binding sites; again, raising the possibility of cooperatively acting Ap and Smad binding sites. We directly tested these scenarios with our point mutants. Only the putative HD-A site proved to be critical. Further, out of the six possible Mad binding sites, only one of them gave a strong loss of function phenotype, and this one had no consensus Medea site nor a juxtaposed HD site. Our genetic analyses of these minimal *cis*-elements gave rise to unexpected results and led us to the non-conventional FMRFa BMP-RE and our conclusion that Ap and Smads synergize between separable *cis*-elements. Therefore, the informatics approach led us to the correct sequences, but our initial interpretation of the function of those sequences, based on previous literature, was incorrect.

Without the benefit of known regulatory inputs and a previously published deletion series, how could we have most efficiently identified the correct sequences and predict/test their inputs? One would clearly first use bioinformatics methods to predict putative TF binding sites (discussed more fully below) and convenient islands of high evolutionary conservation. This would likely identify the essential sequences (although not necessarily in all cases; as discussed for the lack of rigid transcription factor binding site conservation within the stripe 2 enhancer), but also the less essential and numerous non-functional ones. Therefore, we would still be left

with the issue of which sequences were the essential ones. Thus, instead of performing all point mutants, as we did initially prior to re-doing the deletion series, in retrospect, simple tiling deletions of ~40 bp each could have been used to generate a deletion series in the Tv4-enhancer (doing our best to retain the inclusivity of conservation islands within the 40bp constructs, so as to minimize bisecting an essential element). Of these ~11 constructs, at least two would have shown strong loss of function since they would have spanned the HD-RE and BMP-RE. In the case of the Tv4 enhancer, this would have very rapidly focused us on the HD-RE and BMP-RE regions, for directed substitution analysis. Due to the approximately 3 months between planning a construct and analysis of the first transgenic animal, batch production of these deletions would be strategic. If a very large enhancer was being analyzed, division of larger conservation regions, with subsequent ~40bp tiled deletions may prove more fruitful. Potentially, such an approach would find minimal fragments in fewer constructs and transgenic lines produced.

The approaches outlined above would identify the broad regions of a gene's regulatory regions, and whether critical transcription factor binding sites were dispersed or clustered. In all cases, however, we are left with the problem of what defines a "minimal" enhancer (the smallest fragment capable to drive expression in a given pattern). Successively smaller fragments are often weaker than larger fragments, as we found when isolating the HD-RE and BMP-RE. Moreover, if a minimal *cis*-element only drives reporter activity when it is concatemerized, is the minimal enhancer actually 'faithful', or is it an artifact? We favour the view, that in such cases, if the transcription factors that act at that minimal element are clearly capable of generating the correct highly specific spatiotemporal expression pattern, then this cannot reflect mere artifact; but is an actual reflection of the spatiotemporal information contained within that *cis*-element (if not the expression level information). These considerations are further compounded by potential

biases when picking loci for site-directed integration. We specifically chose attP2 in our study because it is well known to have high expression with low background expression of transgenes in the CNS. This choice, however, may have biased our results to only detecting positive regulatory sequences rather than repressive ones. The Taghert lab previously reported three regions of repressive activity in the Tv4 FMRFa enhancer, but we detected none in the form of ectopic reporter expressing cells. If our experiments were repeated with an attP-docking site that gave some basal reporter activity, would we have potentially detected repressive regions?

4.3.1.2.2 Chromatin and Positional Effects of Reporter Constructs

Intensive 'bashing' of enhancers still experimentally utilizes random or site-directed integration of reporter constructs into non-native genomic contexts. Notably, those used in *C. elegans* often generate extrachomosomal arrays of reporter constructs with variable lengths of concatemerization during analysis, and traditional P-element based transgenesis methods in *Drosophila* suffer from strong and variable influences from the surrounding genomic context. Local chromatin effects are well known to affect reporter expression. Indeed, these positional effects have even been used to identify response elements to chromatin modifying factors such as Polycomb (i.e. mini-white gene expression and variegation in the *Drosophila* eye after mobilizing a reporter through the genome (Okulski et al., 2011). It is therefore expected that the context in which the reporter is analyzed would have a profound effect on reporter expression. Prediction of enhancers in the current 'omics' era has been facilitated by identification of chromatin marks enriched in enhancer regions such as H3K27ac and H3K27me3 (Zhu et al., 2013). But at the same time, expression of a reporter is often highly reproducible despite being in loci with a non-native chromatin locus, since several different randomly integrated transgenes often give highly similar expression patterns. These considerations focus on the 'spill over'

influence of chromatin effects onto the integrated constructs. An alternative view can also be taken; regulatory sequences may well generate their own epigenetic effects. The very fact that enhancers can function with high fidelity in non-native genomic loci strongly suggest that enhancers, as well as the TFs that bind them, generate their own chromatin environments. Interaction in *trans* with other regulatory loci of the genome (the three dimensional architecture) is now known to be widespread, as regularly demonstrated by chromatin-chromatin capture assays (Reviewed in (Razin et al., 2013)). This may explain the robustness of enhancer activity in non-native loci since these higher order interactions can "reach" across the genome to form appropriate chromatin domains in three dimensions despite being in the wrong one-dimensional location when linear DNA sequence is considered.

One future avenue of study may be to integrate an attP docking site into a well-studied part of the genome using homologous recombination. Previous studies have "hopped" attB containing transposons through the *Drosophila* genome and used mini-white reporter expression to detect differences in local chromatin behaviour in response to variations in the composition of polycomb response elements that were integrated in to the attP side (Okulski et al., 2011). With respect to any given defined enhancer, the local chromatin (e.g. histone modifications) could then be characterized through ChIP of regions adjacent to the attB site in the presence or absence of enhancers integrated in to the attP site using the PhiC31 integrase system. Different mutated versions of an enhancer could then be integrated. Gain or loss of known functional elements could then be correlated with alterations in local chromatin marks. Rather than simply studying chromatin marks associated with known enhancers, one can actually measure changes in chromatin composition in response to changes in enhancer structure yielding a better mechanistic insight rather than simply correlative data that lines up ChIP-seq data with a reference genome.

4.3.1.2.3 Enhancer Analysis in the CRISPR Era

Recent work on the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) and CRISPR-associated (Cas) genes now allows researchers to induce double stranded or single stranded DNA breaks at targeted loci in the genome. This system may circumvent many of the problems I have outlined above with reporter constructs since mutagenesis of enhancers can be done at their endogenous locus. In the applied form of CRISPR\Cas9, synthetic guide RNAs (sgRNAs) and the Cas9 nuclease are co-expressed in germline or somatic cells (Gratz et al., 2014; Jinek et al., 2012; Port et al., 2014). The sgRNA targets a complementary sequence of 20 nucleotides at which the Cas9 cuts the DNA. Double Stranded DNA Breaks (DSBs) can be repaired by the host Non-Homologous End Joining (NHEJ) pathway, which can result in insertions and/or deletions (indels). If a template with homology to the region of the DSB is supplied, the DSB may be repaired by the homology-directed repair (HDR) pathway. In HDR the homologous template can be part of genomic DNA, or injected/transformed in to the targeted cells (Overballe-Petersen et al., 2013)). Use of a mutated Cas9 protein that causes only single stranded breaks can be used to ensure only HDR and not NHEJ is used to prevent indels (Cong et al., 2013). Already, multiple groups have reported success in using CRISPR/Cas9 to mutagenize the *Drosophila* genome (Gratz et al., 2014; Port et al., 2014). Multiple variations of CRISPR/Cas9 have already been reported for expression of sgRNAs and Cas9 in the germline for germline mutagenesis as well as Gal4 inducible systems (Xue et al., 2014). The sgRNA may be delivered as the RNA itself injected in to the germline during embryo genesis, an injected plasmid that transiently expresses the RNA, or vector integrated in to the genome (Ren et al., 2014).

However, use of CRISPR/Cas9 has its own special considerations to be taken in to account. First, if a gene of interest is to have its regulatory regions interrogated, a method of detecting the gene's product is required. This could be anything from immunohistochemical detection of the gene product itself, to RNA *in situ* hybridization or using CRISPR/Cas9 and HDR to integrate a GFP knock-in reporter construct in to the coding sequence.

For characterization of large regulatory regions on the order of hundreds of base pairs in size, use of two sgRNAs could be used to cause DSBs on either side of a putative enhancer and the intervening sequence lost as an episomal fragment. Generally this would be done in the germline and the progeny screened for mutations by PCR to detect the deletion. Alternatively, a construct with homology arms that cause a deletion when recombined into the target locus could be used in conjunction with HDR to recombine in a positive marker (e.g. *mini-white* or a fluorescent reporter driven by something like the actin promoter) to delete a span of regulatory sequence similar to what is conventionally done in mice for knock-outs (Ren et al., 2014). This would make screening for mutations easier in a practical sense but integration of a transgene would alter the surrounding regulatory sequence and be a confounding factor in defining local regulatory regions. Alternatively, substitution mutagenesis rather than simple deletion could be performed by using HDR.

The integration of a reporter gene to indicate successful deletion or insertion of regulatory sequence is an example of the Observer Effect (often improperly referred to as the Heisenberg uncertainty principle) that many lines of research deal with regularly - the act of observing a phenomenon changes the behaviour of the phenomenon. Even though this system presents distinct advantages over simple reporter based systems, it is not without its own limitations.

For interrogation of single transcription factor binding sites or small regulatory regions (e.g. the HD-RE and BMP-RE of the *FMRFa* Tv4 enhancer*)* a single guide RNA could be used to cause small indels on the order of single to tens of base pairs at a single locus. Screening for these small indels is comparatively more difficult from a technical point of view since resolution of PCR products with variations of tens of nucleotides is difficult even using acrylamide gels. Insertion of a reporter sequence would disrupt the spacing in the enhancer, which could generate artifactual data, especially if the behaviour of the enhancer was enhanceosome-like. This technical problem may be overcome by using locked nuclei acid primers that substantially increase primer annealing specificity and allow differentiation in PCR based assays of single nucleotide polymorphisms (Mouritzen et al., 2003). This approach of causing indels in a putative regulatory region could be extended by causing single DSBs in a DNA-repair impaired mutant (e.g. the *Blm* helicase mutant I used to generate CG32105 alleles in Appendix A) to generate larger deletions with variations in size and span. In effect, one would screen for indels through a putative regulatory region in a manner similar to how p-element imprecise excisions are performed (McVey, 2010; Ou, 2013). If little is known about the regulatory regions of a gene, this "unbiased" allelic series could yield rich data.

Ultimately, a hybrid of the attP/attB system I have used in my thesis and CRISPR/Cas9 mutagenesis may become the most useful approach for analyzing enhancers in the future. Once the rough span of an enhancer is defined, CRISPR/Cas9 could be used to cause a DSB at the locus and an attP site integrated through HDR with homology arms that delete the enhancer. So in effect an attP site is left where the enhancer once was. One could then integrate wild-type and mutated versions of the enhancer into its endogenous locus and measure endogenous gene

expression and/or reporter activity. If given the chance to start this whole project over with today's current technology, this is the method I would utilize.

4.3.2 Identification of Other FMRFa *Trans*-Regulators

In this thesis, I comprehensively assigned known regulators of FMRFa that are transcription factors and transcription co-factors into two broad categories; those that regulate FMRFa expression in the Tv4 neurons through the 445 bp Tv4-specific enhancer, and those that do not regulate this gene. The factors that I show act directly or indirectly through either the HD-RE or BMP-RE (Ap, Eya, Dac, pMad/BMP signalling, Medea, Svp) still fail to account for the highly specific expression pattern of the full 445 bp FMRFa enhancer and its minimal *cis*-elements.

Identification of the transcription factors that act through the *FMRFa* BMP-RE will potentially explain its exquisite specificity. Analysis of multiple other Smad regulated genes, mostly in vertebrates to date, have shown that Smads rarely bind and act at a response element without other non-Smad transcription factors (Marques, 2005). I used *in silico* methods to predict other transcription factors that may bind at the *FMRFa* BMP-RE. These methods initially used the JASPAR database of known transcription factor position weight matrices (Portales-Casamar et al., 2010). However, recently released data of libraries of transcription factor binding motifs based on newer high throughput methods of determining transcription factor binding sites, such as SELEX and Bacterial 1 hybrid, has dramatically increased the number of TFs with detailed position weight matrix data for about 650 TFs (Shazman et al., 2014). Use of such tools that compute the likelihood of a certain motif occurring in a region of DNA versus the whole genome, such as FIMO within the Meme Suite, are being used by the Allan lab to refine the list of candidate transcription factors that may act at the Tv4-enhancer (and the HD-RE and BMP-RE specifically) (Grant et al., 2011). Sequences that are predicted to be present and are

statistically enriched in the Tv4-enhancer will be given top priority, and those TFs will be screened for the role in Tv4-enhancer activity by *UAS-RNAi*-based knockdown of their expression.

4.3.2.1 Tv4 Enhancer RNAi Screen

An RNAi screen is being initiated using an extremely bright GFP reporter vector that utilizes the same 445bp Tv4-enhancer characterized in Chapter 2, but has been transposase-hopped into a locus that allows for extremely strong reporter expression. This published reporter is substantially brighter than the $TV^{WT\text{-}nEYFP}$ used in this study and can be detected directly through the body wall, obviating the need for laborious dissection, and hence greatly accelerating the screening process for essential TFs (Benito-Sipos et al., 2011). Since deletion analysis found that the Tv4-enhancer is contingent on the activity of the 39 base pairs comprising the BMP-RE, strong loss of GFP reporter activity in this screen will most likely indicate a loss of regulators of the FMRFa BMP-RE, so long as the Tv4 neuron is born and differentiates appropriately. Since *ap* is expressed only after Tv4 is born, ap^{Gal4} will be used to drive the *UAS-RNAi* so that effects on Tv4 specification and early differentiation are avoided (Allan et al., 2005; Allan et al., 2003). *UAS-dicer2* will be used to enhance the RNAi effects. A secondary screen will use the *4xBMP-RE-YFP* reporter on the 'hits' from the 445 bp enhancer screen in order to determine if the factors identified in the first screen indeed act on the BMP-RE.

Transcription factors will be prioritized by first screening those than have been bioinformatically predicted to potentially bind to the FMRFa BMP-RE sequence. More than one *UAS-RNAi* construct from different libraries will be used on these candidates since RNAi constructs have different efficacy depending on where they are targeted and how the double stranded RNA is specifically produced. The next set of RNAi lines to be tested will be

transcription factors identified as enriched in the CNS. Similar to Chapter 2, loss of function alleles to the putative regulators will be obtained or generated to verify the role of that factor. Mutational analyses of potential binding sites for the factor will further confirm its activity at the BMP-RE. If no hits are detected in this prioritized list of TFs, all 753 putative sequence-specific TFs in the *Drosophila* genome will be screened.

4.3.3 Identification of Other Atypical BMP Response Elements

The FMRFa BMP-RE is a previously uncharacterized type of BMP Response element. Does it represent a nervous system solution to BMP-regulated gene expression, or is a unique genomic event? In Chapter 3, I identified 27 unique highly conserved sequences in the *Drosophila melanogaster* genome, *in silico,* that conform to the sequence and spacing of the Mad and Medea sites in the *FMRFa* BMP-RE. 19 of these BMP-AE2 were found to be near genes enriched in the CNS (See Table 1). All 19 BMP-AE2 will be cloned (with flanking 2kb) and placed in the pThunderbird nEGFP reporter system and integrated into attP2. The reporters will be initially screened for BMP signalling responsiveness by crossing the reporters to a drug-inducible BMP signalling knock-down system already in use in the lab (based on the Geneswitch system). Geneswitch is a chimeric GAL4 protein fused to the human progesterone receptor ligand-binding domain that becomes active in the presence of the steroid RU486 (Osterwalder et al., 2001). *Elav* is a regulatory region from the *embryonic lethal abnormal vision* gene that drives Geneswitch protein expression to all neurons postmitotically, including the Tv4 neurons. BMP knockdown in the CNS is achieved by simultaneously driving *UAS-dad* and *UAS-tkv*^{DN} after feeding larvae food containing RU486, which activates the GAL4. Reporters that are down-regulated in animals fed RU486 compared to animals on normal food will be recombined into *wit* and *Medea* mutant backgrounds to further investigate the necessity of BMP and *Medea* dependence. I will

also determine if these reporters are *shn* or *brk* responsive using gain and loss of function experiments similar to those described in Chapter 3.

The sequences flanking these other AE2-like BMP responsive elements will be compared to the FMRFa BMP-RE sequence. Expression and BMP-dependence in the Tv4 neuron will also be determined. Commonalities in flanking sequences may expedite the identification of transcription factors cooperatively binding with Mad and Medea.

4.3.4 Applying Lessons Learned From *Drosophila* BMP- Response Elements to other Model Systems.

The Smads and their transcriptional activities and binding sites are extremely well conserved. Lessons from *Drosophila* BMP signalling have already proven highly pertinent to mammals when initial components of the BMP pathways were being identified (Ross and Hill, 2008b; Schmierer and Hill, 2007). BMP signaling is critical in the development of many organ systems, as well as many aspects of adult tissue homeostasis. As in *Drosophila,* loss of BMP Pathway function is pleiotropic and can be embryonically lethal or presents with marked abnormalities, especially in the heart, lungs, and mesoderm derived tissues. Adult tissues also rely on BMP signaling for homeostasis, and examples include fracture repair initiation and pulmonary vascular remodeling (reviewed in (Wang et al., 2014)). Despite Smad-binding sites being better defined in mammalian model systems at a biochemical level because of genome-wide ChIP studies, *bonafide* BMP response elements in vertebrates are less well characterized (Fei et al., 2010; Morikawa et al., 2011). In effect, researchers are left with a "sloppy" position weight matrix of Smad binding sites (it seems from these studies nearly any GC-rich region seems to attract Smads(!)) with little indication of what parts of the genome are actual functional gene regulatory regions with BMP responsiveness. This problem is exacerbated by 'laundry lists' of

microarray or RNA-seq data of up or down regulated genes in BMP gain or loss off function situations (Fei et al., 2010; James et al., 2010; Morikawa et al., 2011). (Freeman et al., 2012). Determining which genes are directly affected by BMP signalling and what are secondary effects will be key to understanding the exact role of BMP signalling in vertebrate physiology. My preliminary data show that Brinker may not be involved in the BMP-responsiveness of AE2-like sequences in the CNS. Intriguingly, there is no obvious ortholog of Brinker in vertebrates. This suggests that BMP signalling in the *Drosophila* CNS may be more similar to BMP signalling in mammals than the better-studied *Drosophila* wing disc. Therefore, if bioinformatically predicted *Drosophila* AE2-like sites are verified through reporter based assays as proposed in Section 4.3.3, a rich resource of trusted "real" BMP response elements would be generated, which can then be used to predict BMP response elements in vertebrate genomes (the same can be done for AE and SE-like elements in the *Drosophila* genome). Simple position weight matrices of verified BMP response elements and/or more complex models based on algorithms such as support vector machines that "distinguish" between functional and non-functional elements (Cui et al., 2014).

Collectively verification and analysis of BMP Response Elements in *Drosophila* represents the establishment of a foundation for better understanding the regulatory role of BMP signalling in both vertebrates and invertebrates in both developmental roles and pathophysiology.

BIBLIOGRAPHY

Aberle, H., Haghighi, A. P., Fetter, R. D., McCabe, B. D., Magalhaes, T. R. and Goodman, C. S. (2002). wishful thinking encodes a BMP type II receptor that regulates synaptic growth in Drosophila. *Neuron* **33**, 545-558.

Adams, M. D., McVey, M. and Sekelsky, J. J. (2003). Drosophila BLM in double-strand break repair by synthesis-dependent strand annealing. *Science* **299**, 265-267.

Affolter, M. and Basler, K. (2007). The Decapentaplegic morphogen gradient: from pattern formation to growth regulation. *Nat Rev Genet* **8**, 663-674.

Ai, X., Cappuzzello, J. and Hall, A. K. (1999). Activin and bone morphogenetic proteins induce calcitonin gene-related peptide in embryonic sensory neurons in vitro. *Mol Cell Neurosci* **14**, 506-518.

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.

Allan, D. W., St Pierre, S. E., Miguel-Aliaga, I. and Thor, S. (2003). Specification of neuropeptide cell identity by the integration of retrograde BMP signaling and a combinatorial transcription factor code. *Cell* **113**, 73-86.

Allan, D. W. and Thor, S. (2015). Transcriptional selectors, masters, and combinatorial codes: regulatory principles of neural subtype specification. *Wiley Interdiscip Rev Dev Biol*.

Allen, M. J., Shan, X., Caruccio, P., Froggett, S. J., Moffat, K. G. and Murphey, R. K. (1999). Targeted expression of truncated glued disrupts giant fiber synapse formation in Drosophila. *J Neurosci* **19**, 9374-9384.

Andrews, J. C., Fernandez, M. P., Yu, Q., Leary, G. P., Leung, A. K., Kavanaugh, M. P., Kravitz, E. A. and Certel, S. J. (2014). Octopamine neuromodulation regulates Gr32a-linked aggression and courtship pathways in Drosophila males. *PLoS Genet* **10**, e1004356.

Andrioli, L. P., Vasisht, V., Theodosopoulou, E., Oberstein, A. and Small, S. (2002). Anterior repression of a Drosophila stripe enhancer requires three position-specific mechanisms. *Development* **129**, 4931-4940.

Angley, C., Kumar, M., Dinsio, K. J., Hall, A. K. and Siegel, R. E. (2003). Signaling by bone morphogenetic proteins and Smad1 modulates the postnatal differentiation of cerebellar cells. *J Neurosci* **23**, 260-268.

Arca, B., Zabalou, S., Loukeris, T. G. and Savakis, C. (1997). Mobilization of a Minos transposon in Drosophila melanogaster chromosomes and chromatid repair by heteroduplex formation. *Genetics* **145**, 267-279.

Arnosti, D. N., Barolo, S., Levine, M. and Small, S. (1996). The eve stripe 2 enhancer employs multiple modes of transcriptional synergy. *Development* **122**, 205-214.

Arnosti, D. N. and Kulkarni, M. M. (2005). Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* **94**, 890-898.

Ball, R. W., Warren-Paquin, M., Tsurudome, K., Liao, E. H., Elazzouzi, F., Cavanagh, C., An, B. S., Wang, T. T., White, J. H. and Haghighi, A. P. (2010). Retrograde BMP signaling controls synaptic growth at the NMJ by regulating trio expression in motor neurons. *Neuron* **66**, 536-549.

