CIS-REGULATORY INTEGRATION OF INTRINSIC TRANSCRIPTION FACTORS WITH TARGET-DERIVED SIGNALS IN NEURONAL DIFFERENTIATION

by

Chung Yiu Jonathan Tang

B.Sc., The University of British Columbia, 2007

A THESIS SUMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

The Faculty of Graduate Studies

(Cell and Developmental Biology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

August 2009

ABSTRACT

We now know that expression of the appropriate terminal differentiation genes (TDG) , such as neuropeptides, neurotransmitters, ion channels, et cetera, in maturing neurons is controlled by cell-specific combinations of transcription factors and by signals secreted from the neurons' target cells. However, it is unclear how these two regulatory inputs are integrated inside neurons. In *Drosophila*, target-derived Bone Morphogenetic Protein (BMP) signals and a well-characterized combinatorial code of transcription factors activate expression of the *FMRFa* gene in the Tv neurons. Here, I performed a *cis*-regulatory analysis of *FMRFa* in order to understand how the two factors functionally intersect. Mutant analysis reveals that 4 of the 7 known *FMRFa* regulators, Apterous, BMP signalling, Dachshund and Zfh1, all regulate the expression of the Tv enhancer, a 446 bp *cis*-regulatory element that faithfully reproduces *FMRFa* expression in the Tv neurons. Within the Tv enhancer, I identified a functional module, termed HD/BRE-A, that is predicted to respond to both BMP signalling and the homeodomain transcription factor Apterous. I also verified that Apterous and the BMP factors, Mad and Medea, can directly bind to HD/BRE-A in *vitro*. Furthermore, transgenic analyse indicate that the positioning between the Apterous and Medea binding site is critical for Tv enhancer activation, suggesting a strict physical requirement for simultaneous association of these factors. Taken together, my results supports a model of *cis*-regulatory integration of BMP signaling and homeodomain transcription factors in the regulation of TDGs.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGEMENTS

DEDICATION

Dedicated to my parents, Joe and Mary.

1. Introduction

1.1. Neuronal specification of BMP-dependent genes in the *Drosophila* nervous system

1.1.1. Gaps exist in our knowledge of neuronal terminal differentiation

Neuronal identity is marked by the gene expression profiles of neurons. When a neuron exits the cell cycle, it has been fated to become a specific neuronal subtype, but it has not completely achieved terminal differentiation, as defined by the expression of genes important for mature neuronal function. These genes, hereafter referred to as terminal differentiation genes (TDGs), include neuropeptides, neurotransmitters, neurotransmitter biosynthetic enzymes, ion channels, et cetera. In general, TDGs mediate neuronal communication through their influence on synaptic transmission, neuropeptide/neurotransmitter processing and secretion, etc. The activation of these genes occurs throughout the life of a neuron and depends on the neuron's intrinsic complement of transcription factors (TFs), as well as extrinsic signals derived from target cells. The integration of these two inputs results in the cell-specific expression of TDGs in the nervous system. Despite this knowledge, we know very little about <u>how</u> these two inputs integrate to specify TDG expression. Since extrinsic target-derived signalling and intrinsic transcription factors primarily regulate gene activity at the transcriptional level, a *cis*-regulatory analysis of target genes may reveal general mechanisms of TDG specification in post-mitotic neurons.

1.1.2. Significance

To date, almost nothing is known about how post-mitotic neurons convert target-derived signals into specific profiles of TDG expression. It is crucial to gain a better understanding of this process because many brain diseases arise from defects in synaptic connections. These defects can in turn result in the dysregulation of TDGs. For example, the expression of neuropeptides and neuropeptide-processing enzymes are disrupted in the brains of Alzheimer's disease patients (Saito et al., 2005; Saito et al., 2003). Likewise, the onset of Schizophrenia has been linked to abnormal changes in the expression of neuropeptides and neurotransmitters (Boules et al., 2007; Lewis and Levitt, 2002; Stephan et al., 2006). Thus, by studying the interaction between TFs and target-derived signals in their regulation of TDGs, we can gain

1

insights into the mechanism behind dysregulation of TDG expression. This new knowledge can be applied to the design of preventive and therapeutic treatments of many brain diseases.