Barolo, S., Carver, L. A. and Posakony, J. W. (2000). GFP and beta-galactosidase transformation vectors for promoter/enhancer analysis in Drosophila. *Biotechniques* **29**, 726, 728, 730, 732.

Baumgardt, M., Karlsson, D., Salmani, B. Y., Bivik, C., MacDonald, R. B., Gunnar, E. and Thor, S. (2014). Global programmed switch in neural daughter cell proliferation mode triggered by a temporal gene cascade. *Dev Cell* **30**, 192-208.

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

Baumgardt, M., Miguel-Aliaga, I., Karlsson, D., Ekman, H. and Thor, S. (2007). Specification of neuronal identities by feedforward combinatorial coding. *PLoS Biol* **5**, e37.

Benito-Sipos, J., Ulvklo, C., Gabilondo, H., Baumgardt, M., Angel, A., Torroja, L. and Thor, S. (2011). Seven up acts as a temporal factor during two different stages of neuroblast 5-6 development. *Development* **138**, 5311-5320.

Benveniste, R. J. and Taghert, P. H. (1999). Cell type-specific regulatory sequences control expression of the Drosophila FMRF-NH2 neuropeptide gene. *J Neurobiol* **38**, 507-520.

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.

Bergman, O., Westberg, L., Nilsson, L. G., Adolfsson, R. and Eriksson, E. (2010). Preliminary evidence that polymorphisms in dopamine-related transcription factors LMX1A, LMX1B and PITX3 are associated with schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* **34**, 1094-1097.

Berke, B., Wittnam, J., McNeill, E., Van Vactor, D. L. and Keshishian, H. (2013). Retrograde BMP signaling at the synapse: a permissive signal for synapse maturation and activity-dependent plasticity. *J Neurosci* **33**, 17937-17950.

Billeter, J. C., Villella, A., Allendorfer, J. B., Dornan, A. J., Richardson, M., Gailey, D. A. and Goodwin, S. F. (2006). Isoform-specific control of male neuronal differentiation and behavior in Drosophila by the fruitless gene. *Curr Biol* **16**, 1063-1076.

Birkholz, O., Rickert, C., Berger, C., Urbach, R. and Technau, G. M. (2013). Neuroblast pattern and identity in the Drosophila tail region and role of doublesex in the survival of sex-specific precursors. *Development* **140**, 1830-1842.

Bischof, J., Maeda, R. K., Hediger, M., Karch, F. and Basler, K. (2007). An optimized transgenesis system for Drosophila using germ-line-specific phiC31 integrases. *Proc Natl Acad Sci U S A* **104**, 3312-3317.

Blitz, I. L. and Cho, K. W. (2009). Finding partners: how BMPs select their targets. *Dev Dyn* **238**, 1321-1331.

Boland, G. M., Perkins, G., Hall, D. J. and Tuan, R. S. (2004). Wnt 3a promotes proliferation and suppresses osteogenic differentiation of adult human mesenchymal stem cells. *J Cell Biochem* **93**, 1210-1230.

Boley, N., Wan, K. H., Bickel, P. J. and Celniker, S. E. (2014). Navigating and mining modENCODE data. *Methods* **68**, 38-47.

Bonini, N. M., Leiserson, W. M. and Benzer, S. (1998). Multiple roles of the eyes absent gene in Drosophila. *Dev Biol* **196**, 42-57.

Bossing, T., Udolph, G., Doe, C. Q. and Technau, G. M. (1996). The embryonic central nervous system lineages of Drosophila melanogaster. I. Neuroblast lineages derived from the ventral half of the neuroectoderm. *Dev Biol* **179**, 41-64.

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.

Brody, T. and Odenwald, W. F. (2000). Programmed transformations in neuroblast gene expression during Drosophila CNS lineage development. *Dev Biol* **226**, 34-44.

Brown, J. B., Boley, N., Eisman, R., May, G. E., Stoiber, M. H., Duff, M. O., Booth, B. W., Wen, J., Park, S., Suzuki, A. M., et al. (2014). Diversity and dynamics of the Drosophila transcriptome. *Nature* **512**, 393-399.

Brugger, S. M., Merrill, A. E., Torres-Vazquez, J., Wu, N., Ting, M. C., Cho, J. Y., Dobias, S. L., Yi, S. E., Lyons, K., Bell, J. R., et al. (2004). A phylogenetically conserved cis-regulatory module in the Msx2 promoter is sufficient for BMP-dependent transcription in murine and Drosophila embryos. *Development* **131**, 5153-5165.

Bucelli, R. C., Gonsiorek, E. A., Kim, W. Y., Bruun, D., Rabin, R. A., Higgins, D. and Lein, P. J. (2008). Statins decrease expression of the proinflammatory neuropeptides calcitonin gene-related peptide and substance P in sensory neurons. *J Pharmacol Exp Ther* **324**, 1172-1180.

Bulger, M. and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**, 327-339.

Burz, D. S., Rivera-Pomar, R., Jackle, H. and Hanes, S. D. (1998). Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the Drosophila embryo. *EMBO J* **17**, 5998-6009.

Calo, E. and Wysocka, J. (2013). Modification of enhancer chromatin: what, how, and why? *Molecular cell* **49**, 825-837.

Campbell, G. and Tomlinson, A. (1999). Transducing the Dpp morphogen gradient in the wing of Drosophila: regulation of Dpp targets by brinker. *Cell* **96**, 553-562.

Castellanos, M. C., Tang, J. C. and Allan, D. W. (2013). Female-biased dimorphism underlies a female-specific role for post-embryonic Ilp7 neurons in Drosophila fertility. *Development* **140**, 3915-3926.

Cau, E. and Blader, P. (2009). Notch activity in the nervous system: to switch or not switch? *Neural Dev* **4**, 36.

Certel, S. J., Savella, M. G., Schlegel, D. C. and Kravitz, E. A. (2007). Modulation of Drosophila male behavioral choice. *Proc Natl Acad Sci U S A* **104**, 4706-4711.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106-1117.

Choi, O. R. and Engel, J. D. (1988). Developmental regulation of beta-globin gene switching. *Cell* **55**, 17-26.

Chou, H. J., Lai, D. M., Huang, C. W., McLennan, I. S., Wang, H. D. and Wang, P. Y. (2013). BMP4 is a peripherally-derived factor for motor neurons and attenuates glutamate-induced excitotoxicity in vitro. *PLoS ONE* **8**, e58441.

Clark, A. G., Eisen, M. B., Smith, D. R., Bergman, C. M., Oliver, B., Markow, T. A., Kaufman, T. C., Kellis, M., Gelbart, W., Iyer, V. N., et al. (2007). Evolution of genes and genomes on the Drosophila phylogeny. *Nature* **450**, 203-218.

Cohen, B., McGuffin, M. E., Pfeifle, C., Segal, D. and Cohen, S. M. (1992). apterous, a gene required for imaginal disc development in Drosophila encodes a member of the LIM family of developmental regulatory proteins. *Genes & development* **6**, 715-729.

Cole, S. H., Carney, G. E., McClung, C. A., Willard, S. S., Taylor, B. J. and Hirsh, J. (2005). Two functional but noncomplementing Drosophila tyrosine decarboxylase genes: distinct roles for neural tyramine and octopamine in female fertility. *J Biol Chem* **280**, 14948-14955.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823.

Core, L. J., Martins, A. L., Danko, C. G., Waters, C. T., Siepel, A. and Lis, J. T. (2014). Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nat Genet* **46**, 1311-1320.

Coulombe, J. N. and Kos, K. (1997). Target tissue influence on somatostatin expression in the avian ciliary ganglion. *Ann N Y Acad Sci* **814**, 209-225.

Crocker, A. and Sehgal, A. (2008). Octopamine regulates sleep in drosophila through protein kinase A-dependent mechanisms. *J Neurosci* **28**, 9377-9385.

Crocker, J., Abe, N., Rinaldi, L., McGregor, A. P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., et al. (2015). Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160**, 191-203.

Cui, S., Youn, E., Lee, J. and Maas, S. J. (2014). An improved systematic approach to predicting transcription factor target genes using support vector machine. *PLoS One* **9**, e94519.

Curtiss, J. and Heilig, J. S. (1998). DeLIMiting development. *BioEssays : news and reviews in molecular, cellular and developmental biology* **20**, 58-69.

da Silva, S. and Wang, F. (2011). Retrograde neural circuit specification by target-derived neurotrophins and growth factors. *Curr Opin Neurobiol* **21**, 61-67.
Dai, J. X., Johnson, R. L. and Ding, Y. Q. (2009). Manifold functions of the Nail-Patella Syndrome gene Lmx1b in vertebrate development. *Dev Growth Differ* **51**, 241-250.
Dasen, J. S. (2009). Transcriptional networks in the early development of sensory-motor circuits. *Curr Top Dev Biol* **87**, 119-148.
de Nooij, J. C., Doobar, S. and Jessell, T. M. (2013). Etv1 inactivation reveals proprioceptor subclasses that reflect the level of NT3 expression in muscle targets. *Neuron* **77**, 1055-1068.
De Vos, K. J., Grierson, A. J., Ackerley, S. and Miller, C. C. (2008). Role of axonal transport in neurodegenerative diseases. *Annu Rev Neurosci* **31**, 151-173.
Demarque, M. and Spitzer, N. C. (2010). Activity-dependent expression of Lmx1b regulates specification of serotonergic neurons modulating swimming behavior. *Neuron* **67**, 321-334.
Deppmann, C. D. and Ginty, D. D. (2006). Retrograde control of neural circuit formation. *Cell* **127**, 1306-1307.
di Sanguinetto, S. A., Dasen, J. S. and Arber, S. (2008). Transcriptional mechanisms controlling motor neuron diversity and connectivity. *Curr Opin Neurobiol* **18**, 36-43.
Ding, Y. Q., Marklund, U., Yuan, W., 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.
Duncan, I. M. (1982). Polycomblike: a gene that appears to be required for the normal expression of the bithorax and antennapedia gene complexes of Drosophila melanogaster. *Genetics* **102**, 49-70.
DuVal, M. G., Gilbert, M. J., Watson, D. E., Zerulla, T. C., Tierney, K. B. and Allison, W. T. (2014). Growth differentiation factor 6 as a putative risk factor in neuromuscular degeneration. *PLoS ONE* **9**, e89183.
Eade, K. T. and Allan, D. W. (2009). Neuronal phenotype in the mature nervous system is maintained by persistent retrograde bone morphogenetic protein signaling. *J Neurosci* **29**, 3852-3864.
Eade, K. T., Fancher, H. A., Ridyard, M. S. and Allan, D. W. (2012). Developmental transcriptional networks are required to maintain neuronal subtype identity in the mature nervous system. *PLoS Genet* **8**, e1002501.
Evans, N. C., Swanson, C. I. and Barolo, S. (2012). Sparkling insights into enhancer structure, function, and evolution. *Current topics in developmental biology* **98**, 97-120.
Fei, T., Xia, K., Li, Z., Zhou, B., Zhu, S., Chen, H., Zhang, J., Chen, Z., Xiao, H., Han, J. D., et al. (2010). Genome-wide mapping of SMAD target genes reveals the role of BMP signaling in embryonic stem cell fate determination. *Genome Res* **20**, 36-44.
Feng, X. H. and Derynck, R. (2005). Specificity and versatility in tgf-beta signaling through Smads. *Annual review of cell and developmental biology* **21**, 659-693.
Ford, E. and Thanos, D. (2010). The transcriptional code of human IFN-beta gene expression. *Biochimica et biophysica acta* **1799**, 328-336.
Freeman, T. J., Smith, J. J., Chen, X., Washington, M. K., Roland, J. T., Means, A. L., Eschrich, S. A., Yeatman, T. J., Deane, N. G. and Beauchamp, R. D. (2012). Smad4-mediated signaling inhibits intestinal neoplasia by inhibiting expression of beta-catenin. *Gastroenterology* **142**, 562-571 e562.
Galant, R., Walsh, C. M. and Carroll, S. B. (2002). Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites. *Development* **129**, 3115-3126.
Gantz, V. M. and Bier, E. (2015). Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science* **348**, 442-444.
Gao, S. and Laughon, A. (2006). Decapentaplegic-responsive silencers contain overlapping mad-binding sites. *J Biol Chem* **281**, 25781-25790.
---- (2007). Flexible interaction of Drosophila Smad complexes with bipartite binding sites. *Biochim Biophys Acta* **1769**, 484-496.

Gao, S., Steffen, J. and Laughon, A. (2005). Dpp-responsive silencers are bound by a trimeric Mad-Medea complex. *J Biol Chem* **280**, 36158-36164.
Golic, K. G. (1991). Site-specific recombination between homologous chromosomes in Drosophila. *Science* **252**, 958-961.
Gomis, R. R., Alarcon, C., He, W., Wang, Q., Seoane, J., Lash, A. and Massague, J. (2006). A FoxO-Smad synexpression group in human keratinocytes. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 12747-12752.
Goold, C. P. and Davis, G. W. (2007). The BMP ligand Gbb gates the expression of synaptic homeostasis independent of synaptic growth control. *Neuron* **56**, 109-123.
Goto, T., Macdonald, P. and Maniatis, T. (1989). Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **57**, 413-422.
Grant, C. E., Bailey, T. L. and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017-1018.
Gratz, S. J., Cummings, A. M., Nguyen, J. N., Hamm, D. C., Donohue, L. K., Harrison, M. M., Wildonger, J. and O'Connor-Giles, K. M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. *Genetics* **194**, 1029-1035.
Gratz, S. J., Harrison, M. M., Wildonger, J. and O'Connor-Giles, K. M. (2015). Precise Genome Editing of Drosophila with CRISPR RNA-Guided Cas9. *Methods in molecular biology* **1311**, 335-348.
Gratz, S. J., Ukken, F. P., Rubinstein, C. D., Thiede, G., Donohue, L. K., Cummings, A. M. and O'Connor-Giles, K. M. (2014). Highly specific and efficient CRISPR/Cas9-catalyzed homology-directed repair in Drosophila. *Genetics* **196**, 961-971.
Grieder, N. C., Nellen, D., Burke, R., Basler, K. and Affolter, M. (1995). Schnurri is required for Drosophila Dpp signaling and encodes a zinc finger protein similar to the mammalian transcription factor PRDII-BF1. *Cell* **81**, 791-800.
Grocott, T., Frost, V., Maillard, M., Johansen, T., Wheeler, G. N., Dawes, L. J., Wormstone, I. M. and Chantry, A. (2007). The MH1 domain of Smad3 interacts with Pax6 and represses autoregulation of the Pax6 P1 promoter. *Nucleic Acids Res* **35**, 890-901.
Grosskortenhaus, R., Robinson, K. J. and Doe, C. Q. (2006). Pdm and Castor specify late-born motor neuron identity in the NB7-1 lineage. *Genes Dev* **20**, 2618-2627.
Groth, A. C., Fish, M., Nusse, R. and Calos, M. P. (2004). Construction of transgenic Drosophila by using the site-specific integrase from phage phiC31. *Genetics* **166**, 1775-1782.
Guha, U., Gomes, W. A., Samanta, J., Gupta, M., Rice, F. L. and Kessler, J. A. (2004). Target-derived BMP signaling limits sensory neuron number and the extent of peripheral innervation in vivo. *Development* **131**, 1175-1186.
Gunawardena, S. and Goldstein, L. S. (2004). Cargo-carrying motor vehicles on the neuronal highway: transport pathways and neurodegenerative disease. *J Neurobiol* **58**, 258-271.
Haase, G., Dessaud, E., Garces, A., de Bovis, B., Birling, M., Filippi, P., Schmalbruch, H., Arber, S. and deLapeyriere, O. (2002). GDNF acts through PEA3 to regulate cell body positioning and muscle innervation of specific motor neuron pools. *Neuron* **35**, 893-905.
Habecker, B. A., Symes, A. J., Stahl, N., Francis, N. J., Economides, A., Fink, J. S., Yancopoulos, G. D. and Landis, S. C. (1997). A sweat gland-derived differentiation activity acts through known cytokine signaling pathways. *J Biol Chem* **272**, 30421-30428.
Hadzic, T., Park, D., Abruzzi, K. C., Yang, L., Trigg, J. S., Rohs, R., Rosbash, M. and Taghert, P. H. (2015). Genome-wide features of neuroendocrine regulation in Drosophila by the basic helix-loop-helix transcription factor DIMMED. *Nucleic Acids Res* **43**, 2199-2215.
Haerry, T. E., Khalsa, O., O'Connor, M. B. and Wharton, K. A. (1998). Synergistic signaling by two BMP ligands through the SAX and TKV receptors controls wing growth and patterning in Drosophila. *Development* **125**, 3977-3987.

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.

Hammonds, A. S., Bristow, C. A., Fisher, W. W., Weiszmann, R., Wu, S., Hartenstein, V., Kellis, M., Yu, B., Frise, E. and Celniker, S. E. (2013). Spatial expression of transcription factors in Drosophila embryonic organ development. *Genome Biol* **14**, R140.

Hegarty, S. V., O'Keeffe, G. W. and Sullivan, A. M. (2013). BMP-Smad 1/5/8 signalling in the development of the nervous system. *Prog Neurobiol* **109**, 28-41.

Henriquez, J. P., Krull, C. E. and Osses, N. (2011). The Wnt and BMP families of signaling morphogens at the vertebrate neuromuscular junction. *International journal of molecular sciences* **12**, 8924-8946.

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.

Hippenmeyer, S., Kramer, I. and Arber, S. (2004). Control of neuronal phenotype: what targets tell the cell bodies. *Trends Neurosci* **27**, 482-488.

Hobert, O., Carrera, I. and Stefanakis, N. (2010). The molecular and gene regulatory signature of a neuron. *Trends Neurosci* **33**, 435-445.

Hobert, O. and Westphal, H. (2000). Functions of LIM-homeobox genes. *Trends in genetics : TIG* **16**, 75-83.

Hodge, L. K., Klassen, M. P., Han, B. X., Yiu, G., Hurrell, J., Howell, A., Rousseau, G., Lemaigre, F., Tessier-Lavigne, M. and Wang, F. (2007). Retrograde BMP signaling regulates trigeminal sensory neuron identities and the formation of precise face maps. *Neuron* **55**, 572-586.

Holland, L. Z., Carvalho, J. E., Escriva, H., Laudet, V., Schubert, M., Shimeld, S. M. and Yu, J. K. (2013). Evolution of bilaterian central nervous systems: a single origin? *EvoDevo* **4**, 27.

Hoyer, S. C., Eckart, A., Herrel, A., Zars, T., Fischer, S. A., Hardie, S. L. and Heisenberg, M. (2008). Octopamine in male aggression of Drosophila. *Curr Biol* **18**, 159-167.

Hu, M. C. and Rosenblum, N. D. (2005). Smad1, beta-catenin and Tcf4 associate in a molecular complex with the Myc promoter in dysplastic renal tissue and cooperate to control Myc transcription. *Development* **132**, 215-225.

Huang, A. M., Rehm, E. J. and Rubin, G. M. (2009). Quick preparation of genomic DNA from Drosophila. *Cold Spring Harbor protocols* **2009**, pdb prot5198.

Hudson, J. B., Podos, S. D., Keith, K., Simpson, S. L. and Ferguson, E. L. (1998). The Drosophila Medea gene is required downstream of dpp and encodes a functional homolog of human Smad4. *Development* **125**, 1407-1420.

Huminiecki, L., Goldovsky, L., Freilich, S., Moustakas, A., Ouzounis, C. and Heldin, C. H. (2009). Emergence, development and diversification of the TGF-beta signalling pathway within the animal kingdom. *BMC Evol Biol* **9**, 28.

Inoue, H., Imamura, T., Ishidou, Y., Takase, M., Udagawa, Y., Oka, Y., Tsuneizumi, K., Tabata, T., Miyazono, K. and Kawabata, M. (1998). Interplay of signal mediators of decapentaplegic (Dpp): molecular characterization of mothers against dpp, Medea, and daughters against dpp. *Mol Biol Cell* **9**, 2145-2156.

Isshiki, T., Pearson, B., Holbrook, S. and Doe, C. Q. (2001). Drosophila neuroblasts sequentially express transcription factors which specify the temporal identity of their neuronal progeny. *Cell* **106**, 511-521.

Istrail, S. and Davidson, E. H. (2005). Logic functions of the genomic cis-regulatory code. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4954-4959.

Jacob, J., Maurange, C. and Gould, A. P. (2008). Temporal control of neuronal diversity: common regulatory principles in insects and vertebrates? *Development* **135**, 3481-3489.

James, C. G., Stanton, L. A., Agoston, H., Ulici, V., Underhill, T. M. and Beier, F. (2010). Genome-wide analyses of gene expression during mouse endochondral ossification. *PLoS One* **5**, e8693.

Jazwinska, A., Kirov, N., Wieschaus, E., Roth, S. and Rushlow, C. (1999). The Drosophila gene brinker reveals a novel mechanism of Dpp target gene regulation. *Cell* **96**, 563-573.
Jiang, J. and Levine, M. (1993). Binding affinities and cooperative interactions with bHLH activators delimit threshold responses to the dorsal gradient morphogen. *Cell* **72**, 741-752.
Jin, W., Takagi, T., Kanesashi, S. N., Kurahashi, T., Nomura, T., Harada, J. and Ishii, S. (2006). Schnurri-2 controls BMP-dependent adipogenesis via interaction with Smad proteins. *Dev Cell* **10**, 461-471.
Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821.
Kamiya, Y., Miyazono, K. and Miyazawa, K. (2008). Specificity of the inhibitory effects of Dad on TGF-beta family type I receptors, Thickveins, Saxophone, and Baboon in Drosophila. *FEBS letters* **582**, 2496-2500.
Kanai, M. I., Okabe, M. and Hiromi, Y. (2005). seven-up Controls switching of transcription factors that specify temporal identities of Drosophila neuroblasts. *Dev Cell* **8**, 203-213.
Karlsson, D., Baumgardt, M. and Thor, S. (2010). Segment-specific neuronal subtype specification by the integration of anteroposterior and temporal cues. *PLoS Biol* **8**, e1000368.
Kelly, C. E., Thymiakou, E., Dixon, J. E., Tanaka, S., Godwin, J. and Episkopou, V. (2013). Rnf165/Ark2C enhances BMP-Smad signaling to mediate motor axon extension. *PLoS Biol* **11**, e1001538.
Kim, J., Johnson, K., Chen, H. J., Carroll, S. and Laughon, A. (1997). Drosophila Mad binds to DNA and directly mediates activation of vestigial by Decapentaplegic. *Nature* **388**, 304-308.
Kirkpatrick, H., Johnson, K. and Laughon, A. (2001). Repression of dpp targets by binding of brinker to mad sites. *J Biol Chem* **276**, 18216-18222.
Kitisin, K., Saha, T., Blake, T., Golestaneh, N., Deng, M., Kim, C., Tang, Y., Shetty, K., Mishra, B. and Mishra, L. (2007). Tgf-Beta signaling in development. *Sci STKE* **2007**, cm1.
Klose, M. K., Dason, J. S., Atwood, H. L., Boulianne, G. L. and Mercier, A. J. (2010). Peptide-induced modulation of synaptic transmission and escape response in Drosophila requires two G-protein-coupled receptors. *J Neurosci* **30**, 14724-14734.
Knoblich, J. A. (2008). Mechanisms of asymmetric stem cell division. *Cell* **132**, 583-597.
Kohwi, M. and Doe, C. Q. (2013). Temporal fate specification and neural progenitor competence during development. *Nature reviews. Neuroscience* **14**, 823-838.
Kohwi, M., Hiebert, L. S. and Doe, C. Q. (2011). The pipsqueak-domain proteins Distal antenna and Distal antenna-related restrict Hunchback neuroblast expression and early-born neuronal identity. *Development* **138**, 1727-1735.
Kulkarni, M. M. and Arnosti, D. N. (2003). Information display by transcriptional enhancers. *Development* **130**, 6569-6575.
Lamba, P., Khivansara, V., D'Alessio, A. C., Santos, M. M. and Bernard, D. J. (2008). Paired-like homeodomain transcription factors 1 and 2 regulate follicle-stimulating hormone beta-subunit transcription through a conserved cis-element. *Endocrinology* **149**, 3095-3108.
Langley, C. H., Stevens, K., Cardeno, C., Lee, Y. C., Schrider, D. R., Pool, J. E., Langley, S. A., Suarez, C., Corbett-Detig, R. B., Kolaczkowski, B., et al. (2012). Genomic variation in natural populations of Drosophila melanogaster. *Genetics* **192**, 533-598.
Lelli, K. M., Slattery, M. and Mann, R. S. (2012). Disentangling the many layers of eukaryotic transcriptional regulation. *Annu Rev Genet* **46**, 43-68.
Lenz, O., Xiong, J., Nelson, M. D., Raizen, D. M. and Williams, J. A. (2015). FMRFamide signaling promotes stress-induced sleep in Drosophila. *Brain Behav Immun* **47**, 141-148.
Lewis, D. A. and Levitt, P. (2002). Schizophrenia as a disorder of neurodevelopment. *Annu Rev Neurosci* **25**, 409-432.
Li, X., Chen, Z. and Desplan, C. (2013). Temporal patterning of neural progenitors in Drosophila. *Curr Top Dev Biol* **105**, 69-96.

Li, X., Nie, S., Chang, C., Qiu, T. and Cao, X. (2006). Smads oppose Hox transcriptional activities. *Exp Cell Res* **312**, 854-864.

Li, X. and Noll, M. (1994). Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the Drosophila embryo. *EMBO J* **13**, 400-406.

Liang, H. L., Xu, M., Chuang, Y. C. and Rushlow, C. (2012). Response to the BMP gradient requires highly combinatorial inputs from multiple patterning systems in the Drosophila embryo. *Development*.

Lin, M. C., Park, J., Kirov, N. and Rushlow, C. (2006). Threshold response of C15 to the Dpp gradient in Drosophila is established by the cumulative effect of Smad and Zen activators and negative cues. *Development* **133**, 4805-4813.

Liu, A. and Niswander, L. A. (2005). Bone morphogenetic protein signalling and vertebrate nervous system development. *Nature reviews. Neuroscience* **6**, 945-954.

Lodish, H. F. (2004). *Molecular cell biology* (5th edn). New York: W. H. Freeman.

Lopez-Coviella, I., Follettie, M. T., Mellott, T. J., Kovacheva, V. P., Slack, B. E., Diesl, V., Berse, B., Thies, R. S. and Blusztajn, J. K. (2005). Bone morphogenetic protein 9 induces the transcriptome of basal forebrain cholinergic neurons. *Proc Natl Acad Sci U S A* **102**, 6984-6989.

Lotz, K., Pyrowolakis, G. and Jentsch, S. (2004). BRUCE, a giant E2/E3 ubiquitin ligase and inhibitor of apoptosis protein of the trans-Golgi network, is required for normal placenta development and mouse survival. *Molecular and cellular biology* **24**, 9339-9350.

Ludwig, M. Z., Patel, N. H. and Kreitman, M. (1998). Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change. *Development* **125**, 949-958.

Majdazari, A., Stubbusch, J., Muller, C. M., Hennchen, M., Weber, M., Deng, C. X., Mishina, Y., Schutz, G., Deller, T. and Rohrer, H. (2013). Dendrite complexity of sympathetic neurons is controlled during postnatal development by BMP signaling. *J Neurosci* **33**, 15132-15144.

Malik, S. and Roeder, R. G. (2005). Dynamic regulation of pol II transcription by the mammalian Mediator complex. *Trends Biochem Sci* **30**, 256-263.

Mardon, G., Solomon, N. M. and Rubin, G. M. (1994). dachshund encodes a nuclear protein required for normal eye and leg development in Drosophila. *Development* **120**, 3473-3486.

Marques, G. (2005). Morphogens and synaptogenesis in Drosophila. *J Neurobiol* **64**, 417-434.

Marques, G., Bao, H., Haerry, T. E., Shimell, M. J., Duchek, P., Zhang, B. and O'Connor, M. B. (2002). The Drosophila BMP type II receptor Wishful Thinking regulates neuromuscular synapse morphology and function. *Neuron* **33**, 529-543.

Marques, G., Haerry, T. E., Crotty, M. L., Xue, M., Zhang, B. and O'Connor, M. B. (2003). Retrograde Gbb signaling through the Bmp type 2 receptor wishful thinking regulates systemic FMRFa expression in Drosophila. *Development* **130**, 5457-5470.

Marty, T., Muller, B., Basler, K. and Affolter, M. (2000). Schnurri mediates Dpp-dependent repression of brinker transcription. *Nat Cell Biol* **2**, 745-749.

Marty, T., Vigano, M. A., Ribeiro, C., Nussbaumer, U., Grieder, N. C. and Affolter, M. (2001). A HOX complex, a repressor element and a 50 bp sequence confer regional specificity to a DPP-responsive enhancer. *Development* **128**, 2833-2845.

Massague, J. and Chen, Y. G. (2000). Controlling TGF-beta signaling. *Genes & development* **14**, 627-644.

Massague, J., Seoane, J. and Wotton, D. (2005). Smad transcription factors. *Genes & development* **19**, 2783-2810.

Maurange, C., Cheng, L. and Gould, A. P. (2008). Temporal transcription factors and their targets schedule the end of neural proliferation in Drosophila. *Cell* **133**, 891-902.

McCabe, B. D., Hom, S., Aberle, H., Fetter, R. D., Marques, G., Haerry, T. E., Wan, H., O'Connor, M. B., Goodman, C. S. and Haghighi, A. P. (2004). Highwire regulates presynaptic BMP signaling essential for synaptic growth. *Neuron* **41**, 891-905.

McCabe, B. D., Marques, G., Haghighi, A. P., Fetter, R. D., Crotty, M. L., Haerry, T. E., Goodman, C. S. and O'Connor, M. B. (2003). The BMP homolog Gbb provides a retrograde signal that regulates synaptic growth at the Drosophila neuromuscular junction. *Neuron* **39**, 241-254.

McVey, M. (2010). In vivo analysis of Drosophila BLM helicase function during DNA double-strand gap repair. *Methods Mol Biol* **587**, 185-194.

McVey, M., Larocque, J. R., Adams, M. D. and Sekelsky, J. J. (2004). Formation of deletions during double-strand break repair in Drosophila DmBlm mutants occurs after strand invasion. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 15694-15699.

Merika, M., Williams, A. J., Chen, G., Collins, T. and Thanos, D. (1998). Recruitment of CBP/p300 by the IFN beta enhanceosome is required for synergistic activation of transcription. *Molecular cell* **1**, 277-287.

Merino, C., Penney, J., Gonzalez, M., Tsurudome, K., Moujahidine, M., O'Connor, M. B., Verheyen, E. M. and Haghighi, P. (2009). Nemo kinase interacts with Mad to coordinate synaptic growth at the Drosophila neuromuscular junction. *J Cell Biol* **185**, 713-725.

Metaxakis, A., Oehler, S., Klinakis, A. and Savakis, C. (2005). Minos as a genetic and genomic tool in Drosophila melanogaster. *Genetics* **171**, 571-581.

Mettler, U., Vogler, G. and Urban, J. (2006). Timing of identity: spatiotemporal regulation of hunchback in neuroblast lineages of Drosophila by Seven-up and Prospero. *Development* **133**, 429-437.

Middleton, C. A., Nongthomba, U., Parry, K., Sweeney, S. T., Sparrow, J. C. and Elliott, C. J. (2006). Neuromuscular organization and aminergic modulation of contractions in the Drosophila ovary. *BMC Biol* **4**, 17.

Miguel-Aliaga, I., Allan, D. W. and Thor, S. (2004). Independent roles of the dachshund and eyes absent genes in BMP signaling, axon pathfinding and neuronal specification. *Development* **131**, 5837-5848.

Miguel-Aliaga, I., Thor, S. and Gould, A. P. (2008). Postmitotic specification of Drosophila insulinergic neurons from pioneer neurons. *PLoS Biol* **6**, e58.

Mills, J. C. and Taghert, P. H. (2012). Scaling factors: transcription factors regulating subcellular domains. *Bioessays* **34**, 10-16.

Mitchell, K. J., Doyle, J. L., Serafini, T., Kennedy, T. E., Tessier-Lavigne, M., Goodman, C. S. and Dickson, B. J. (1996). Genetic analysis of Netrin genes in Drosophila: Netrins guide CNS commissural axons and peripheral motor axons. *Neuron* **17**, 203-215.

Mlodzik, M., Hiromi, Y., Weber, U., Goodman, C. S. and Rubin, G. M. (1990). The Drosophila seven-up gene, a member of the steroid receptor gene superfamily, controls photoreceptor cell fates. *Cell* **60**, 211-224.

Molenaar, M., van de Wetering, M., Oosterwegel, M., Peterson-Maduro, J., Godsave, S., Korinek, V., Roose, J., Destree, O. and Clevers, H. (1996). XTcf-3 transcription factor mediates beta-catenin-induced axis formation in Xenopus embryos. *Cell* **86**, 391-399.

Monastirioti, M., Linn, C. E., Jr. and White, K. (1996). Characterization of Drosophila tyramine beta-hydroxylase gene and isolation of mutant flies lacking octopamine. *J Neurosci* **16**, 3900-3911.

Morikawa, M., Koinuma, D., Tsutsumi, S., Vasilaki, E., Kanki, Y., Heldin, C. H., Aburatani, H. and Miyazono, K. (2011). ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif. *Nucleic Acids Res* **39**, 8712-8727.

Mouritzen, P., Nielsen, A. T., Pfundheller, H. M., Choleva, Y., Kongsbak, L. and Moller, S. (2003). Single nucleotide polymorphism genotyping using locked nucleic acid (LNA). *Expert review of molecular diagnostics* **3**, 27-38.

Mullen, A. C., Orlando, D. A., Newman, J. J., Loven, J., Kumar, R. M., Bilodeau, S., Reddy, J., Guenther, M. G., DeKoter, R. P. and Young, R. A. (2011). Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* **147**, 565-576.

Muller, B., Hartmann, B., Pyrowolakis, G., Affolter, M. and Basler, K. (2003). Conversion of an extracellular Dpp/BMP morphogen gradient into an inverse transcriptional gradient. *Cell* **113**, 221-233.

Naka, H., Nakamura, S., Shimazaki, T. and Okano, H. (2008). Requirement for COUP-TFI and II in the temporal specification of neural stem cells in CNS development. *Nat Neurosci* **11**, 1014-1023.

Nguyen, G. H., Dexheimer, T. S., Rosenthal, A. S., Chu, W. K., Singh, D. K., Mosedale, G., Bachrati, C. Z., Schultz, L., Sakurai, M., Savitsky, P., et al. (2013). A small molecule inhibitor of the BLM helicase modulates chromosome stability in human cells. *Chem Biol* **20**, 55-62.

Nishi, R. (2003). Target-mediated control of neural differentiation. *Prog Neurobiol* **69**, 213-227.

Noyes, M. B., Christensen, R. G., Wakabayashi, A., Stormo, G. D., Brodsky, M. H. and Wolfe, S. A. (2008). Analysis of homeodomain specificities allows the family-wide prediction of preferred recognition sites. *Cell* **133**, 1277-1289.

O'Connor, M. B., Umulis, D., Othmer, H. G. and Blair, S. S. (2006). Shaping BMP morphogen gradients in the Drosophila embryo and pupal wing. *Development* **133**, 183-193.

O'Keefe, D. D., Thor, S. and Thomas, J. B. (1998). Function and specificity of LIM domains in Drosophila nervous system and wing development. *Development* **125**, 3915-3923.

Okulski, H., Druck, B., Bhalerao, S. and Ringrose, L. (2011). Quantitative analysis of polycomb response elements (PREs) at identical genomic locations distinguishes contributions of PRE sequence and genomic environment. *Epigenetics & chromatin* **4**, 4.

Ormerod, K. G., Krans, J. L. and Mercier, A. J. (2015). Cell-selective modulation of the Drosophila neuromuscular system by a neuropeptide. *J Neurophysiol* **113**, 1631-1643.

Osterwalder, T., Yoon, K. S., White, B. H. and Keshishian, H. (2001). A conditional tissue-specific transgene expression system using inducible GAL4. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 12596-12601.

Ou, H. L. (2013). Gene knockout by inducing P-element transposition in Drosophila. *Genetics and molecular research : GMR* **12**, 2852-2857.

Overballe-Petersen, S., Harms, K., Orlando, L. A., Mayar, J. V., Rasmussen, S., Dahl, T. W., Rosing, M. T., Poole, A. M., Sicheritz-Ponten, T., Brunak, S., et al. (2013). Bacterial natural transformation by highly fragmented and damaged DNA. *Proc Natl Acad Sci U S A* **110**, 19860-19865.

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. *Current biology : CB* **21**, 1515-1524.

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

Parks, A. L., Cook, K. R., Belvin, M., Dompe, N. A., Fawcett, R., Huppert, K., Tan, L. R., Winter, C. G., Bogart, K. P., Deal, J. E., et al. (2004). Systematic generation of high-resolution deletion coverage of the Drosophila melanogaster genome. *Nat Genet* **36**, 288-292.

Parrish, J. Z., Kim, M. D., Jan, L. Y. and Jan, Y. N. (2006). Genome-wide analyses identify transcription factors required for proper morphogenesis of Drosophila sensory neuron dendrites. *Genes & development* **20**, 820-835.

Patel, T. D., Jackman, A., Rice, F. L., Kucera, J. and Snider, W. D. (2000). Development of sensory neurons in the absence of NGF/TrkA signaling in vivo. *Neuron* **25**, 345-357.

Patel, T. D., Kramer, I., Kucera, J., Niederkofler, V., Jessell, T. M., Arber, S. and Snider, W. D. (2003). Peripheral NT3 signaling is required for ETS protein expression and central patterning of proprioceptive sensory afferents. *Neuron* **38**, 403-416.

Pavelock, K. A., Girard, B. M., Schutz, K. C., Braas, K. M. and May, V. (2007). Bone morphogenetic protein down-regulation of neuronal pituitary adenylate cyclase-activating polypeptide and reciprocal effects on vasoactive intestinal peptide expression. *J Neurochem* **100**, 603-616.

Peterson, A. J. and O'Connor, M. B. (2014). Strategies for exploring TGF-beta signaling in Drosophila. *Methods* **68**, 183-193.

Pfeiffer, B. D., Truman, J. W. and Rubin, G. M. (2012). Using translational enhancers to increase transgene expression in Drosophila. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 6626-6631.

Pignoni, F., Hu, B., Zavitz, K. H., Xiao, J., Garrity, P. A. and Zipursky, S. L. (1997). The eye-specification proteins So and Eya form a complex and regulate multiple steps in Drosophila eye development. *Cell* **91**, 881-891.

Port, F., Chen, H. M., Lee, T. and Bullock, S. L. (2014). Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in Drosophila. *Proc Natl Acad Sci U S A* **111**, E2967-2976.

Portales-Casamar, E., Thongjuea, S., Kwon, A. T., Arenillas, D., Zhao, X., Valen, E., Yusuf, D., Lenhard, B., Wasserman, W. W. and Sandelin, A. (2010). JASPAR 2010: the greatly expanded open-access database of transcription factor binding profiles. *Nucleic Acids Res* **38**, D105-110.

Pyrowolakis, G., Hartmann, B., Muller, B., Basler, K. and Affolter, M. (2004). A simple molecular complex mediates widespread BMP-induced repression during Drosophila development. *Dev Cell* **7**, 229-240.

Qiu, Y., Cooney, A. J., Kuratani, S., DeMayo, F. J., Tsai, S. Y. and Tsai, M. J. (1994). Spatiotemporal expression patterns of chicken ovalbumin upstream promoter-transcription factors in the developing mouse central nervous system: evidence for a role in segmental patterning of the diencephalon. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 4451-4455.

Qiu, Y., Pereira, F. A., DeMayo, F. J., Lydon, J. P., Tsai, S. Y. and Tsai, M. J. (1997). Null mutation of mCOUP-TFI results in defects in morphogenesis of the glossopharyngeal ganglion, axonal projection, and arborization. *Genes & development* **11**, 1925-1937.

Raftery, L. A. and Sutherland, D. J. (1999). TGF-beta family signal transduction in Drosophila development: from Mad to Smads. *Dev Biol* **210**, 251-268.

Ramos, A. I. and Barolo, S. (2013). Low-affinity transcription factor binding sites shape morphogen responses and enhancer evolution. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **368**, 20130018.

Razin, S. V., Gavrilov, A. A., Ioudinkova, E. S. and Iarovaia, O. V. (2013). Communication of genome regulatory elements in a folded chromosome. *FEBS letters* **587**, 1840-1847.

Ren, X., Yang, Z., Xu, J., Sun, J., Mao, D., Hu, Y., Yang, S. J., Qiao, H. H., Wang, X., Hu, Q., et al. (2014). Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. *Cell reports* **9**, 1151-1162.

Rogulja, D. and Irvine, K. D. (2005). Regulation of cell proliferation by a morphogen gradient. *Cell* **123**, 449-461.

Ross, S. and Hill, C. S. (2008a). How the Smads regulate transcription. *The international journal of biochemistry & cell biology* **40**, 383-408.

---- (2008b). How the Smads regulate transcription. *Int J Biochem Cell Biol* **40**, 383-408.

Rushlow, C., Colosimo, P. F., Lin, M. C., Xu, M. and Kirov, N. (2001). Transcriptional regulation of the Drosophila gene zen by competing Smad and Brinker inputs. *Genes Dev* **15**, 340-351.

Ryder, E., Ashburner, M., Bautista-Llacer, R., Drummond, J., Webster, J., Johnson, G., Morley, T., Chan, Y. S., Blows, F., Coulson, D., et al. (2007). The DrosDel deletion collection: a Drosophila genomewide chromosomal deficiency resource. *Genetics* **177**, 615-629.

Salinas, P. C. (2005). Signaling at the vertebrate synapse: new roles for embryonic morphogens? *J Neurobiol* **64**, 435-445.

Saller, E. and Bienz, M. (2001). Direct competition between Brinker and Drosophila Mad in Dpp target gene transcription. *EMBO Rep* **2**, 298-305.

Satoh, S., Tang, K., Iida, A., Inoue, M., Kodama, T., Tsai, S. Y., Tsai, M. J., Furuta, Y. and Watanabe, S. (2009). The spatial patterning of mouse cone opsin expression is regulated by bone morphogenetic protein signaling through downstream effector COUP-TF nuclear receptors. *J Neurosci* **29**, 12401-12411.

Savage, C., Das, P., Finelli, A. L., Townsend, S. R., Sun, C. Y., Baird, S. E. and Padgett, R. W. (1996). Caenorhabditis elegans genes sma-2, sma-3, and sma-4 define a conserved family of transforming growth factor beta pathway components. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 790-794.

Sawala, A., Sutcliffe, C. and Ashe, H. L. (2012). Multistep molecular mechanism for bone morphogenetic protein extracellular transport in the Drosophila embryo. *Proc Natl Acad Sci U S A* **109**, 11222-11227.

Schmid, A., Chiba, A. and Doe, C. Q. (1999). Clonal analysis of Drosophila embryonic neuroblasts: neural cell types, axon projections and muscle targets. *Development* **126**, 4653-4689.

Schmierer, B. and Hill, C. S. (2007). TGFbeta-SMAD signal transduction: molecular specificity and functional flexibility. *Nat Rev Mol Cell Biol* **8**, 970-982.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* **9**, 671-675.

Schneider, L. E., Roberts, M. S. and Taghert, P. H. (1993). Cell type-specific transcriptional regulation of the Drosophila FMRFamide neuropeptide gene. *Neuron* **10**, 279-291.

Schwank, G., Yang, S. F., Restrepo, S. and Basler, K. (2012). Comment on "Dynamics of dpp signaling and proliferation control". *Science* **335**, 401; author reply 401.

Seibert, J., Volland, D. and Urbach, R. (2009). Ems and Nkx6 are central regulators in dorsoventral patterning of the Drosophila brain. *Development* **136**, 3937-3947.

Sekelsky, J. J., Newfeld, S. J., Raftery, L. A., Chartoff, E. H. and Gelbart, W. M. (1995). Genetic characterization and cloning of mothers against dpp, a gene required for decapentaplegic function in Drosophila melanogaster. *Genetics* **139**, 1347-1358.

Seoane, J., Le, H. V., Shen, L., Anderson, S. A. and Massague, J. (2004). Integration of Smad and forkhead pathways in the control of neuroepithelial and glioblastoma cell proliferation. *Cell* **117**, 211-223.

Shazman, S., Lee, H., Socol, Y., Mann, R. S. and Honig, B. (2014). OnTheFly: a database of Drosophila melanogaster transcription factors and their binding sites. *Nucleic Acids Res* **42**, D167-171.

Shen, W. and Mardon, G. (1997). Ectopic eye development in Drosophila induced by directed dachshund expression. *Development* **124**, 45-52.

Shi, X., Yang, X., Chen, D., Chang, Z. and Cao, X. (1999). Smad1 interacts with homeobox DNA-binding proteins in bone morphogenetic protein signaling. *J Biol Chem* **274**, 13711-13717.