1.2. Background/Literature Review

1.2.1. Neuronal terminal differentiation

The acquisition of entire sets of TDGs is a prolonged process that occurs over the lifetime of a neuron and depends on intrinsic and extrinsic factors (Edlund and Jessell, 1999; Hippenmeyer et al., 2004; Koo and Pfaff, 2002). Intrinsically, TFs, inherited from neuronal precursors or activated at the time of a neuron's birth, can dictate a neuron's differentiation program. TFs are DNA binding proteins involved in gene regulation. They act by binding to specific sequences, or motifs, in the *cis*-regulatory regions of genes. This in turns lead to the activation, de-repression or silencing of genes, depending on the molecular context. Thus, the cell-specific expression of many neuronal genes can be attributed to the selective activation of TFs in subset of neurons (Hobert and Westphal, 2000; Hunter and Rhodes, 2005; Jessell, 2000; Shirasaki and Pfaff, 2002). Homeodomain-type TFs play a particularly critical role in this process (Briscoe and Novitch, 2008; Dasen et al., 2005; Landgraf and Thor, 2006; Shirasaki and Pfaff, 2002; Skeath and Thor, 2003; Thaler et al., 2004). A simplistic view of TDG regulation in post-mitotic neuron comes from studies in *C.*elegans. In this organism, which has only 302 neurons, only one or two TFs are required for a neuron's entire repertoire of TDGs (Etchberger et al., 2007; Wenick and Hobert, 2004). However, in higher organisms such as Drosophila (~300,000 neurons) and mammals (billions of neurons), combinations of 5-10 TFs may be required to activate a TDG in specific cells (Allan et al., 2005; Allan et al., 2003; Jorgensen et al., 2004). Some of these TFs may cluster into modules at the *cis*-regulatory regions of TDGs, thereby increasing the specificity of gene expression in the nervous system (Jorgensen et al., 2004). Others TFs may control TDG expression through indirect ways, such as through the control of axon pathfinding, and the ability to sense target-derived signals (Allan et al., 2003).

Besides TFs, TDG activation can also require extrinsic signals derived from target cells. These retrograde signals help to shape a neuron's properties to ensure functional compatibility between pre- and post-synaptic cells. Signalling molecules such as TGF-β/Bone Morphogenetic Proteins (BMP), neurotrophins, and cytokines are all target-derived signals important for the

proper differentiation of maturing neurons (Allan et al., 2003; Ernsberger and Rohrer, 1999; Hippenmeyer et al., 2004; Nishi, 2003; Xu and Hall, 2006). They act via the activation of signalling pathways which affect many gene regulatory mechanisms in a cell. Notably, the same signalling pathway may control the transcriptional activation of different TDGs in different neuronal contexts, suggesting interplay between target-derived signalling and cell-specific transcription factor codes (Coulombe and Kos, 1997; Coulombe and Nishi, 1991; Pavelock et al., 2007). Thus, a key to understanding the diversification of TDG expressions in the nervous system is the mechanistic knowledge of the interplay between intrinsic and extrinsic factors.

1.2.2. *Drosophila* as a model to study neuronal terminal differentiation

Attempts to address these problems require manipulation of signalling pathways and TF expression. This has been met with many technical problems in vertebrates, as mutations of TFs and signalling components (pertinent to this study) often result in embryonic lethality or aberrant specification of neuronal populations (Harrison et al., 1999; Lechleider et al., 2001; Monuki et al., 2001; Oppenheim, 1989; Pfaff et al., 1996; Sheng et al., 1996; Sirard et al., 1998; Sofroniew et al., 2001; Thaler et al., 1999; Tremblay et al., 2001; Yang et al., 1998). *Drosophila melanogaster* is an ideal model organism for this objective because disruption of signalling pathways and TFs pertinent to this study do not always lead to embryonic lethality or neuronal lineage re-specification (Allan et al., 2005; Allan et al., 2003; Benveniste et al., 1998; Miguel-Aliaga et al., 2004). In cases where embryonic lethality does occur, measures can be taken to allow for embryo survival up to the developmental stage of interest.

Drosophila possesses a much simpler nervous system than mammals, making it easier to study individual neurons. To illustrate, the late embryonic/larval Drosophila ventral nerve cord (VNC) consists of ~10,000 neurons, whereas mammalian nervous systems are estimated to have between millions to billions of neurons. Moreover, the genes and mechanisms of neuronal terminal differentiation are well-conserved between invertebrates and vertebrates (Arendt, 2005; Arendt and Nubler-Jung, 1996, 1999; Thor and Thomas, 2002), including the role of target-derived signals Bone Morphogenetic Protein (BMP) signalling (Allan et al., 2003; Hodge et al., 2007; Nishi, 2003) (Figure 1.1). Furthermore, the availability of a wide variety of genetic tools and databases make it very convenient to study individual neurons in *Drosophila*. To illustrate, a large collection of GAL4 drivers is available in the fly community. These drivers allow for the

cell-specific expression of UAS-transgene or RNAi in overexpression or knockdown experiments (Brand and Perrimon, 1993; Dietzl et al., 2007; Duffy, 2002). Lastly, the recent advent of an optimized transgenesis system utilizing the phiC31 integrase has reduced the time and workload needed to generate numerous transgenic flies (Bischof et al., 2007). This system allows for site-specific integration of transgenes into genomic DNA, making it ideal for direct comparison of enhancer-reporter variants.

1.2.3. Neuronal diversity mediated by single target-derived signals

One of the key questions surrounding neuronal differentiation has to do with how a common target-derived signal induces different responses in different neurons. A survey across vertebrate species showed that signals conducted through the same pathway can be interpreted in different ways by different neurons (Coulombe and Kos, 1997; Coulombe and Nishi, 1991; Pavelock et al., 2007). This apparent dilemma may be explained by differences in the complement of TFs that are expressed in different neurons. However, it remains a mystery as to how neuron-specific TFs integrate with the same target-derived signals to turn on different TDGs. To address this issue, our laboratory has recently identified numerous BMP-dependent TDGs that are expressed in different subsets of neurons, and regulated by different combinations of cell-specifically-expressed TFs.

1.2.4. The BMP pathway

BMP ligands are well-conserved members of the TGF-β family of signalling molecules (Schmierer and Hill, 2007). In neurons, BMPs are important for intercellular communication, synaptic growth, and neurotransmitter release in vertebrates (Ai et al., 1999; Guha et al., 2004; Hall et al., 2002; Hodge et al., 2007; Lee-Hoeflich et al., 2004; Pavelock et al., 2007; Shen et al., 2004; Sun et al., 2007; Withers et al., 2000; Xu and Hall, 2006) and in *Drosophila* (Aberle et al., 2002; Allan et al., 2003; Eaton and Davis, 2005; Marques et al., 2002; McCabe et al., 2003; Rawson et al., 2003). Canonical transduction through the BMP signalling pathway is well characterized and well conserved (Fig.1.1) (Affolter et al., 2001; Massague and Gomis, 2006; Massague et al., 2005; Massague and Wotton, 2000; Schmierer and Hill, 2007; Shi and Massague, 2003). In *Drosophila* neurons, BMP signalling is initiated when the target-derived BMP ligand Glass bottom boat (Gbb) binds to the type-II BMP receptor Wishful Thinking (Wit)

(Allan et al., 2003). This ligand-receptor interaction results in the recruitment and phosphorylation of the type-I BMP receptor Saxophone (Sax) or Thickveins (Tkv). The type-I receptor then proceeds to phosphorylate the cytoplasmic Smads, Mothers against decapentalplegic (Mad), to generate phosphoMad (pMad) (Fig. 1.1). Phosphorylation of Mad is believed to expose its nuclear localization signal and allow for nuclear localization through interactions with Importin-B (Xiao et al., 2000). pMad can interact with the common-non-phosphorylated Smad, Medea, to form the pMad/Medea complex. The pMad/Medea complex then translocates into the nucleus, where it elicits transcriptional response. pMad/Medea acts as a TF, which binds Mad (GRCGNC) and Medea (GTCT) binding motifs (together called BMP response element, or BRE) (Gao and Laughon, 2007; von Bubnoff et al., 2005). pMad and Medea are well described to bind DNA somewhat weakly and often require interactions with additional TFs (Affolter et al., 2001; Attisano and Wrana, 2000; Gao and Laughon, 2007; Massague and Wotton, 2000; Shi and Massague, 2003; Wotton and Massague, 2001). Such interactions not only increase DNA binding affinity but also increase target gene specificity. Therefore, it is conceivable that the BMP signal may be diversified by the interaction of pMad/Medea with different TFs in different neurons.
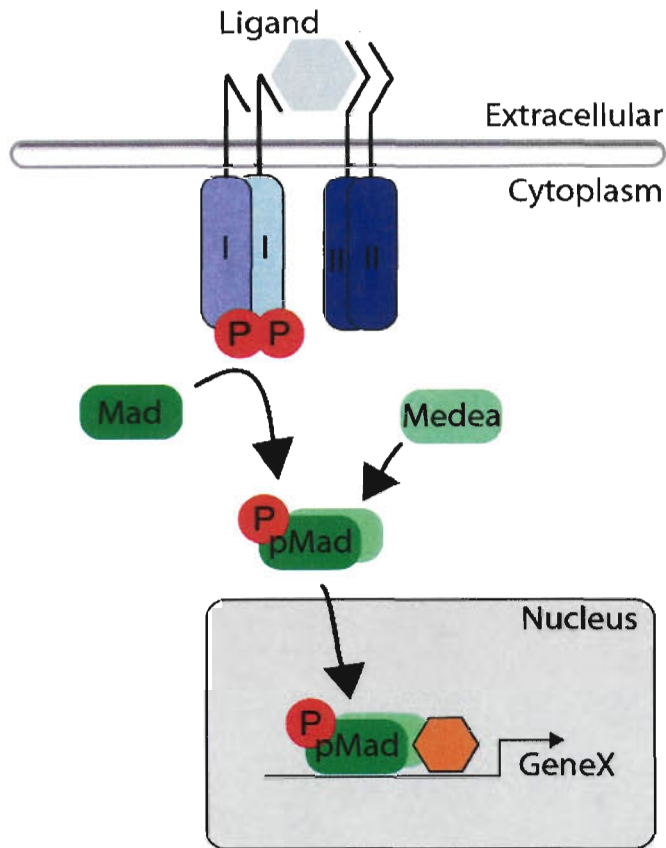
Figure 1.1. The BMP signalling pathway. A BMP ligand is bound to the tetrameric receptor complex comprised of Type II and Type I BMP receptors. Upon binding to the ligand, the type II receptor phosphorylates the Type I receptor, which then goes on to phosphorylate Mad, generating pMad. Phosphorylation exposes the NLS of Mad and allows for its nuclear localization through interactions with Importin-B. pMad then binds to Medea to become pMad/Medea, which is translocated into the nucleus and induces transcriptional response.