Shi, Y. and Massague, J. (2003). Mechanisms of TGF-beta signaling from cell membrane to the nucleus. *Cell* **113**, 685-700.

Sivasankaran, R., Vigano, M. A., Muller, B., Affolter, M. and Basler, K. (2000). Direct transcriptional control of the Dpp target omb by the DNA binding protein Brinker. *EMBO J* **19**, 6162-6172.

Skeath, J. B. and Thor, S. (2003). Genetic control of Drosophila nerve cord development. *Curr Opin Neurobiol* **13**, 8-15.

Small, S., Blair, A. and Levine, M. (1992). Regulation of even-skipped stripe 2 in the Drosophila embryo. *EMBO J* **11**, 4047-4057.

Song, N. N., Xiu, J. B., Huang, Y., Chen, J. Y., Zhang, L., Gutknecht, L., Lesch, K. P., Li, H. and Ding, Y. Q. (2011). Adult raphe-specific deletion of Lmx1b leads to central serotonin deficiency. *PLoS One* **6**, e15998.

Spana, E. P. and Doe, C. Q. (1996). Numb antagonizes Notch signaling to specify sibling neuron cell fates. *Neuron* **17**, 21-26.

Stanke, M., Duong, C. V., Pape, M., Geissen, M., Burbach, G., Deller, T., Gascan, H., Otto, C., Parlato, R., Schutz, G., et al. (2006). Target-dependent specification of the neurotransmitter

phenotype: cholinergic differentiation of sympathetic neurons is mediated in vivo by gp 130 signaling. *Development* **133**, 141-150.

Stark, A., Lin, M. F., Kheradpour, P., Pedersen, J. S., Parts, L., Carlson, J. W., Crosby, M. A., Rasmussen, M. D., Roy, S., Deoras, A. N., et al. (2007). Discovery of functional elements in 12 Drosophila genomes using evolutionary signatures. *Nature* **450**, 219-232.

Suszko, M. I., Antenos, M., Balkin, D. M. and Woodruff, T. K. (2008). Smad3 and Pitx2 cooperate in stimulation of FSHbeta gene transcription. *Mol Cell Endocrinol* **281**, 27-36.

Szutorisz, H., Dillon, N. and Tora, L. (2005). The role of enhancers as centres for general transcription factor recruitment. *Trends Biochem Sci* **30**, 593-599.

Takaesu, N. T., Herbig, E., Zhitomersky, D., O'Connor, M. B. and Newfeld, S. J. (2005). DNA-binding domain mutations in SMAD genes yield dominant-negative proteins or a neomorphic protein that can activate WG target genes in Drosophila. *Development* **132**, 4883-4894.

Tear, G., Seeger, M. and Goodman, C. S. (1993). To cross or not to cross: a genetic analysis of guidance at the midline. *Perspectives on developmental neurobiology* **1**, 183-194.

Teleman, A. A. and Cohen, S. M. (2000). Dpp gradient formation in the Drosophila wing imaginal disc. *Cell* **103**, 971-980.

Therianos, S., Leuzinger, S., Hirth, F., Goodman, C. S. and Reichert, H. (1995). Embryonic development of the Drosophila brain: formation of commissural and descending pathways. *Development* **121**, 3849-3860.

Tran, K. D. and Doe, C. Q. (2008). Pdm and Castor close successive temporal identity windows in the NB3-1 lineage. *Development* **135**, 3491-3499.

Tripodi, M., Filosa, A., Armentano, M. and Studer, M. (2004). The COUP-TF nuclear receptors regulate cell migration in the mammalian basal forebrain. *Development* **131**, 6119-6129.

Trompouki, E., Bowman, T. V., Lawton, L. N., Fan, Z. P., Wu, D. C., DiBiase, A., Martin, C. S., Cech, J. N., Sessa, A. K., Leblanc, J. L., et al. (2011). Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* **147**, 577-589.

Uhler, J., Garbern, J., Yang, L., Kamholz, J. and Mellerick, D. M. (2002). Nk6, a novel Drosophila homeobox gene regulated by vnd. *Mech Dev* **116**, 105-116.

Ulvklo, C., MacDonald, R., Bivik, C., Baumgardt, M., Karlsson, D. and Thor, S. (2012). Control of neuronal cell fate and number by integration of distinct daughter cell proliferation modes with temporal progression. *Development* **139**, 678-689.

Umulis, D., O'Connor, M. B. and Blair, S. S. (2009). The extracellular regulation of bone morphogenetic protein signaling. *Development* **136**, 3715-3728.

Veverytsa, L. and Allan, D. W. (2011). Retrograde BMP signaling controls Drosophila behavior through regulation of a peptide hormone battery. *Development* **138**, 3147-3157.

---- (2012). Temporally tuned neuronal differentiation supports the functional remodeling of a neuronal network in Drosophila. *Proc Natl Acad Sci U S A*.

---- (2013). Subtype-specific neuronal remodeling during Drosophila metamorphosis. *Fly (Austin)* **7**, 78-86.

Walker, J., Curtis, V. and Murray, R. M. (2002). Schizophrenia and bipolar disorder: similarities in pathogenic mechanisms but differences in neurodevelopment. *Int Clin Psychopharmacol* **17 Suppl 3**, S11-19.

Walsh, C. M. and Carroll, S. B. (2007). Collaboration between Smads and a Hox protein in target gene repression. *Development* **134**, 3585-3592.

Wang, R. N., Green, J., Wang, Z., Deng, Y., Qiao, M., Peabody, M., Zhang, Q., Ye, J., Yan, Z., Denduluri, S., et al. (2014). Bone Morphogenetic Protein (BMP) signaling in development and human diseases. *Genes Dis* **1**, 87-105.

Watanabe, K., Kamiya, D., Nishiyama, A., Katayama, T., Nozaki, S., Kawasaki, H., Watanabe, Y., Mizuseki, K. and Sasai, Y. (2005). Directed differentiation of telencephalic precursors from embryonic stem cells. *Nat Neurosci* **8**, 288-296.

Weiss, A., Charbonnier, E., Ellertsdottir, E., Tsirigos, A., Wolf, C., Schuh, R., Pyrowolakis, G. and Affolter, M. (2010). A conserved activation element in BMP signaling during Drosophila development. *Nat Struct Mol Biol* **17**, 69-76.

Weiszmann, R., Hammonds, A. S. and Celniker, S. E. (2009). Determination of gene expression patterns using high-throughput RNA in situ hybridization to whole-mount Drosophila embryos. *Nat Protoc* **4**, 605-618.

Wieschaus, E., Nusslein-Volhard, C. and Kluding, H. (1984). Kruppel, a gene whose activity is required early in the zygotic genome for normal embryonic segmentation. *Dev Biol* **104**, 172-186.

Winter, S. E. and Campbell, G. (2004). Repression of Dpp targets in the Drosophila wing by Brinker. *Development* **131**, 6071-6081.

Witsell, A., Kane, D. P., Rubin, S. and McVey, M. (2009). Removal of the bloom syndrome DNA helicase extends the utility of imprecise transposon excision for making null mutations in Drosophila. *Genetics* **183**, 1187-1193.

Xiao, L., Michalski, N., Kronander, E., Gjoni, E., Genoud, C., Knott, G. and Schneggenburger, R. (2013). BMP signaling specifies the development of a large and fast CNS synapse. *Nat Neurosci* **16**, 856-864.

Xu, M., Kirov, N. and Rushlow, C. (2005). Peak levels of BMP in the Drosophila embryo control target genes by a feed-forward mechanism. *Development* **132**, 1637-1647.

Xu, P. and Hall, A. K. (2007). Activin acts with nerve growth factor to regulate calcitonin gene-related peptide mRNA in sensory neurons. *Neuroscience* **150**, 665-674.

Xu, X., Yin, Z., Hudson, J. B., Ferguson, E. L. and Frasch, M. (1998). Smad proteins act in combination with synergistic and antagonistic regulators to target Dpp responses to the Drosophila mesoderm. *Genes Dev* **12**, 2354-2370.

Xue, Z., Wu, M., Wen, K., Ren, M., Long, L., Zhang, X. and Gao, G. (2014). CRISPR/Cas9 mediates efficient conditional mutagenesis in Drosophila. *G3* **4**, 2167-2173.

Yamaguchi, H., Zhou, C., Lin, S. C., Durand, B., Tsai, S. Y. and Tsai, M. J. (2004). The nuclear orphan receptor COUP-TFI is important for differentiation of oligodendrocytes. *Dev Biol* **266**, 238-251.

Yan, C. H., Levesque, M., Claxton, S., Johnson, R. L. and Ang, S. L. (2011). Lmx1a and lmx1b function cooperatively to regulate proliferation, specification, and differentiation of midbrain dopaminergic progenitors. *J Neurosci* **31**, 12413-12425.

Yao, L. C., Blitz, I. L., Peiffer, D. A., Phin, S., Wang, Y., Ogata, S., Cho, K. W., Arora, K. and Warrior, R. (2006). Schnurri transcription factors from Drosophila and vertebrates can mediate Bmp signaling through a phylogenetically conserved mechanism. *Development* **133**, 4025-4034.

Yavatkar, A. S., Lin, Y., Ross, J., Fann, Y., Brody, T. and Odenwald, W. F. (2008). Rapid detection and curation of conserved DNA via enhanced-BLAT and EvoPrinterHD analysis. *BMC Genomics* **9**, 106.

Ying, Q. L., Nichols, J., Chambers, I. and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115**, 281-292.

Yu, K., Sturtevant, M. A., Biehs, B., Francois, V., Padgett, R. W., Blackman, R. K. and Bier, E. (1996). The Drosophila decapentaplegic and short gastrulation genes function antagonistically during adult wing vein development. *Development* **122**, 4033-4044.

Zawel, L., Dai, J. L., Buckhaults, P., Zhou, S., Kinzler, K. W., Vogelstein, B. and Kern, S. E. (1998). Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol Cell* **1**, 611-617.

Zhang, H., Levine, M. and Ashe, H. L. (2001). Brinker is a sequence-specific transcriptional repressor in the Drosophila embryo. *Genes Dev* **15**, 261-266.

Zhang, T., Branch, A. and Shen, P. (2013). Octopamine-mediated circuit mechanism underlying controlled appetite for palatable food in Drosophila. *Proc Natl Acad Sci U S A* **110**, 15431-15436.

Zhou, B., Chen, L., Wu, X., Wang, J., Yin, Y. and Zhu, G. (2008). MH1 domain of SMAD4 binds N-terminal residues of the homeodomain of Hoxc9. *Biochim Biophys Acta* **1784**, 747-752.

Zhu, B., Pennack, J. A., McQuilton, P., Forero, M. G., Mizuguchi, K., Sutcliffe, B., Gu, C. J., Fenton, J. C. and Hidalgo, A. (2008). Drosophila neurotrophins reveal a common mechanism for nervous system formation. *PLoS Biol* **6**, e284.

Zhu, Y., Sun, L., Chen, Z., Whitaker, J. W., Wang, T. and Wang, W. (2013). Predicting enhancer transcription and activity from chromatin modifications. *Nucleic Acids Res* **41**, 10032-10043.

APPENDIX A- MUTAGENESIS AND PRELIMINARY CHARACTERIZATION OF TWO *DROSOPHILA* LMX1 ORTHOLOGUES; *CG4328* AND *CG32105*.

A.1 Introduction

Transcription factors act in complex cascades throughout neurogenesis to combinatorially direct the identity and function of a vast diversity of specialized neurons of the central nervous system (CNS). Understanding the transcriptional mechanisms that control specification and differentiation is a major goal of developmental biology. Thus, identifying the transcription factors that determine neuronal identity has long been a focus of many researchers in attempting to understand development of the nervous system. In the primary body of this thesis, I present work that investigates the role of the LIM-Homeodomain transcription factor (LIM-HD TF), Apterous, in regulating the differentiation of FMRFa-expressing Tv4 neurons. LIM-HD TFs have proven to function as critical determinants of neuronal subtype identity throughout metazoans (Curtiss and Heilig, 1998; Hobert and Westphal, 2000).

The LIM-HD TF family is remarkably well-conserved and subdivides into the *apterous*, *Lhx6/7*, Islet, Lim3, Lin-11 and LMX groups (Hobert and Westphal, 2000) (See **Figure S 1**). The *Drosophila* members of these groups are well studied, except for the two *Drosophila* members of the LMX group, *CG4328* and *CG32105*, for which mutants had not been isolated and studied. To start to characterize the function of these two genes in *Drosophila*, I generated mutants of *CG4328* and *CG32105*. To date, I have identified an egg-laying phenotype in *CG4328* mutants that appears to be primarily due to aberrant innervation of octopaminergic neurons and did a direct comparison of two strategies for mobilizing transposable elements to delete the coding sequence of *CG32105*.

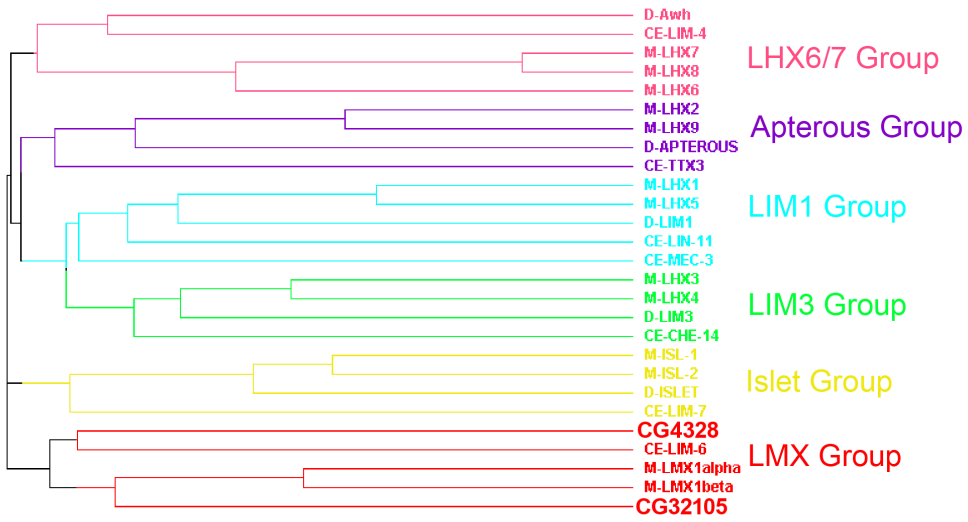


Figure S 1. Relationship of the *CG4328* and *CG32105* protein sequence to other known LIM-Homeodomain homologues from representative model organisms

(A) Major phylogenetic groupings of LIM-Homeodomain transcription factors based on percent sequence similarity and conventional naming from the literature. Prefixes "CE", "D", and "M" indicate *C. elegans*, *D. melanogaster* and *M. musculus*, respectively. All known LIM-Homeodomain proteins from each species are represented.

Lmx1 genes are known to play key roles in vertebrate development. Autosomal dominant loss of function mutations of *lmx1b* causes Nail-patella syndrome which is characterized by ventralization of dorsal limb surfaces, improper joint patterning and defective glomeruli in the kidney (Dai et al., 2009). When LIM-HD TFs are expressed in the CNS of animals, they have typically been found to regulate key aspect of both specification and post-mitotic neuronal differentiation, such as axon guidance and neurotransmitter biosynthesis; two developmental processes absolutely required for neuronal communication. Failure to properly specify, differentiate or maintain neuronal identity has been implicated in degenerative disorders and mental illness, where death or dysfunction of neurons occurs (De Vos et al., 2008; Gunawardena and Goldstein, 2004; Lewis and Levitt, 2002; Walker et al., 2002). Vertebrate midbrain serotonergic neurons express *lmx1b* (Demarque and Spitzer, 2010; Yan et al., 2011) and absolutely require the activity of *lmx1b* for their specification (Ding et al., 2003) and appropriate differentiation (Demarque and Spitzer, 2010) as well as maintenance of neurotransmitter identity in the adult (Song et al., 2011). Notably, polymorphisms in intronic regions of the human *lmx1b* locus have been correlated with schizophrenia, a disease that has been attributed in part to improper dopaminergic and serotonergic signaling (Bergman et al., 2010).

High throughput expression *in-situ* hybridization studies by the Berkeley *Drosophila* Genome Project showed that *CG4328* and *CG32105* are almost exclusively expressed in the VNC of the developing *Drosophila* embryo. High throughput RNA-seq studies from the modENCODE project show that the transcripts appear to be highly restricted to the CNS throughout life for both *CG4328* and *CG32105* (Hammonds et al., 2013). High throughout RNAi studies identified *CG4328* as being involved in sensory neuron morphogenesis (Parrish et al., 2006). The expression patterns in the embryo of both *CG4328* and *CG32105* appeared to be non-

neuroblast like and elaborated over time, therefore we hypothesized that one or both of these genes would be involved in the differentiation of post-mitotic neurons in a manner similar to *apterous* in the Tv neurons.

This work has been conducted as a time-limited 'back-burner' project alongside the major work of the thesis presented in Chapters 2 and 3. The work here is preliminary data from this ongoing project.

A.2 Materials and Methods

A.2.1 Fly Strains

Strains used for *CG4328* experiments: *w;;pBac{WH}f00253/TM6B*, *yw;;P{XP}d00806* contained the FRT sites used to delete *CG4328* and were from the Exelixis collection at Harvard Medical School. Flippase was provided by *P{hsFLP}1* ,y^1,w^{1118}; ; *Dr /TM3,*ry^*Sb^1 (Golic, 1991). Lines were balanced with $w^{1118;}$; *Dr /TM3,Sb* and then $w^{1118;}$; *Dr /TM3,Sb,Ser,twiGAL4,UAS-2xEGFP*. *w; tdc2-GAL4* was a gift from the Goodwin lab (Cole et al., 2005). *UAS-CD4::tdGFP* and *UAS-nls.EGFP* were from the Bloomington stock center (Bloomington, IN). Mutants were kept over *CyO,Act-GFP TM3,Ser,Act-GFP* or *CyO,twiGAL4,UAS-2xEGFP* or *TM3,Sb,Ser,twiGAL4,UAS-2xEGFP*.

Strains used for *CG31205* experiments: w^{1118};; *Mi{ET1}CG32105,* w^{1118}; Blm^{N1}*/TM3,* Sb^1 , $w^{1118;}$; *Dr /TM3,Sb,* w^{1118};; *Sco/SM6a, P{hsILMiT}2.4, w;;Sp/CyO, w;* Dr/*SM6a-TM6B,Tb* all from the Bloomington Stock Center (Bloomington, IN). Flies were maintained at 25°C, 70% humidity**.**

A.2.2 Heat Shock Protocol

To induce either Flippase for the Flip/FRT deletion or the Transposase to mobilize the *Minos* element flies were crossed and parents allowed to lay eggs for three days in normal vials before being flipped in to another vial. Vials were then heat shocked for one hour at 37 ^{o}C in each day for three consecutive days using a water bath. Between each heat shock vials were allowed to recover at 25^{o}C.

A.2.3 Primers Used to Screen Flip/Frt Deletion of *CG4328*

Genomic DNA was extracted using the "Quick Prep" method (Huang et al., 2009). PCR performed with NEB Taq polymerase according to the manufacturers recommended protocol (New England Biolabs, Ipswich, MA.) Amplicons were resolved on a 1% agarose gel using standard electrophoresis techniques with TAE buffer. Sequencing primers were CG4328-seq.L#1- ATGCACAATTGCTTCATTTT; CG4328-seq.R#1-TGCTGCGATATGAACTCTTT

A.2.4 Primers Used to Screen *CG32105 Minos* Excision Events

PCR was conducted with iPoof DNA polymerase according to the manufacturer's recommended protocol (Biorad, Philadelphia, PA.)

Initial screen:
F_CG32105.small- AGCGGAGTTACACTGGAACAAA
R_CG32105.small- GGTATGGGTCGCATTACCAG

Sequencing:
CG32105_B.F- TTTGGCTTAAGTTCGGTGGTAT
CG32105_B.R- ATCTCCTTCTCCAGATCGTGAC
CG32105_C.F- ACACTTTGCAGAGGTCGAG
CG31205_C.R- GTGCCACATATGCTGTTTAAT
CG32105_D.F-CACCGTCAAATTAGCCCACT
CG32105_D.R-TCGTCATCCAACTGGTCAAA

A.2.5 RNA *In-Situ* Hybridization

Alkaline Phosphatase RNA *In-Situ* Hybridization was performed as previously described (Weiszmann et al., 2009). A full length cDNA clone IP01440 of *CG4328* was used as a template from the BDGP to produce a Digoxigenin labeled RNA probe. A fluorescent variation of this protocol was developed by using Sheep Anti-Digoxigenin-POD at 1:5000 (Roche) and the Cy3 TSA-Plus kit from PerkinElmer (Waltham, MA) was used to generate signal under manufacturer recommended conditions. Anti-GFP staining was performed after TSA development.

A.2.6 Egg Laying Assays

Virgin females were collected and aged to 3-5 days post-eclosion in vials only with other females. Vials of 3 virgin females were crossed to 5 week old w^{1118} males for 24 hours. After mating, groups of three mated females were removed and placed in cages with grape juice-agar media plates and a very thin layer yeast paste for egg collection. Flies were not exposed to CO_2 when plates were switched. Plates were changed every 24 hours and the number of eggs counted on each plate. Plates were stored at 4^oC before counting to prevent hatching of the eggs. Egg collections were performed for seven days. Data is represented as the average of eggs, per fly, per day.

A.2.7 Oviduct Dissections and Immunohistochemistry

Primary antibodies: Rat anti-TβH (1:50, M. Monastirioti); Rabbit anti-TDC2 (S. Goodwin); Mouse anti-GFP (1:100 Abcam) Mouse anti-Discs large (1:50; clone 4F3 DSHB.) Secondary antibodies: were Donkey anti-Rat Cy3, Donkey anti-Rabbit Cy5, Donkey anti-Mouse 488. Standard protocols were used for fixation (Eade and Allan, 2009).

A.2.8 Imaging and Data Analysis

Images were acquired with an Olympus FV1000 confocal microscope with settings that avoided pixel intensity saturation. Representative images of Tv-neurons being compared in figures were contrast enhanced together in Adobe Photoshop CS5 (Adobe Systems, Mountain View, CA). All statistical analysis and graphing were performed using Prism 5 (GraphPad Software, San Diego, CA). Student's two-tailed *t*-test was used to compare groups. Differences between groups were considered statistically significant when $p<0.05$. Data presented as mean ±SEM

A.2.8.1 Oviduct Imaging and Image Analysis

The common oviduct was imaged through its entire diameter. Maximum sum intensity projections were the generated to give a two dimensional representation of the oviduct. *Tdc2-Gal4 ->UAS-CD4::tdGFP* labeled axons were traced using the line trace function in ImageJ (Schneider et al., 2012). Anti-Discs large (DLG) staining of the post-synaptic specialization of the NMJ was counted using the region of interest tool. Total 2-dimension area of the oviduct was measured by tracing the outside of the tissue. Axon length is represented as linear-axon length in μm per μm^2 of Z-projected projected oviduct. The number of DLG puncta were similarly divided by the area of the oviduct in μm^2.

A.3 Results

A.3.1 Characterization of Early *CG4328* Expression Patterns

Publicly available high throughput expression pattern data suggested that *CG4328* was expressed in the VNC of the developing *Drosophila* embryo. To confirm the fidelity of these data, we initially characterized the expression of *CG4328* through mid-embryonic stages during neurogenesis and neuronal differentiation (Stages 13 and Sate 16 respectively). Alkaline phosphatase staining was used to replicate conditions used in the BDGP database and then modified to utilize fluorescence labeling. *CG4328* expression was not detectable in the embryo before ~10 hours after egg laying (data not shown). Expression was restricted to the CNS with expression appearing in pairs of cells along the dorsal midline which then expands to more lateral neuronal populations as development proceeds (See **Figure S 2**).

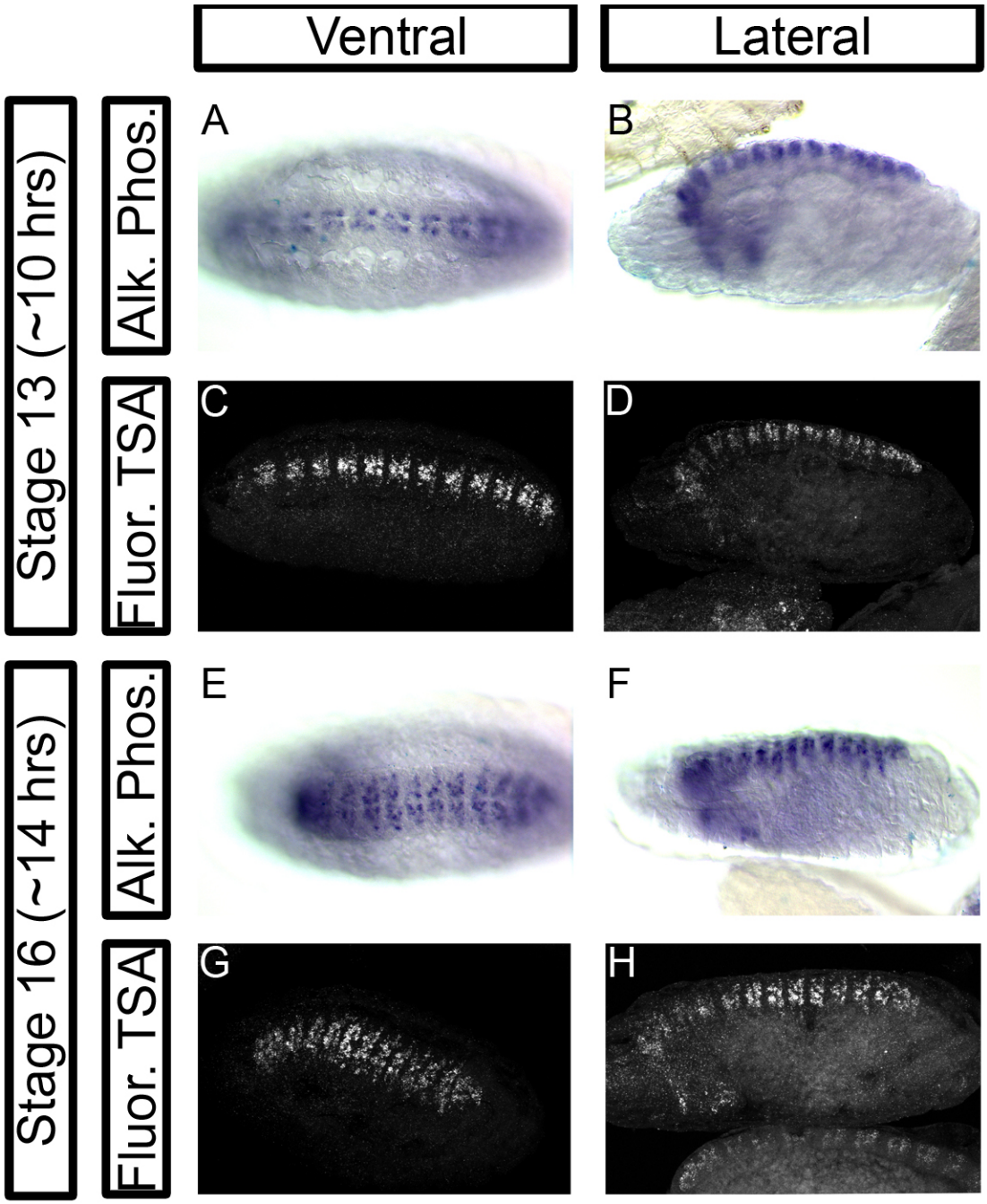
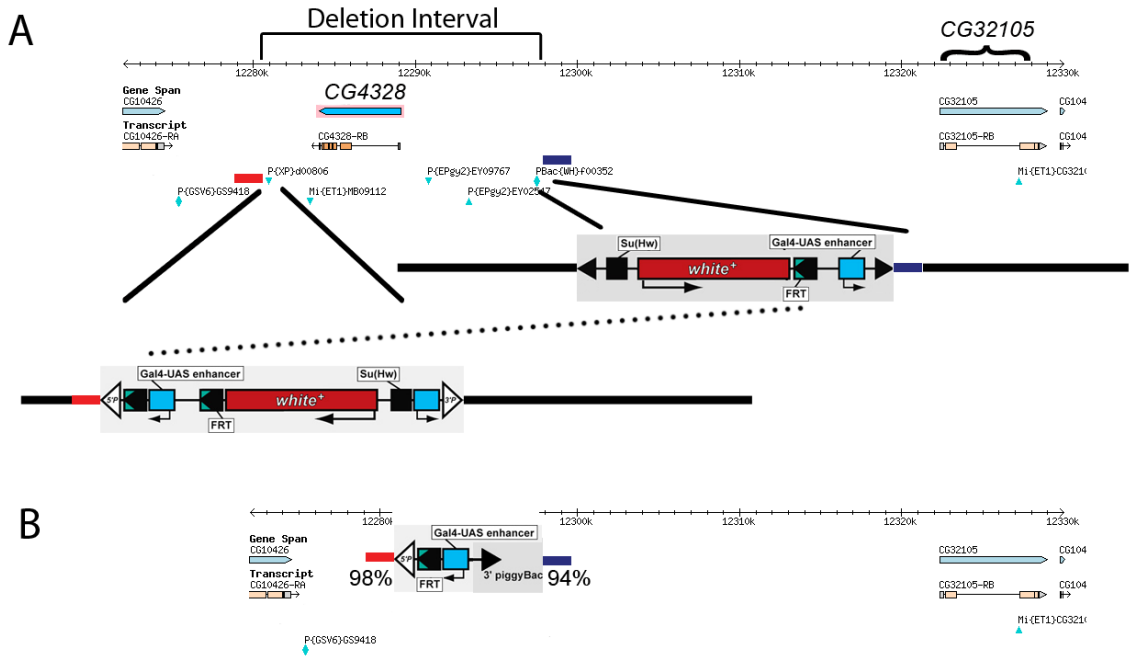


Figure S 2. Expression of *CG4328* transcript in the *Drosophila* embryonic central nervous system as detected by RNA in situ hybridization.

(A-D) *CG4328* transcript is detectable in Stage 13 (10 hrs) embryos along the midline of the ventral nerve cord. **(E-H)** Expression expands more laterally by Stage 16 (14 hrs.) Expression can be detected with an Alkaline Phosphatase chromogenic reaction or a fluorescent Horseradish peroxidase catalyzed Fluorescent Tyramide Signal Amplification reaction.

A.3.2 Flip/FRT Mediated Deletion of *CG4328*

A Flip/FRT based deletion strategy was used to generate a deletion in *CG4328*, using published methods (Parks et al., 2004). A piggyBac and p-element each containing an FRT site flank the *CG4328* locus with only the CG4328 gene between these two transposons. Flippase mediated recombination between the two FRT sites, in *trans,* generated a ~ 16 kbp deletion that was confirmed by PCR in the progeny. The PCR primers flanked the "outside" of the two transposons so that a predicted 2.3 kb product would be produced only upon successful deletion of the sequence between the FRT sites. This PCR product was fully sequenced to confirm the molecular identity of the deletion (See **Figure S 3**). Thus, I confirmed that the entirety of the *CG4328* gene locus was deleted, thereby generating a genetic null.



C

5'flanking genomic DNA; P-element inverted repeat; P[XP]sequence; FRT Site

pBac[WH] Sequence; piggyBac inverted repeat; 3'flanking Genomic DNA

5'ATGCACAATTGCTTCATTTTGCAAAAGGATTTTATTAAACCCGAAGCCGAGTGAAGTTGAAACCGAAGCAAAGTGTTGGTC
GATAAAGGCAAGTGGCAGGCTGTTTGGCGATGGCAGGAGTGGATCGAAGGATGTGGGTGCGAGTGTCCTGGATGCTTTTCCAAGC
ATACCATGATGAAATAACATAAGGTGGTCCCGTCGATAGCCGAAGCTTACCGAAGTATACACTTAAATTCAGTGCACGTTTGCTT
GTTGAGAGGAAAGGTTGTGTGCGGACGAATTTTTTTTTGAAAACATTAACCCTTACGTGGAATAAAAAAAAATGAAATATTGCAA
ATTTTGCTGCAAAGCTGTGACTGGAGTAAAATTAATTCACGTGCCGAAGTGTGCTATTAAGAGAAAATTGTGGGAGCAGAGCCTT
GGGTGCAGCCTTGGTGAAAACTCCCAAATTTGTGATACCCACTTTAATGATTCGCAGTGGAAGGCTGCACCTGCAAAAGGTCAGA
CATTTAAAAGGAGGCGACTCAACGCAGATGCCGTACCTAGTAAAGTGATAGAGCCTGAACCAGAAAAGATAAAAGAAGGCTATA
CCAGTGGGAGTACACAAACAGAGTAAGTTTGAATAGTAAAAAAAATCATTTATGTAAACAATAACGTGACTGTGCGTTAGGTCCT
CCCGGGCTGCAGAGATCTTGCGACGTGAACAGTGAGCTGTATGTGCGCAGCTCGCGTTGCATTTTCGGAAGCGCTCGTTTTCGGAA
ACGCTTTGAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCAGAGCGCTTTTGAAAACCAAAAGCGCTCTGA
AGACGCACTTTCAAAAAACCAAAAACGCACCGGACTGTAACGAGCTAGATCTTCAAGGGCGAATTGGGCGGCCGCTCTAGCCCC
CCCTCGAATGTTCTCTCTCTTCTCTTCTCTCTCTCTTTCTCGAATGTTCTCTCTCTTCTCTTCTCTCTCTCTTTCTCGAGGTCATCAAG
CTTAGGCCTCCAAGGCGGAGTACTGTCCTCCGGGCTGGCGGAGTACTGTCCTCCGGCAAGGTCGGAGTACTGTCCTCCGACACTA
GAGGTCGGAGTACTGTCCTCCGACGCAAGGCGGAGTACTGTCCTCCGGGCTGCGGAGTACTGTCCTCCGGCAAGGTCGGAGTACT
GTCCTCCGACACTAGAGGTCGGAGTACTGTCCTCCGACGCAAGGTCGGAGTACTGTCCTCCGACACTAGAGGTCGGAGTACTGTC
CTCCGACGCAAGGTCGGAGTACTGTCCTCCGACACTAGAGGTCGGAGTACTGTCCTCCGACGCAAGGCGGAGTACTGTCCTCCGG
GCTGGCGGAGTACTGTCCTCCGGCAAGGGTCGAGGGTACCGAGAGAGCGCCGGAGTATAAATAGAGGCGCTTCGTCTACGGAGC
GACAATTCAATTCAAACAAGCAAAGTGAACACGTCGAGATCTGACAATGTTCAGTGCAGAGACTCGGCTACGCCTCGTGGACTTT
GAAGTTGACCAACAATGTTTATTCTTACCTCTAATAGTCCTCTGTGGCAAGGTCAAGATTCTGTTAGAAGCCAATGAAGAACCTGG
TTGTTCAATAACATTTTGTTCGTCTAATATTTCACTACCGCTTGACGTTGGCTGCACTTCATGTACCTCATCTATAAACGCTTCTTCT
GTATCGCTCTGGACGTCATCTTCACTTACGTGATCTGATATTTCACTGTCAGAATCCTCACCAACAAGCTCGTCATCGCTTTGCAGA
AGAGCAGAGAGGATATGCTCATCGTCTAAAGAACTACCCATTTTATTATATATTAGTCACGATATCTATAACAAGAAAATATATAT
ATAATAAGTTATCACGTAAGTAGAACATGAAATAACAATATAATTATCGTATGAGTTAAATCTTAAAAGTCACGTAAAAGATAAT
CATGCGTCATTTTGACTCACGCGGTCGTTATAGTTCAAAATCAGTGACACTTACCGCATTGACAAGCACGCCTCACGGGAGCTCCA
AGCGGCGACTGAGATGTCCTAAATGCACAGCGACGGATTCGCGCTATTTAGAAAGAGAGAGCAATATTTCAAGAATGCATGCGTC
AATTTTACGCAGACTATCTTTCTAGGGTTAAAGACAACTTAAACGGTCCCCCGAGGAATTCGCATCCTTTTTTTTTTTTTTTTTGG
AAAAGTTTCCTTTAATATCCTGACTGATTTCGAGCTGTGCGGTCGTCGAATGGCCTCCTTGCTGTCATAACAACAATAACTGCATTT
TTAGCACCTCATTAAAAAAGAGTTCATATCGCA-3'

Figure S 3. Flip/FRT based deletion strategy to generate the *CG4328*^{cbk}-null allele.

(A) Genomic region deleted by Flippase induced recombination between two FRT sites in *trans*. **(B)** The entirety of the *CG4328* coding sequence and 16kbp of surrounding genomic DNA is deleted from the region between the two FRT containing transposons. No other genes are deleted. A hybrid p-element-piggyBac footprint is left after the Flip/FRT deletion. (**C**) The region surrounding the deletion locus was sequenced.

A.3.3 Preliminary Characterization of *CG4328* Mutants

CG4328 mutants are viable as homozygotes. However, when attempting to generate a stable intercross between homozygous flies, we consistently found that fewer progeny eclosed from the vial, than the number of parents put into the vial. Visual inspection suggested that far fewer eggs were being laid than by wild type flies or by the strains used to generate the flip-FRT deletion. This observation has led us to call this allele of *CG4328,* '*continental breakfast*' (*cbk*) ($CG4328^{cbk}$), since eggs are rarely encountered.

We hypothesized that *CG4328* may have a fertility phenotype. Work by our laboratory and that of others showed that the *Drosophila* oviduct is innervated by two neuronal populations, octopaminergic neurons and the Insulin Like Peptide 7 (dILP7)-expressing motoneurons (Castellanos et al., 2013). The soma of both neuronal populations are found along the midline of the *Drosophila* CNS, and both of these neuron populations are required for egg laying. Octopamine itself is required for correct rhythmicity of oviduct contraction (Middleton et al., 2006). Mutation of either *tyramine beta-hydroxylase* (*TBH*), the enzyme that catalyzes the last step in octopamine biosynthesis, or *tyrosine decarboxylase 2* (*Tdc2*) whose product synthesizes the octopamine precursor tyramine from tyrosine, results in animals that survive to adulthood and mate, but display a near complete loss of egg-laying (Cole et al., 2005; Monastirioti et al., 1996). Similarly, dILP7-neuron innervation of the oviduct is necessary for egg-laying, with loss of activity or death of dILP7-neurons resulting in a total loss of egg-laying (Castellanos et al., 2013). Female post-embryonic dILP7-neurons are the sole glutamatergic motoneuron of the oviduct; they terminate in fast excitatory type I-like neuromuscular junctions (NMJs) on the oviduct that can be immunohistochemically detected with anti-Discs large (Dlg) staining of the post-synaptic specialization of the NMJ, while octopaminergic neurons form neuromodulatory

type-II-like NMJs at the oviduct with no post-synaptic Dlg accumulation (Castellanos et al., 2013; Middleton et al., 2006).

To test if the low fertility phenotype we observed in homozygous *CG4328*^{cbk} animals was due to female egg-laying defects, we performed egg-laying assays wherein females were mated to w^{1118} males, and the number of eggs laid per day per fly was counted daily for up to 7 days. *CG4328*^{cbk} homozygotes displayed a 85% reduction in the average number of eggs laid per female per day compared to heterozygous controls (See **Figure S 4**) At least some of the eggs that were laid, however, were fertilized because larvae were observed on the embryo collection plates.

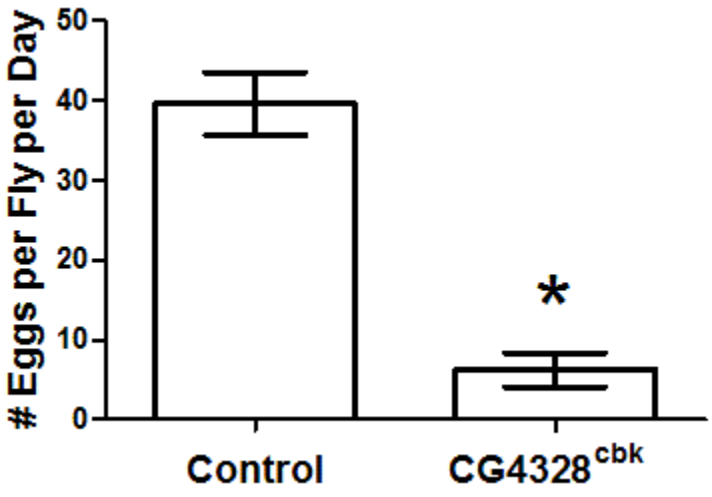


Figure S 4. Loss of *CG4328* function results in a loss of egg laying

Average number of eggs per fly per day averaged over one week. 3-5 day old females were mated 24 hours before being placed in cages with three mated females per age. *=$p<10^{-4}$ Student's two-tailed *t*-test n=17 cages for control, n=10 for mutants.

Given the profound loss of egg laying ability in females and the exclusive CNS expression of *CG4328*, we assessed whether loss of *CG4328* function may affect innervation of the oviduct. A Gal4 regulatory fusion construct from the *Tdc2* gene drives Gal4 activity specifically in the octopaminergic neurons of the *Drosophila* CNS (Cole et al., 2005). I used *Tdc2-Gal4* to drive *UAS-nlsGFP* in embryos to label the nuclei of octopaminergic neurons and performed fluorescent RNA *in-situ* hybridization using probes complementary to the *CG4328* transcript. *CG4328* transcript was detected surrounding and overlapping *Tdc2-Gal4>UAS-nlsGFP* labeled nuclei in Stage 16 embryos (See **Figure S 5**).

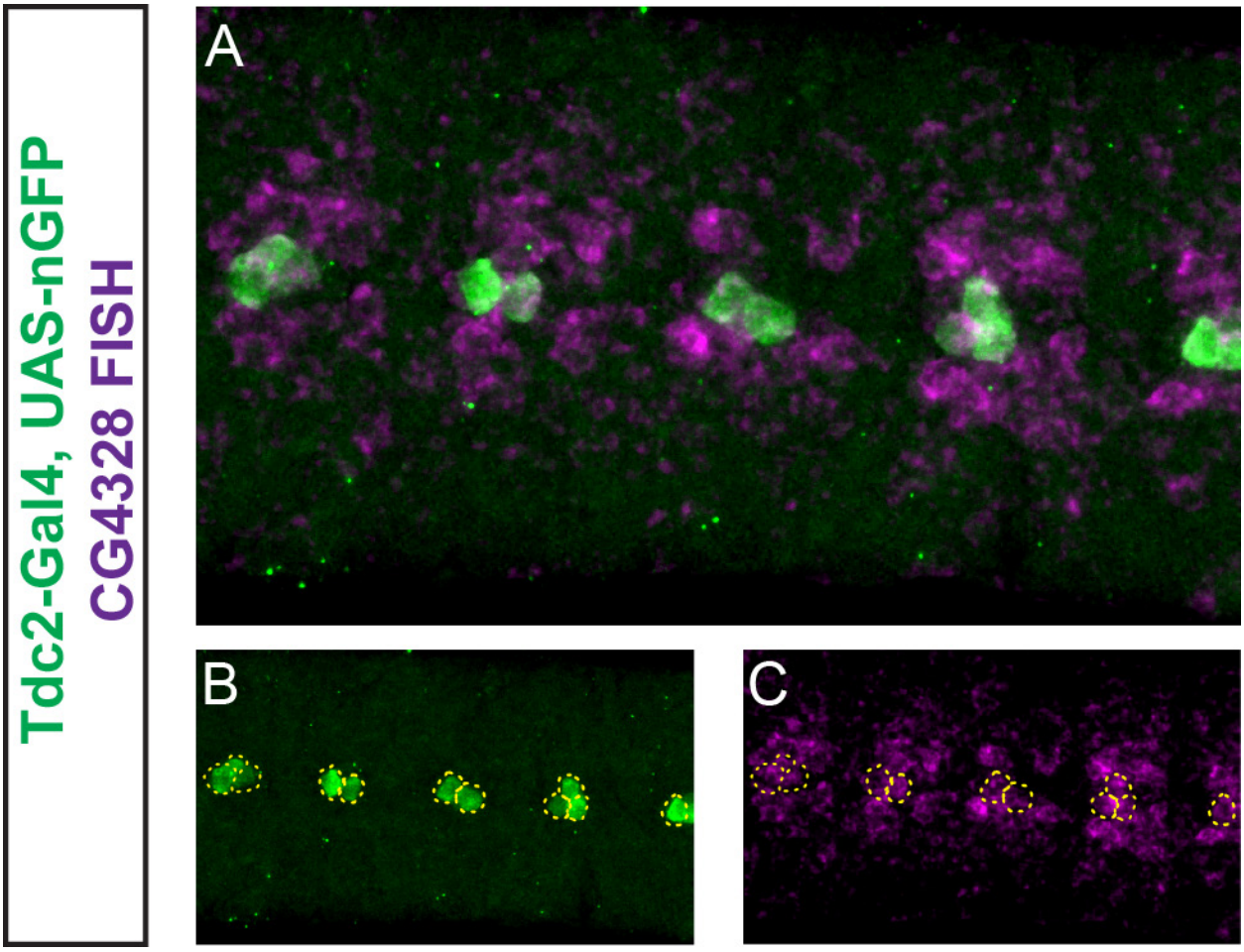


Figure S 5. *CG4328* is detectable in octopaminergic neurons in the embryo.