1.2.5. Functional intersection between HD TFs and the BMP/TGF-β pathway

Homeodomain-containing (HD) TFs make up a family of DNA-binding proteins that is involved in all stages of developmental hierarchy in a wide range of species, and notably play a critical role in nervous system development (Briscoe and Novitch, 2008; Dasen et al., 2005; Landgraf and Thor, 2006; Shirasaki and Pfaff, 2002; Skeath and Thor, 2003; Thaler et al., 2004). Common to this family of proteins is an evolutionarily conserved DNA-binding domain that generally binds to the TAATNN consensus sequence (Berger et al., 2008; Noyes et al., 2008). Amongst HD TFs, additional specificity is conferred by their differential preferences for variants of this basic HD-binding sequence (Berger et al., 2008; Noyes et al., 2008; Svingen and Tonissen, 2006).

Recent studies have pointed to a role for HD TFs as co-factors for the BMP pathway (Brugger et al., 2004; Faresse et al., 2008; Grocott et al., 2007; Lamba et al., 2008; Li et al., 2006; Suszko et al., 2008; Walsh and Carroll, 2007; Zhou et al., 2008). Enhancer analysis on the BMP-dependent *msx2* gene in mouse showed that a phylogenetically conserved sequence containing closely associated BRE and HD sequences was able to induce BMP-dependent gene expression not only in its host species, but also in *Drosophila* (Brugger et al., 2004). Similarly, a tight coupling of BRE and HD was also found in the BMP-dependent *sal* gene in *Drosophila* (Walsh and Carroll, 2007). The spacing between BRE and HD is crucial for function, as demonstrated by the loss of reporter activity when these two motifs are physically separated (Walsh and Carroll, 2007). Using a series of bandshift assays and loss-of-function experiments, Walsh and Carroll (2007) further showed that the *sal* regulator *Ubx*, a HD TF, can indeed bind to the *sal* enhancer through the aforementioned HD motif. Taken together, these findings suggest an evolutionarily conserved partnership between homeodomain and BMP (via pMad and Medea) signalling in gene regulation. However, it is still unclear what the relationship is between these two TFs in BMP-dependent neuronal differentiation, and which HD factors may interact with BMP signals in the terminal steps of neuron-specific differentiation.