(**A**) Fluorescent *in-situ* hybridization (FISH) signal from a DIG-labeled RNA probe co-localizes with *Tdc2-Gal4>UAS-nlsGFP* labeled nuclei in Stage 16 embryos. (**B)** anti-GFP staining (**C**) fluorescent signal from TSA detection of RNA probe anti-sense to *CG4328* transcript**.**

A.3.4. Loss of *CG4328* Function Results in Oviduct Innervation Defects.

Since preliminary evidence suggested that *CG4328* is expressed in octopaminergic cells, we used *Tdc2-Gal4* and membrane localized *UAS-CD4::tdGFP* to label octopaminergic axons in the oviducts of 3-5 day old mated females, in $CG4328^{cbk}$ heterozygotes or homozygotes. Total *Tdc2-Gal4* labeled axon length per unit area of the oviduct was reduced by 63% in *CG4328* mutants compared to heterozygous controls (See **Figure S 6C**). In the same tissue, anti-Dlg staining was performed to detect type-I NMJs. A 46% decrease in the number of Dlg-positive NMJs per unit area was observed in the oviduct (See **Figure S 6D**).

Loss of octopaminergic innervation could potentially be due to improper axon guidance of octopaminergic neurons or lack of neurogenesis and/or cell death in *CG4328* mutants. Therefore, we examined *Tdc2-Gal4 ->UAS-CD4::tdGFP* expression, anti-TDC2 and anti-TBH in the VNCs of $CG4328^{cbk}$ mutants. No difference in the number of *Tdc2-Gal4* expressing, anti-TDC2, or anti-TBH neurons was observed between $CG4328^{cbk}$ mutants and heterozygous controls, nor was the intensity of the Gal4 or antibody staining noticeably different (See **Figure S 7**). Therefore, loss of GFP labeled axons in the oviduct is likely not due to simple agenesis or death of octopaminergic neurons, but rather to axon guidance defects.

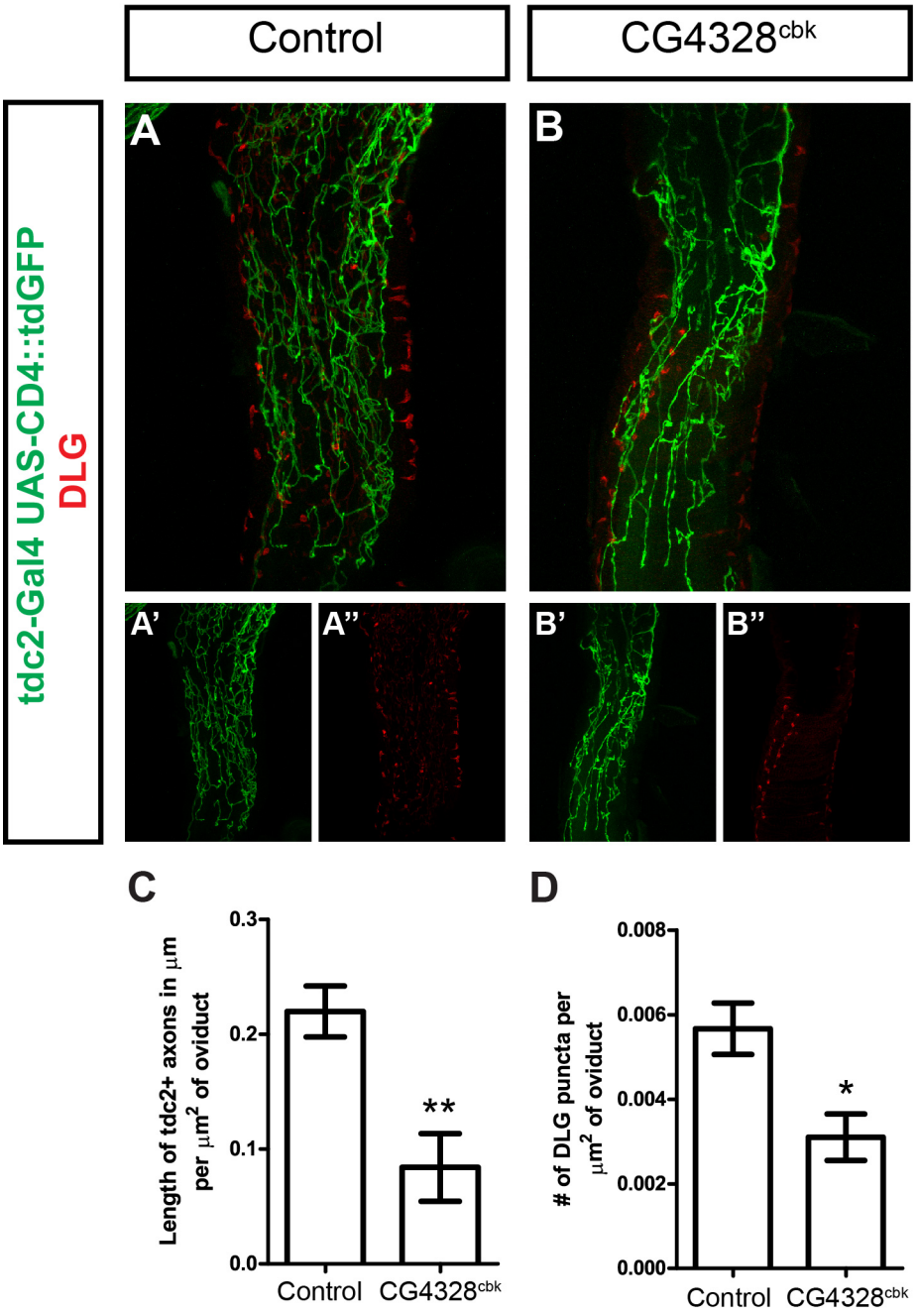


Figure S 6. Innervation of the oviduct is decreased in *CG4328^{cbk}* mutants

(**A-B**) Octopaminergic neurons are labeled by *Tdc2-Gal4, UAS-CD4::tdGFP* in the common oviduct (**A**' and **B'**- green). Anti-DLG staining is used to detect type-I NMJs of dILP7 neurons (**A''** and **B''**- red). (**C**) The total length of all *Tdc2-Gal4* labeled axonal processes in the oviduct was traced and then divided by the total area of oviduct imaged. A 63% reduction in *Tdc2*+ axon length per unit area was observed in homozygous mutants compared to heterozygous controls ($p<0.01$ Student's two-tailed t-test controls n=8, mutant n=6 oviducts). (**D**) The number of anti-Dlg labeled NMJs was counted and divided by the total area of oviduct imaged. A 46% reduction in *Tdc2*+ axon length per unit area was observed in homozygous mutants compared to heterozygous controls ($p<0.05$ Student's two-tailed t-test controls n=8, mutant n=6 oviducts).
Genotypes **Control**: *w; Tcd2-Gal4/+; UAS-CD4::tdGFP,CG4328^{cbk}/+*
Mutant: *w; Tcd2-Gal4/+; UAS-CD4::tdGFP,CG4328^{cbk}/CG4328^{cbk}*

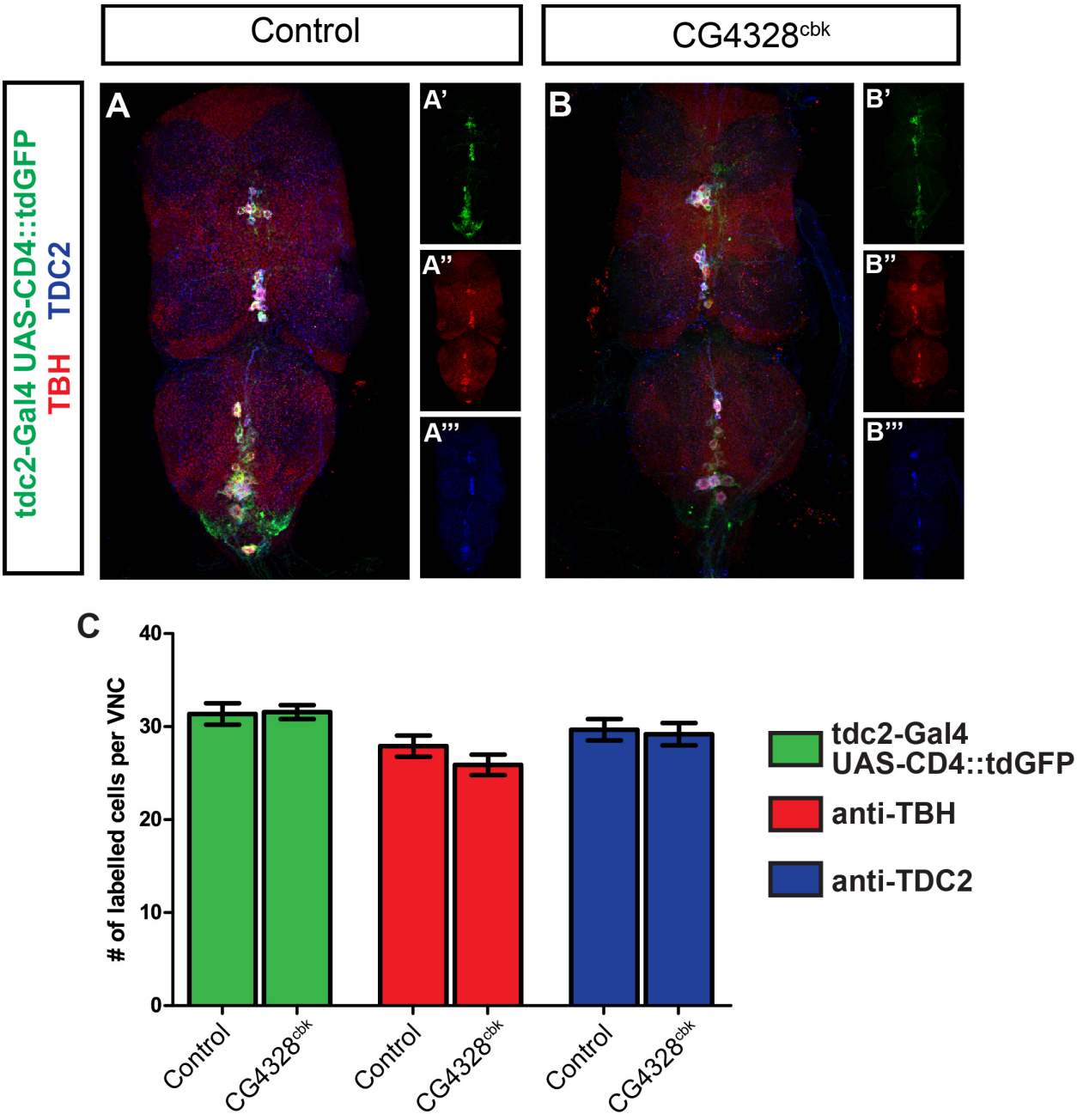


Figure S 7. Octopaminergic cell number is unchanged in *CG4328^{cbk}* mutants

(**A-B**) Adult ventral nerve cords with octopaminergic neurons labeled by *Tdc2-Gal4, UAS-CD4::tdGFP* (green) anti-TBH (red) or anti-TDC2 (blue). (**C**) Graph of cell counts from heterozygous and mutant VNCs. No significant difference in Tdc2+ or TBH+ cell number was detected between controls and mutants ($p>0.05$ Student's two-tailed t-test n=11 heterozygous and n=9 homozygous mutant VNCs per group). Genotypes **Control**: *w; Tcd2-Gal4/+; UAS-CD4::tdGFP,CG4328^{cbk}/+* **Mutant**: *w; Tcd2-Gal4/+; UAS-CD4::tdGFP,CG4328^{cbk}/CG4328^{cbk}*

A.4 Discussion

A.4.1 Egg Laying Phenotype of *CG4328* Mutants

Preliminary characterization of *CG4328* mutants shows a significant loss of egg-laying. *CG4328* appears to be expressed, at least during the late embryonic period, in octopaminergic neurons in addition to a number of more lateral populations that have not yet been identified. Reduced octopaminergic neuron innervation in the oviduct in $CG4328^{cbk}$ mutants provides an anatomical correlate for this behavioural phenotype and suggests a mechanistic connection, since octopaminergic innervation of the oviduct is required for proper oviduct contractility and egg laying (Middleton et al., 2006). We postulate that *CG4328* therefore may play a role in regulating axon guidance of octopaminergic neurons and is vital for the function of the reproductive system. Presence of a normal number of octopaminergic neurons, as detected by three different markers of octopaminergic cells, suggests that *CG4328* does not play a role in the early specification of these cells, as the correct number are born. Instead, *CG4328* likely is regulating axon extension or guidance to the oviduct. Expression of *CG4328* in other cell types in the VNC during development precludes one from concluding that it is only loss of octopaminergic innervation that is mediating reduced egg-laying. Reduction of anti-Dlg labeled NMJs in the oviduct suggests that dILP7 neurons may also be is also affected since they are the only neurons that innervate Dlg-positive NMJs in the oviduct (Castellanos et al., 2013). Additionally, reduced egg-laying may be due to more centrally mediated problems occurring in the VNC or adult brain due to other neuronal developmental defects.

A.4.2 Future Directions

First and foremost, confirming the expression of *CG4328* in octopaminergic neurons and dILP7-neurons during both development and potentially adult life is required. A method that is

more facile than RNA *in-situ* hybridization is needed for experimental detection of *CG4328* at multiple developmental time points will therefore be necessary, since fluorescent *in situ* hybridization is time consuming and often only reveals "fuzzy donuts" around cells of interest, making co-localization with antibody or genetically encoded markers difficult. Two potential avenues will be explored. A polyclonal antibody to *CG4328* could be generated. Alternatively, CRISPR and homology directed repair can be used to integrate a GFP coding sequence or an epitope tag at the N- or C- terminus of the *CG4328* coding sequence. This should generate a faithful reporter construct reproducing the expression pattern of *CG4328*. After one or both of these tools are made, the expression pattern of *CG4328* would be characterized at embryonic, larval and adult stages.

Of primary interest are the octopaminergic neurons. Is *CG4328* expressed by all octopaminergic neurons? This will be tested using available antibodies and genetic reporters, once these are generated. We will also test the requirement of *CG4328* in these other octopaminergic neurons. Octopaminergic neurons play roles in aggression (Andrews et al., 2014; Hoyer et al., 2008), courtship (Certel et al., 2007), sleep (Crocker and Sehgal, 2008), and feeding behaviours (Zhang et al., 2013). We plan to run $CG4328^{cbk}$ mutants through a battery of behavioural tests, to explore which behaviours require CG4328 for normal function.

A.5 Mutagenesis of *CG32105*

Unlike the *CG4328* locus, the *CG32105* locus lacked convenient FRT containing transposons to perfom a facile deletion. However, a publicly available stock with a *Minos MiET1* transposable element inserted in to the first intron near the splice acceptor site was publicly available. I therefore mutagenized the locus by mobilizing the *Minos* element and screening for imprecise excisions. According to published literature, imprecise excisions with *Minos* elements are rare and not usually as large as the more commonly used *p-elements* (Adams et al., 2003; Metaxakis et al., 2005; Witsell et al., 2009). Therefore, we increased the size of excision and frequency of imprecise excision by performing the mobilization in a *Drosophila Bloom helicase* (*Blm,* also known as *mus309*) mutant background. Previous studies have shown that the absense of Blm helicase results in increased *P-element* and *Minos* element excision size by several fold (Adams et al., 2003; Witsell et al., 2009). Here, we found that transposon mobilization of the *Minos* element in a *Blm* mutant background increased deletion size and frequency several fold, generating two verified null alleles of *CG32105*. These results are the second demonstration of the efficacy of *Blm* helicase mutants to increase imprecise excision frequency and size in a well controlled side-by-side comparison of mutagenesis approaches (See **Table S 1** and **Figure S 9**).

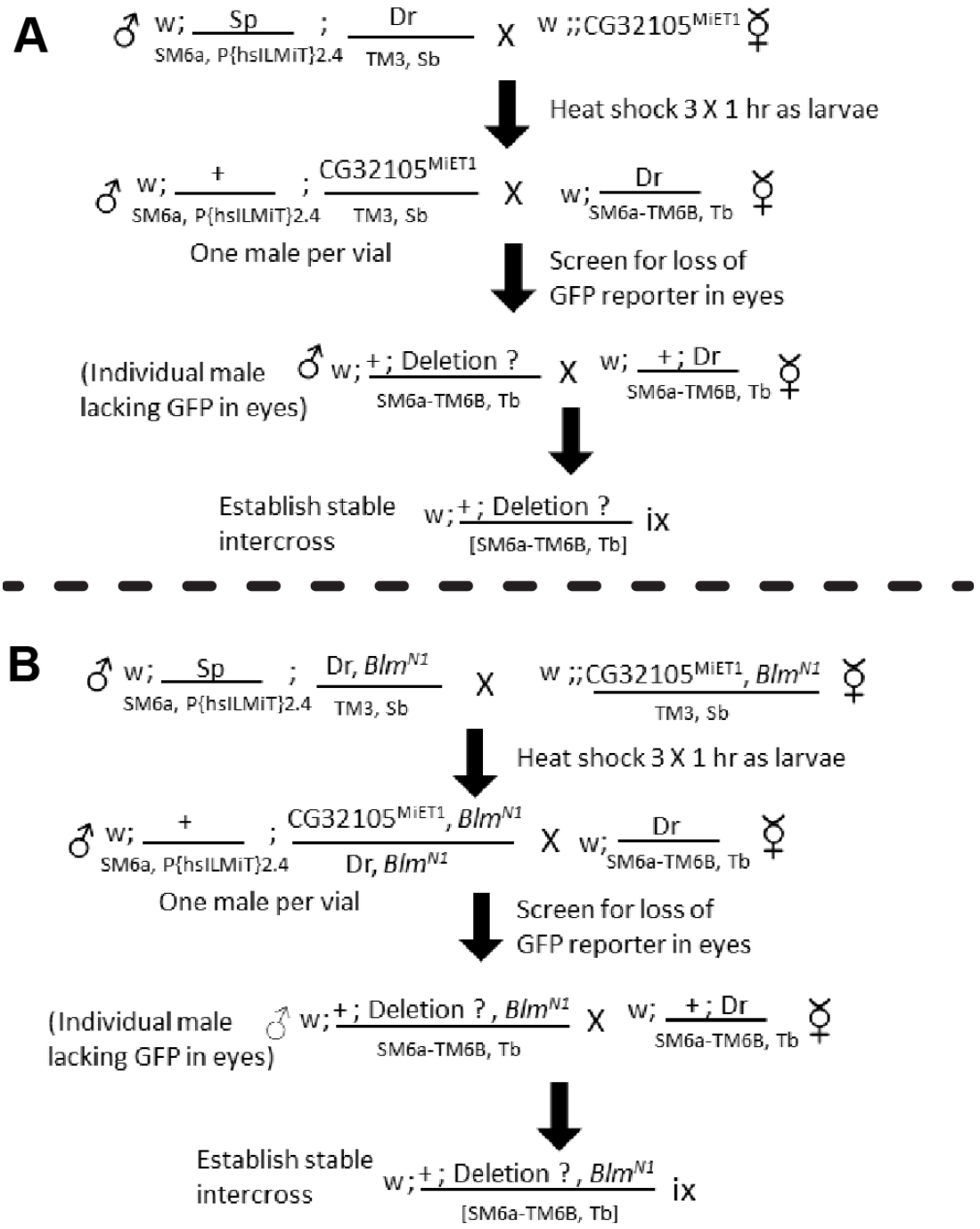


Figure S 8. Summary of crosses used to generate *CG32105* alleles

The first mobilization screen (in the absence of *Blm*) was done using the available heatshock induced *Minos* transposase in an otherwise wild-type background containing the *MiET1* insertion in *CG32105 (Metaxakis et al., 2005)*. The crossing scheme is detailed in Figure S 8. From 200 heat-shocked males, 96 unique mobilizations events of the *MiET1* insertion were recovered, as individual male progeny that showed loss of the transposon-contained eye GFP reporter. All excisions were balanced over the *SM6a-TM6B,Tb* fused balancer chromosome, and all produced homozygous flies. PCR screening for deletions was perfomed on homozygous animals. The first-pass screening PCR primers were situated ~150 bp on each side of transposon insertion site to produce a product of an expected size of 281bp, if a precise excision had occurred. Of these, 91 lines produced PCR product showing a precice excisions had occured. Four lines produced PCR product that, when extension time was increased, contained a MiET1 element with an evidentally non-functional GFP reporter. One line initially produced no PCR product. When primers situated further away from the MiET1 insertion site were used, a 343 base pair deletion that extended into the intron of *CG32105* was found; however, this did not remove any coding sequence. I have called this allele $CG32105^{B}$ since the first allele of the gene was the intial *MiET1* insertion which I call $CG32105^{A}$. Data summarized in **Table S 1** and **Figure S 9**.

Table S 1. Summary of imprecise excisions efficiency for *Minos* element mobilization in the CG32105 locus

	Number of heat shocked males screened	**Number of mobilization events**	**Fraction of crosses that yield successful mobilization events**	**Number of detectable deletions**	**Fraction of mobilizations that yield deletions**
Conventional Hop	200	96	48%	1	1%
***Blm* mutant Hop**	233	115	49%	4	3%

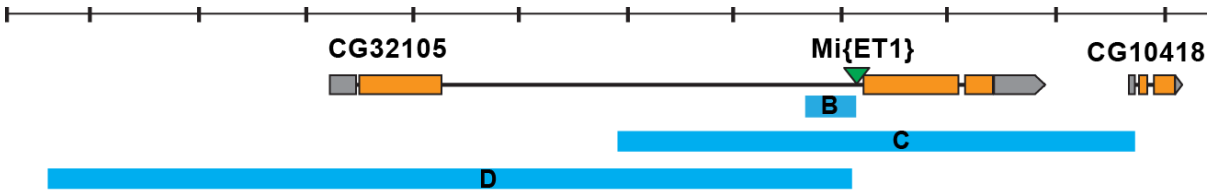


Figure S 9. CG32105 locus showing the location and size of deletions generated by imprecise excision of a Minos transposon.

Letters assigned to the deletions (blue bars) correspond to the allele assigned to the deletion. Grey indicates 5' and 3' UTRs, orange is protein coding sequence. Ruler is 1 kb intervals. Deletion B only removes intronic sequence, while C and D generate null alleles of *CG32105*. Deletion C removes part of the 5'UTR of CG10418 and is lethal when crossed to the deficiency while deletions B and D are not lethal when crossed to the deficiency.

Given the ~1% success rate of generating an imprecise excision and our inability to obtain a coding sequence deletion of *CG32105* to confidentally generate a null allele, I expressed the *Minos* transposase in a Drosophila Blm^{N1} homozygous mutant background (see Figure S 8 for crossing scheme). Of 233 heat-shocked males, 105 progeny were recovered that showed loss of GFP expression in the eye. Lines were balanced and screened as above. All but two lines generated homozygous progeny. The two lines that did not generate homozygotes were crossed to a deficiency that spans the *CG32015* locus and several genes on either side (*Df(3L)ED4483*). Both of these excisions were lethal when heterozygous with the deficiency and did not produce adults. By comparison, the original *Minos* element insertion ($CG32105^{A}$), the known precise excision alleles, and the small intronic deletion $CG32105^{B}$, were all viable when heterozygous with the deficiency.

Therefore, four potential exiscion events, two of which were lethal over a deficency, had been generated. Primer combinations staggered at 3 kb intervals on either side of the *CG32105 MiET1* insertion site were then used to screen for deletions in these four lines, out to a maximum of 12 kb in 5' and 3' directions from the insertion site. Two deletions were mapped, the first one identified, $CG32105^{C}$, is a ~5 kb deletion spanning most of the first intron and all of the second and third exons of *CG32105* (Figure S 9). This deletion also extends into the 5' UTR of the adjacent gene, *CG10418*, which contains a *transformer*-like predicted RNA splicing protein. The second mapped deletion, $CG31205^{D}$, is a ~7 kb deletion that extends unidirectionally from the *MiET1* insertion site towards the 5' end of *CG32105*, deleting all of the first exon, the 5'UTR and ~3 kb upstream of the gene. Unlike the "C" allele, the "D" allele is homozygous viable and not lethal when heterozygous with the deficiency. Therefore, I conclude that the $CG32105^{C}$ allele is lethal in combination with the deficiency due to the partial deletion of the neighbouring

gene, *CG10418*, rather than mutation of *CG32105* itself. Comparison of mutagenesis approaches is shown in Table S 1.

A.6 Discussion

We generated three unique deletions in the *CG32105* locus, two of which delete large portions of the gene and produce genetic nulls. In my own experiments, mobilization of a *Minos* transposon in a *Blm* helicase mutant background increased the number of imprecise excision events four fold, and the size of the resulting deletions ten fold. These results are similar to previous reports of *Blm* helicase increasing imprecise excision frequency and size by at least 5-10 fold (Witsell et al., 2009). Genetic nulls of *CG32105* appear to be viable. Further analysis of *CG32105* expression patterns, and potential phenotypes, is now required.

Transposase activity generates double-strand breaks (DSBs) in the *Drosophila* genome. Approximately 25% of Minos-induced double-strand breaks in animals heterozygous for the insertion are repaired by nonhomologous end joining (NHEJ) and mismatch repair, frequently resulting in a 6-bp insertion, or "footprint," relative to the original target sequence, while the other 75% are proposed to be repaired by homology-directed repair (HDR) (Arca et al., 1997). Recently, the CRISPR-Cas9 (clustered regularly interspaced short palindromic repeats) based genome editing system has been applied to *Drosophila*. Nuclease activity of Cas9 generates double-stranded breaks at sites that are complementary to guide-RNA (gRNA) sequences. Deletional mutagenesis has been shown to be possible through generation of DSBs that are repaired by error-prone NHEJ or HDR (Gantz and Bier, 2015; Gratz et al., 2013; Gratz et al., 2015; Ren et al., 2014). *Drosophila Blm* helicase is required for efficient and accurate HDR (McVey et al., 2004). Therefore, CRISPR-Cas9 mediate mutagenesis of loci, where deletion of

coding sequence is required, may benefit from performing the mutagenesis in a *Blm* mutant background and may make loci "recalcitrant" to CRISPR-Cas9 mutagenesis accessible, or, potentially make deletions large enough to easily screen by conventional PCR methods. However, if alleles of *Blm* are used, one must cross-in and the cross-out the *Blm* alleles. This requires either recombination or double balancing of lines, which can take months. Once an appropriate deletion is generated, the *Blm* mutation then must be out-crossed, further increasing the labor and time commitment. If a number of loci are to be mutated, this approach would be laborious and slow. Recently, a small molecule inbitor of *Blm* helicase acitvity, ML216, has been identified (Nguyen et al., 2013). Potentially, adding this to the food that animals consume during either transposon mobilization and/or gRNA expression may expedite large scale mutagenesis projects since the introduction and then out-crossing the *Blm* allele after the desired mutations are identified is no longer necessary. Further, since loss of *Blm* helicase sensitises the genome to environmental mutagens, use of a small molecule inhibitor removes the potential of spontaneous second-site mutations accumulating in stocks that may be crossed and flipped for multiple generations before use and screening.

APPENDIX B- SUPPLEMENTAL FIGURES

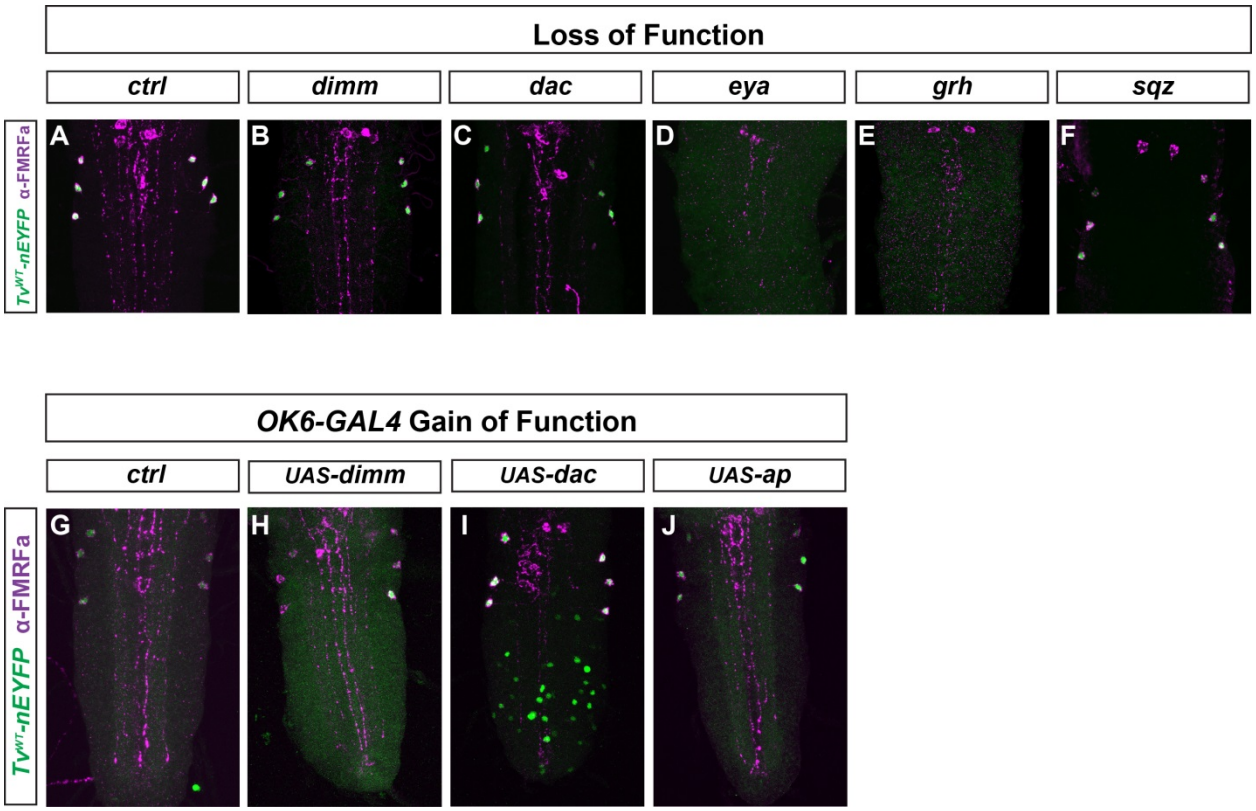


Figure S 10. Expression of the Tv^{WT}-nEYFP reporter is regulated by a subset of transcription factors known to affect FMRFa gene and peptide expression

(A-F) Representative images of the quantitative loss of function data shown in Figure 1. **(G-J)** Representative images of the quantitative gain of function data shown in Figure 1 using *OK6-GAL4* to drive transcription factors known to generate ectopic FMRFa expression elsewhere in the VNC. Whole VNCs were imaged through the entirety of their z-axis.

***Loss of function genotypes:* ctrl** *(Tv^{WT}-nEYFP).* **dimm** *($dimm^{rev4}/dimm^{P1}$; Tv^{WT}-nEYFP)* **dac** *(Df(2L)Exel7066/ dac^{3}; Tv^{WT}-nEYFP),* **eya** *($eya^{Cli\text{-}IID}/eya^{D1}$; Tv^{WT}-nEYFP).* **grh**, *(grh^{IM}/grh^{Df}; Tv^{WT}-nEYFP).* **sqz** *(Tv^{WT}-nEYFP,sqz^{ie}/ Tv^{WT}-nEYFP,sqz^{Df})*

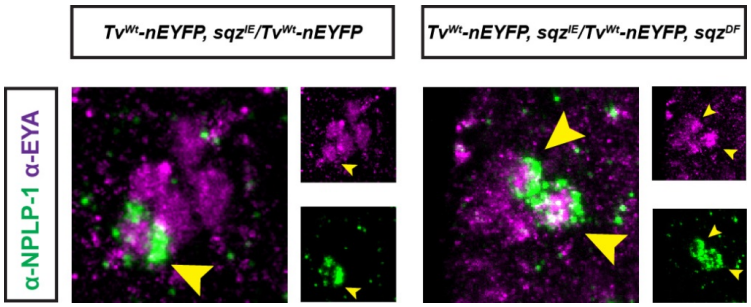
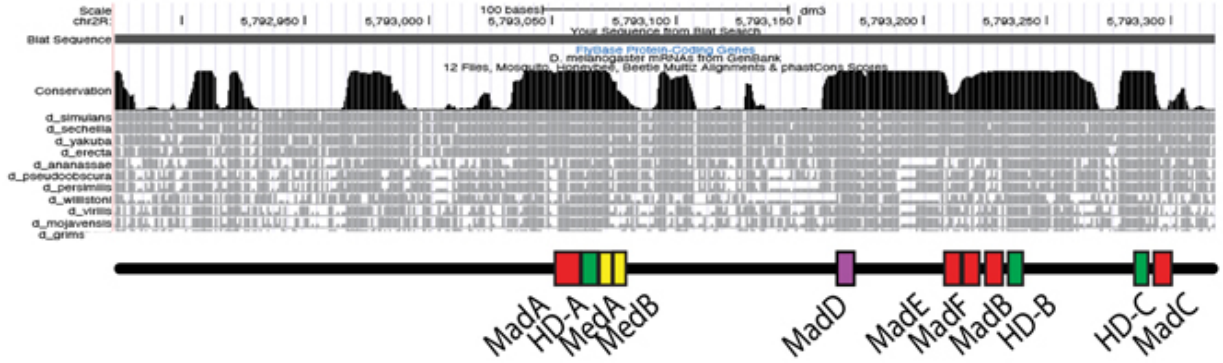


Figure S 11. Confirmation of *sqz* mutant phenotype in Tv^{WT}-nEYFP reporter experiments.

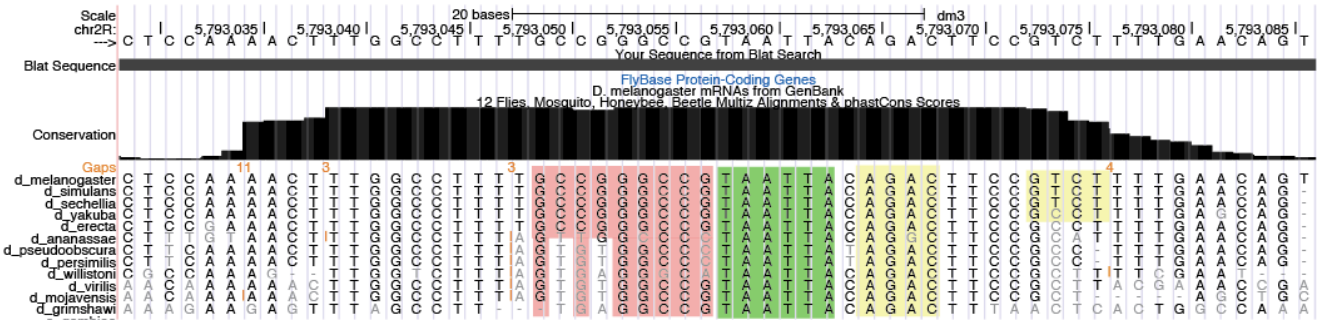
(A&B) Anti-NPLP1 staining in Thoracic Segment 1 (T1) Tv cluster cells shows the expected supernumerary NPLP-1 immunoreactive cells in *sqz* mutants, but not heterozygous controls.

A. Whole Tv4 enhancer showing conserved HD, Mad and Med regions.

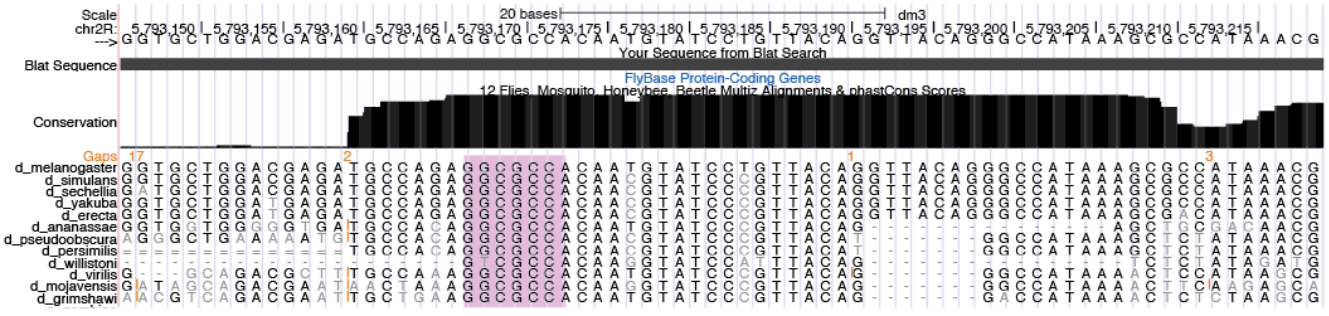


Mad sites are highlighted in Red, Medea sites in Yellow, Homeodomain sites in Green throughout this figure. The only consensus Mad site (MadD) is shown in Purple.

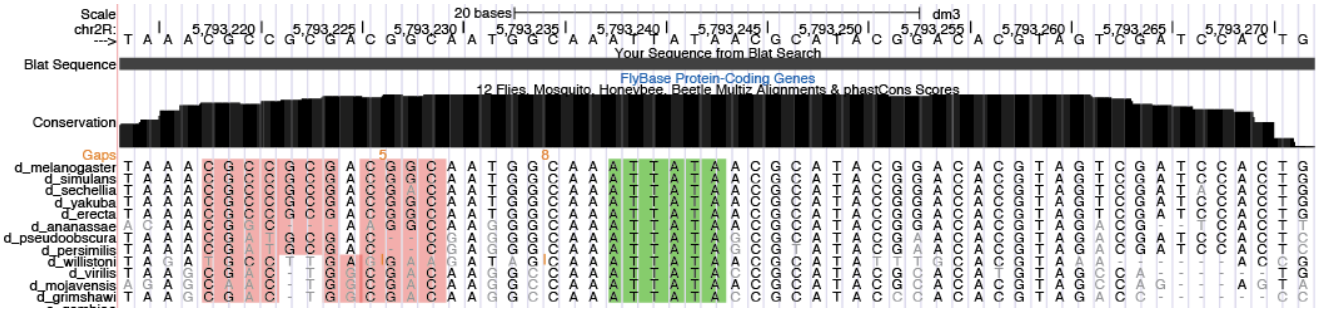
B. Sequence identity of HD-A (green), Mad-A (red), Med-A (yellow) across 12 Drosophila species



C. Sequence identity of Mad-D (purple) across 12 Drosophila species



D. Sequence identity of HD-B (green) and Mad-B (red) across 12 Drosophila species



E. Sequence identity of HD-C (green) and Mad-C (red) across 12 Drosophila species

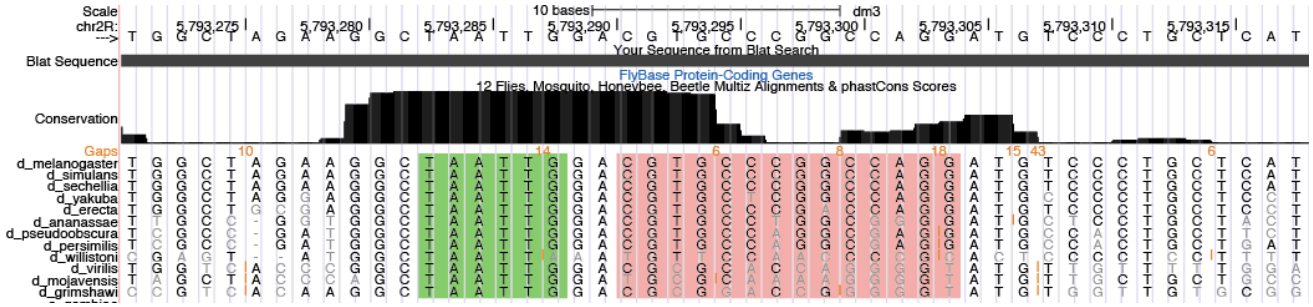


Figure S 12. Sequence identity of conserved regions of Tv4 across multiple Drosophila species

Figure S 13. Tv4 enhancers for all 12 Drosophila species

Capitalized blue letters denote identity to the *D. melanogaster* Tv4 enhancer.

thick underline denotes BMP-RE

double underline denotes HD-RE
red highlight = putative Mad-binding sequence
magenta highlight = putative Mad-binding sequence of Mad-D
green highlight = putative Ap-binding sequence
yellow highlight = putative Medea-binding sequence

>D. melanogaster
CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCA
ACAAGAGCAGTGGACACCCAATAAACCTGTCCAAAAACCCGACACATTTCTGC
CCAGTCATGCGTGGTGGACAATAGCCAAATGCCATTGATGAGACTCGTCTCCA
AAACTTTGGCCTTTTGCCGGGCCGTAATTACAGACTTCCGTCTTTTGAACAGTT
TTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTGGGATGGGGTGG
TTTCGGGTGCTGGACGAGATGCCAGAGGCGCCACAATGTATCCTGTTACAGGT
TACAGGGCCATAAAGCGCCATAAACGGCCGCGACGGCAATGGCAAATTATAACG
CATACGGACACGTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGG
CCAGGATGTCCCTGCTCAT

>D. simulans
acgacttggcctggtcatcaaacccaagccaagatttgaacttgaactagtctcgccttatcacattgccttctaattcgttttcattttc
gattttcgcatctacttatgctaatgaagCCATCTGCAGACGTGGTTTTCGAgCGTATTTATATcG
ATTATGGGTGATCGcCAACgAGcGCAGTGGgCACCCAATAAACCTGgCCAAAAAC
CCGACACATTTCTGCCCAGTCATGCGTGGTGGcCAAaAGCCAAATGCCATTGAT
GAGACTCGTCTCCAAAACTTTGGCCTTTTGCCGGGCCGTAATTACAGACTTCCG
TCTTTTGAACAGTTTcTCAGCCCCACCCAAGAGgCGAGTCTTGAAAAGCTGGCgG
GGATGGGGTGGTTTCGGGTGCTGGACGAGATGCCAGAGGCGCCACAAcGTATC
CcGTTACAGGTTACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGGC
AAATTATAACGCATACGGACACGTAGTCGATCCACTGGCTAGAAGGCTAATTG
GACGTGCCCGGCCAGGATGTCCCTGCTCATcccgcagccgacttaaaagtgggcgtaacataacaat
tgtgggcgatttgcatacaatttggtgctcgagtggccagcaaactggcgtcgaaatgttgctacgtcattgttgctgcctct

>D. sechellia
cgcagttggcctggtcatcaaacccaagccaagatttgaacttgacctagtctcgtcttatcacattgccttctaatttgttttcattttc
gattttcgcatctacttatgctaatgaagCCATCTGCAGACGTGGTTTTCGAgCGTATTTATATcG
ATTATGtGTGATCGcCAACgAGcGCAGTGGACACCCAATAAACCTGTCCAAAAAC
CCGACACATTTCTGCCCAGTCATGCGTGGTGGcCAAaAGCCAAATGaCATTGAT
GAGACTCGTCTCCAAAACTTTGGCCTTTTGCCGGGCCGTAATTACAGACTTCCG
TCTTTTGAACAGTTTcTCAGCCCCACCCAAGAGgCGAGTCTTGAAAAGCTGGCgG
GGatATGGGGTGGTTTCGGaTGCTGGACGAGATGCCAGAGGCGCCACAAcGTAT
CCcGTTACAGGTTACAGGGCCATAAAGCGCCATAAACGCCGCGACGaCAATGGC
AAATTATAACGCATACGGACACGTAGTCGATaCACTGGCTAGAAGGCTAATTGG

ACGTGCCCGGCCAGGATGTCCCTGCTCATcccacagccgacttaaaagtgggcgtaacataacaattg
tgggcgatttgcatacaatttggtgctcgagtggccagcaaactggcgtcgaaatgttgctacggcattgttgctgcctct

>*D. erecta*
cgacttggcctggtcatcaaaccccaaaccaaaatttgaacttgaactagtctgggcttatcacattgccttctaattcgttttcattttc
gacttttgcatctatttatgctaatgaagCCAaCTGCAtACGTGGTTTcCGcgCGTgTTTATATTGAT
TTGGaTGATCGTaAAGAgcGCAGTGGAtACCCAATAAACCTGgCCAAAACCCGAC
ACATTTCTGCCCAGTCATGCGTGcTGGcCAAaAGCgAAATGCCATTGATGAGACT
CGTCTCCgAAACTTTGGCCTTTTGCCGGGCCGTAATTACAGACTTCCGcCTTTTG
AACAGTTTcTCAGCCCCACCCAAGAGgCGAGTCTTGAAAAGCTGGCgGGGATGGt
gtGTGGTTTgGGGTGCTGGAtGAGATGCCAGAGGCGCCACAAcGTATCCcGTTAC
AGGTTACAGGGCCATAAAGCGaCATAAACGCCGCGACGGCAATGGCAAATTAT
AACGCATACGGACACGTAGTCGATCCACTGGCTgcgAGGCTAATTGGACGTGCC
CGaCCAGGATGTCCCTGCTCcTcccacccgcgacttaaaagtgggcgtaacataacaattgtgggcgatttgca
tacaatttggtgctcgagtggccagcagactggcgtcgaaatgttgctacgtcattgttgccgctttt

>*D. yakuba*
aataaacatcaaatttatttaaatttgaagatacttatatcattgagacatagaaaggaattcaaatgcaaatgaatgaacaataaa
ggcttctggagtgagtccatagattacagagcaACGTGGTTTTCtcgaccagatcataatttgtagctgaactattttcaa
ctggccgattggcacacgacttggcctggtcatcaaacccaaagctaaatttgaacttgaactagtctcgcctcatcacattgccttct
tcgccttcattcgttttcatttccgatttttgcatctatttatgctaataaagccaactgcagacgtggtttctGAgCGagTTTAT
ATTGATTtTtGGTGATCGTAAAGAgcGCAGTGGACACaCAATAAAgCTGgCCAAAA
ACCCGACcCATTTCTGCCCAGTCATGCGTGcTGGcCAAgAGaCAAAGCCATTGAT
GAGACTCGTCTCCAAAACTTTGGCCTTTTGCCGGGCCGTAATTACAGACTTCCG
cCTTTTGAgCAGTTTcTCAGCCCCACCCAAGAGgCGAGTCTTGAAAAGCTGGCggg
gatggggTGGGATGGGGTGGgTTgGGGTGCTGGAtGAGATGCCAGAGGCGCCACAA
cGTATCCcGTTACAGGTTACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCA
ATGGCAAATTATAACGCATACGGACACGTAGTCGATCCACTGGCTAGgAGGCTA
ATTGGACGTGCtCGGCCAGGATGcCCCTGCTCcTcccactggcgacttaaaagtgggcgtagcataa
caattgtgggcgatttgcatacaatttggtactcgagtggccagcaggctggcgtcgaaatgttgctacgtcattgttgctgctttt

>*D. ananassae*
ttatgcaaatggagccatcggtctggtggttataaaatttgtatattgatttccgaagatcgtaaagagcgcagctgaacaataatct
tgacggacacgagacgacttggtcttctgaacTTCTGCCCAGTCATGtGTGCCATTGAcggggaaacaact
ggagcttgtcttcgactggtctttgtAACTttgTTGGCCTTtaagTTgGCCCCTAATTACAGgCTTCC
GcaTTTTGAACAGTTTTgcgGCCCCACCCAAGgGgCGAGgCTTaAAAACTGGgcgatT
GGGcactgaacataGGGCTGGACtgggtctggagtggggtggtgggggtgactTGCCAcAGGCGCCAC
AATGTATCCcGTTACAGagctGCGacaACGGCAAggcaagGGCAAATTATAACGCATA
CGGACACGTAGacgtcacttgccggtGGCTAATTGGACGTGCCtGGCggGGATtgcttggctcgt
ccagctcctgctactcctgaagctggcaatcgccttaaaaatgggcgtaccataacaattgtgggcgatttgcatacaatttggctct
caagtggccagaacactggc

>*D. persimilis*
gcaaatgaagccatcagcatgtttaagtatccaaactggtttctattgatttcggacgatcgtaaagtgagcatccaacacattaagc
cagacgacaagcagatacatgacaatggacctGACACATTTCTGCCtcaTCATGCccaGTGGACAAA
GtCATTGATaaaacgggcatagtacgtacgtacaagaaacaacacgcacagatacagagacagaattgggctaactggc

AGACTtGTCTtCAAAACTTTGGCCTTTcaaagtgtGGCCcTAATTAtAGACTTCCGCcT
TTGAACAGTTTcgCAGCCCCACCCAAGAGgaggggcctgagacaaagacagggactgggacagaggg
acagagacaggcctgaaaaatgctTGCCAcAGGCGCCACAAcGTATCCcGTTACAtGGCCATAA
AGCtCtATAAACGatGCGACcgagGGCAAATTATAgCGtATACGaACACGTAGaCGAT
CCACTcgccgatGGCTAATTGGACGTGCCaGGCgAGGctactccttcCCCTGCTgATgccactg
ctgatgccactgctgctgccccggctgctgccccggctactgctccggctgctgctgttccttctggaacttaaaagcgggcgtaaca
taacaattgtgggcgatttgcatac

>*D. pseudoobscura*
ggacctgacacatttctgcctcatcatgcccagtggacaaagtcattgataaaacgggcatagtacgtacgtacaagaaacaacac
gcacatatacagagacagaagtgggctaactggcAGACTtGTCTtCAAAACTTTGGCCTTTcaaagtgtG
GCCcTAATTAtAGACTTCCGcCTTTGAACAGTTTcgCAGCCCCACCCAAGAGgagggg
cctgagacaaagacagggacagggacagggctgaaaaatgctTGCCAcAGGCGCCACAAcGTATCCcGT
TACAtGGCCATAAAGCtCtATAAACGatGCGACcgagGGCAAATTATAgCGCATACGa
ACACGTAGaCGATCCACTcgccgatGGCTAATTGGACGTGCCaGGCgAGGctactccttcC
CCTGCTgATgccactgcttctgccccggctgctgccccggctgctgccccggctgctgctccggctactgctgttccttctgga
acttaaaagtgggcgtaacataacaattgtgggcgatttgcatac

>*D. virilis*
catgcatatttatcgatgcaaatgtctggctattgattttcgacgatcgttagagaacaaaaaaaaaaagaagctaaaatgaaata
agaattgaagcgaactgccagacacacaatggttGACACATTTCTtGCaAGTCATGgtccacatcttggccgga
cgctgggcgggttcttcaaattcagacacgcaaaaaaaataagaaacgggttgaaaacaaAAAACTTGGCCTTTca
aagtgtGGCCGTAATTACAGACTTCCGCTTacgaaaccgaaagtgtccgcccagccaaatggccaagaag
tattgaccgaaGATGGGGTtGGgactgggcgcagTCGGGcaGACGctttgggcTGCCAaAGGCGCC
ACAATGTATCCcGTTACAGGGgCCATAAAaCtCCATAAGCGACtGGCgacAAgGcCA
AATTATAcCGCATACGcACAtGTAGccatggtctaagttgtggacccGGCTAATTGGACGcgcacc
aggggtatggtcaggggtggtggggttgtttttggattctctataaacttaaaagtgggcggcacataacaattgtgggcgatttgc
atacaatttggcgcttggattaccag

>*D. willistoni*
agtcattgataaaaagcacaagaatgtagaagaagatgaagaagaacaaacacgaataaaaaaaaaactactatctcacacag
tcagtgtcgccaaagttggtctttcaaagtgagggcaTAATTACAGACTTCCGCTTTcaattcgaatgttgtgggc
cACAGTTTagAGAGTCcttttgccagcgGATGCCtgtCGCCACAAgGTATCCatTTACAGtctc
tatagatgccttgagaagaggaagatagaagacgaaCAAATTATAACGCATAtttgCACGTAaaaccgagta
tggctaatttttactaagaaagaaaatgttccaaccccgcagcccacccactcccctccgtactgttttctctcgcccaatttcaaaaa
agtgggcgtaacataacaatt

>*D. mojavensis*
catttgttgccagtcatcctcaccgactgggagtgttgctccgaaatgccaacaacaacaacaacagcaacagaaactaagaaac
aaaaaatcgaaagacgaaaacaaatagaagcgactAAACTTGGCCTTTcaaagtgtGGCCGTAATTACA
GACTTCCGctagctgcaaacgcccgcccagccaaaagagagaggggagggcaaagcgagtagctgagagagaagggg
cagagagagagagagagaggagattgggggattgggttatagcagacgaatgggacaactaaAGGCGCCACAAg
GTATCCcGTTACAGGGCCATAAAacttcaataagagcaactgGCGACAAgGcCAAATTATAA
CGCATACGcACACGTAGccagagtagCTGaaGCTGAAGacccaGCTAATTGGAtgcgataaga

caaaaggggtatggcaggggtagttgataacgctgagggtgttggggtattgctgctgccacagctgacaaaacttaaaagtggg
cggcgcataacaatttgcatacaa

>D. grimshawi
cgaacaagcagacaataaatagtgacacatttgtggcaattcaacaatgcctgtctctttgctcagggctgggagagttcttcgaatt
cacacacgccaaagagaaaagacaaagaagagTTTaGCCTTTGaGGCCGTAATTACAGACTTtaact
cactggccaaaaagtgctaaacattcaggccaaaatgcagacactggccaaggaggcggggctgttgctgcgacaggcaacatc
cgacagacggcgcacgccggacgtcaGACGAatcggGATGCtgaAGGCGCCACAATGTATCCcGTT
ACAGGaCCATAAAactctctaaGCGACtGGCgacAAgGcCAAATTATAcCGCATACccACA
CGTAGaccccgtccaagttgtgcacAAGGCTAATTGGACGcggaccgctcttcatggggtatgggttggatatag
gcagatatataaatatatagggggatggggaatggttgtgcgggtcttaacttaaaagaaggcggaacataacaattgtgggtgat

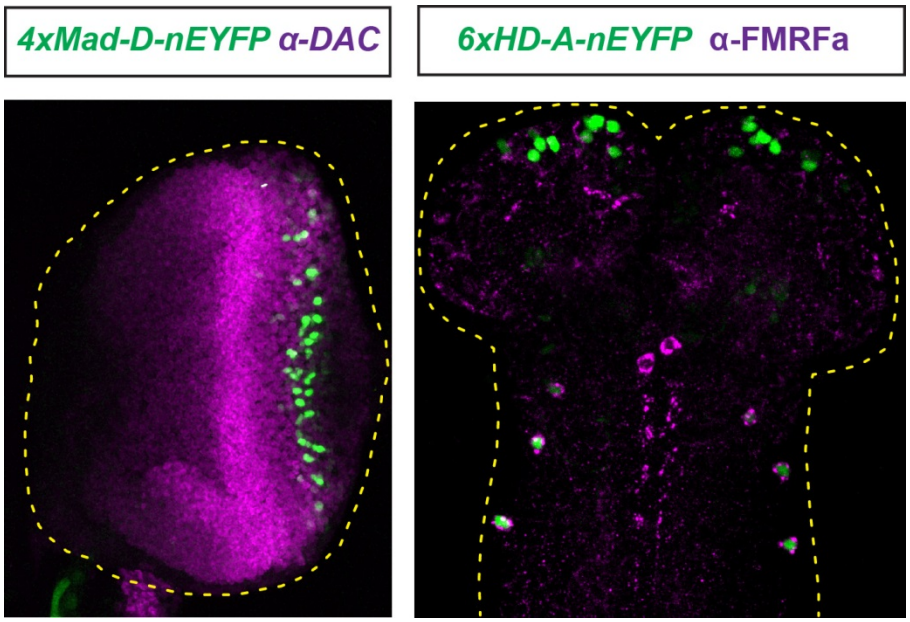


Figure S 14. Restricted and non-overlapping ectopic expression of concatemerized BMP-RE and HD-RE reporters.

(A) Expression of *4xMad-D-nEYFP* in late 3rd Instar larvae eye disc. (**B**) Expression of *6xHD-A-nEYFP* expression in the brain lobes of early L1 larva.

APPENDIX C- SUMMARY OF TV4 ENHANCER MUTATIONS

Table S 2. Summary of Tv4 enhancer mutations

Mutations in Red Restriction enzyme sites or Restriction scar sites in concatemers in Blue

Descriptor	Sequence
445 bp Tv4 FMRFa enhancer	CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCAG TGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACAA TAGCCAAATGCCATTGATGAGACTCGTCTCCCAAACTTTGGCCTTTTGCCGGGCCGTAATTACA GACTTCCGTCTTTTGAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTG GGATGGGGTGGTTTCGGGTGCTGGACGAGGTGCCAGAGGCGCCACAATGTATCCCGTTACAGGT TACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGGCAAATTATAACGCATACGGACAC GTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGGCCAGGATGTCCCTGCTCAT
del5' to MadA->End	AAACTTTGGCCTTTTGCCGGGCCGTAATTACAGACTTCCGTCTTTTGAACAGTTTTTTCAGCCC CACCCAAGAGTCGAGTCTTGAAAAGCTGGCTGGGATGGGGTGGTTTCGGGTGCTGGACGAGATG CCAGAGGCGCCACAATGTATCCCGTTACAGGTTACAGGGCCATAAAGCGCCATAAACGCCGCGA CGGCAATGGCAAATTATAACGCATACGGACACGTAGTCGATCCACTGGCTAGAAGGCTAATTGG ACGTGCCCGGCCAGGATGTCCCTGCTCAT
del5' to MadF->End	AACGCCGCGACGGCAATGGCAAATTATAACGCATACGGACACGTAGTCGATCCACTGGCTAGAA GGCTAATTGGACGTGCCCGGCCAGGATGTCCCTGCTCAT
del-MadA->MedB island (del-HD-RE)	GCCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCA GTGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACA ATAGCCAAATGCCATTGATGAGACTCGTCTCCAAAACAGTTTTTTCAGCCCCACCCAAGAGTCG AGTCTTGAAAAGCTGGCTGGGATGGGGTGGTTTCGGGTGCTGGACGAGATGCCAGAGGCGCCAC AATGTATCCCGTTACAGGTTACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGGCAAA TTATAACGCATACGGACACGTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGGCCA GGATGTCCCTGCTCAT
del-MadD island (del-BMP-RE)	CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCAG TGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACAA TAGCCAAATGCCATTGATGAGACTCGTCTCCAAAACTTTGGCCTTTTGCCGGGCCGTAATTACA GACTTCCGTCTTTTGAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTG GGATGGGGTGGTTTCGGGTGCTGGTTACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAAT GGCAAATTATAACGCATACGGACACGTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCC CGGCCAGGATGTCCCTGCTCAT
Mad-A->Mad-D	AAACTTTGGCCTTTTGCCGGGCCGTAATTACAGACTTCCGTCTTTTGAACAGTTTTTTCAGCCC CACCCAAGAGTCGAGTCTTGAAAAGCTGGCTGGGATGGGGTGGTTTCGGGTGCTGGACGAGATG CCAGAGGCGCCACAATGTATCCCGTTACA
1XHD-RE+1XBMP-RE	TTGCCGGGCCGTAATTACAGACTTCGCTAGAGGACGAGATGCCAGAGGCGCCACAATGTATCC**C** GTTACA
Mad-A->Mad-D GC content of spacer is preserved (every other base is compliment; A>T, T>A, C>G, and G>C)	AAACTTTGGCCTTTTGCCGGGCCGTAATTACAGACTTCCGTCTTTTGATCTGATATATGACCGC GAGCGATGTGACCACTGTAGTATACCAGCCAGCGTTCGCGAGCTATGGCGAGGTGGACGAGATG CCAGAGGCGCCACAATGTATCCCGTTACA
Mad-A->Mad-D noncomplementary transversion of spacer	AAACTTTGGCCTTTTGCCGGGCCGTAATTACAGACTTCCGTCTTTTGCCACTGGGGGGACTAAA ACAAACCTCTGATCTGAGGTCCCCTAGTTAGTTTCGTTTTGTTGGGATTTGTAGGGACGAGATG CCAGAGGCGCCACAATGTATCCCGTTACA
ΔMadA	CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCAG TGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACAA TAGCCAAATGCCATTGATGAGACTCGTCTCCCAAACTTTGGCCTTTTAGTAGTACGTAATTACA GACTTCCGTCTTTTGAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTG GGATGGGGTGGTTTCGGGTGCTGGACGAGGTGCCAGAGGCGCCACAATGTATCCCGTTACAGGT TACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGGCAAATTATAACGCATACGGACAC GTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGGCCAGGATGTCCCTGCTCAT

Descriptor	Sequence
ΔHD-A	CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCAG TGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACAA TAGCCAAATGCCATTGATGAGACTCGTCTCCCAAACTTTGGCCTTTTGCCGGGCCGGAGCTCCA GACTTCCGTCTTTTGAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTG GGATGGGGTGGTTTCGGGTGCTGGACGAGGTGCCAGAGGCGCCACAATGTATCCCGTTACAGGT TACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGGCAAATTATAACGCATACGGACAC GTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGGCCAGGATGTCCCTGCTCAT
ΔMedA+B	CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCAG TGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACAA TAGCCAAATGCCATTGATGAGACTCGTCTCCCAAACTTTGGCCTTTTGCCGGGCCGTAATTACG AACTTCCGTTCTTTGAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTG GGATGGGGTGGTTTCGGGTGCTGGACGAGGTGCCAGAGGCGCCACAATGTATCCCGTTACAGGT TACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGGCAAATTATAACGCATACGGACAC GTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGGCCAGGATGTCCCTGCTCAT
ΔMadD	CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCAG TGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACAA TAGCCAAATGCCATTGATGAGACTCGTCTCCCAAACTTTGGCCTTTTGCCGGGCCGTAATTACA GACTTCCGTCTTTTGAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTG GGATGGGGTGGTTTCGGGTGCTGGACGAGGTGCCAGATAGTAGACAATGTATCCCGTTACAGGT TACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGGCAAATTATAACGCATACGGACAC GTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGGCCAGGATGTCCCTGCTCAT
ΔMadE+F	CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCAG TGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACAA TAGCCAAATGCCATTGATGAGACTCGTCTCCCAAACTTTGGCCTTTTGCCGGGCCGTAATTACA GACTTCCGTCTTTTGAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTG GGATGGGGTGGTTTCGGGTGCTGGACGAGGTGCCAGAGGCGCCACAATGTATCCCGTTACAGGT TACAGGGCCATAGTAGTAGTAGTAGTAGTCGACGGCAATGGCAAATTATAACGCATACGGACAC GTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGGCCAGGATGTCCCTGCTCAT
ΔMadB	CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCAG TGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACAA TAGCCAAATGCCATTGATGAGACTCGTCTCCCAAACTTTGGCCTTTTGCCGGGCCGTAATTACA GACTTCCGTCTTTTGAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTG GGATGGGGTGGTTTCGGGTGCTGGACGAGGTGCCAGAGGCGCCACAATGTATCCCGTTACAGGT TACAGGGCCATAAAGCGCCATAAAGTAGTAGTAGTAGTAATGGCAAATTATAACGCATACGGACAC GTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGGCCAGGATGTCCCTGCTCAT
ΔHD-B	CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCAG TGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACAA TAGCCAAATGCCATTGATGAGACTCGTCTCCCAAACTTTGGCCTTTTGCCGGGCCGTAATTACA GACTTCCGTCTTTTGAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTG GGATGGGGTGGTTTCGGGTGCTGGACGAGGTGCCAGAGGCGCCACAATGTATCCCGTTACAGGT TACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGGCAAGAGCTCACGCATACGGACAC GTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGGCCAGGATGTCCCTGCTCAT
ΔHDC	CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCAG TGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACAA TAGCCAAATGCCATTGATGAGACTCGTCTCCCAAACTTTGGCCTTTTGCCGGGCCGTAATTACA GACTTCCGTCTTTTGAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTG GGATGGGGTGGTTTCGGGTGCTGGACGAGGTGCCAGAGGCGCCACAATGTATCCCGTTACAGGT TACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGGCAAATTATAACGCATACGGACAC GTAGTCGATCCACTGGCTAGAAGGCGAGCTCGACGTGCCCGGCCAGGATGTCCCTGCTCAT
ΔMadC	CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCAACAAGAGCAG TGGACACCCAATAAACCTGTTCAAAAACCCGACACATTTCTGCCCAGTCATGCGTGGTGGACAA TAGCCAAATGCCATTGATGAGACTCGTCTCCCAAACTTTGGCCTTTTGCCGGGCCGTAATTACA GACTTCCGTCTTTTGAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTG GGATGGGGTGGTTTCGGGTGCTGGACGAGGTGCCAGAGGCGCCACAATGTATCCCGTTACAGGT TACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGGCAAATTATAACGCATACGGACAC GTAGTCGATCCACTGGCTAGAAGGCTAATTGGAAGTAGTAGTAGTAGATGTCCCTGCTCAT
6xHD-RE	TTGCCGGGCCGTAATTACAGACTTCGCTAGATTGCCGGGCCGTAATTACAGACTTCGCTAGATT GCCGGGCCGTAATTACAGACTTCGCTAGATTGCCGGGCCGTAATTACAGACTTCGCTAGATTGC CGGGCCGTAATTACAGACTTCGCTAGATTGCCGGGCCGTAATTACAGACTTCGCTAGA

Descriptor	Sequence
4XSpacer element	CTAGAAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTGGGATGGGGTG GTTTCGGGTGCTGCTAGAAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGG CTGGGATGGGGTGGTTTCGGGTGCTGAATTCAACAGTTTTTTCAGCCCCACCCAAGAGTCGAGT CTTGAAAAGCTGGCTGGGATGGGGTGGTTTCGGGTGCTGCTAGAAACAGTTTTTTCAGCCCCAC CCAAGAGTCGAGTCTTGAAAAGCTGGCTGGGATGGGGTGGTTTCGGGTGCT
4xBMP-RE	GGACGAGATGCCAGAGGCGCCACAATGTATCCCGTTACAGCTAGAGGACGAGATGCCAGAGGCG CCACAATGTATCCCGTTACAGCTAGAGGACGAGATGCCAGAGGCGCCACAATGTATCCCGTTAC AGCTAGAGGACGAGATGCCAGAGGCGCCACAATGTATCCCGTTACA
4XMad-E+F	GGTTACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGCTAGAGGTTACAGGGCCATAA AGCGCCATAAACGCCGCGACGGCAATGCTAGAGGTTACAGGGCCATAAAGCGCCATAAACGCCG CGACGGCAATGCTAGAGGTTACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAAT
4XMadF->End	AACGCCGCGACGGCAATGGCAAATTATAACGCATACGGACACGTAGTCGATCCACTGGCTAGAA GGCTAATTGGACGTGCCCGGCCAGGATGGCTAGAAACGCCGCGACGGCAATGGCAAATTATAAC GCATACGGACACGTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGGCCAGGATGGC TAGAAACGCCGCGACGGCAATGGCAAATTATAACGCATACGGACACGTAGTCGATCCACTGGCT AGAAGGCTAATTGGACGTGCCCGGCCAGGATGGCTAGAAACGCCGCGACGGCAATGGCAAATTA TAACGCATACGGACACGTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGGCCAGGA TG