1.2.6. *FMRFa*, a well-characterized BMP-dependent neuropeptide

FMRFa was the first BMP-dependent neuropeptide found in the *Drosophila* ventral nerve cord (VNC) (Allan et al., 2003). *FMRFa* is notably expressed in a group of 6 neurons termed the Thoracic ventral (Tv) neurons (Fig.3.2) (Schneider et al., 1993). At embryonic stage 17, the Tv neurons establishes axonal connection with the neurohemal organ, which in turns secretes the retrogradely-transported BMP-ligand Gbb to activate the expression of *FMRFa* in the Tv neurons. *FMRFa* is regulated by a unique Tv neuron-specific TF code that includes the LIM-HD TF *apterous* (*ap*), the basic Helix-loop-Helix TF *dimmed* (*dimm*), *dachshund* (*dac*), *eyes absent* (*eya*) the Zinc finger TF *squeeze* (*sqz*) and the zinc-finger homeodomain TF *zinc finger homeodomain 1* (*zfh1*) (Allan et al., 2005; Allan et al., 2003; Miguel-Aliaga et al., 2004; Vogler and Urban, 2008) (Fig. 1.2). Amongst these TFs, *apterous* is of particular interest because it is the only TF in the '*FMRFa* combinatorial code' that is required in all TF combinations known to induce ectopic *FMRFa* expression. In addition, these TF combinations fail to trigger ectopic *FMRFa* expression in a *wit* mutant background, suggesting a requirement

7

for BMP signalling in Apterous activation of *FMRFa* (Allan et al., 2005; Allan et al., 2003; Miguel-Aliaga et al., 2004). Whether there is a connection between BMP-signalling and any of the intrinsic *FMRFa* regulators remain unknown.

1.2.7. Tv enhancer

Previously, a 446 bp *FMRFa* enhancer was identified to be sufficient for Tv-neuron expression of FMRFa (Benveniste et al., 1998), hereafter referred to as the Tv enhancer. Loss-of-function genetic experiments indicate that *apterous* and *zfh1* can regulate expression of the Tv enhancer (Benveniste et al., 1998; Vogler and Urban, 2008). In the Tv enhancer, three predicted HD binding sites are bound by Apterous in *vitro* and at least two of them are essential for reporter activity in *vivo*. In this work, these sites will be denoted as HD-A, HD-B and HD-C, in accordance with previous work (Benveniste et al., 1998). Despite this understanding, it is unknown whether the Tv enhancer is also regulated by target-derived BMP signalling, or by other TFs of the *FMRFa* combinatorial code.
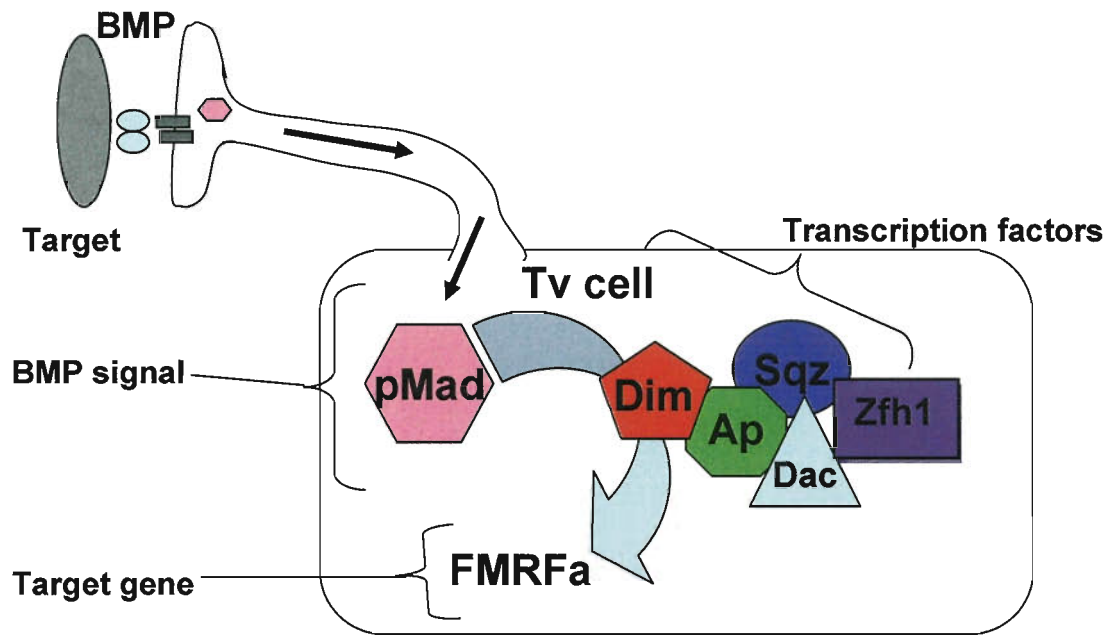
Figure 1.2. Neuronal identity is determined by intrinsic and extrinsic factors. Here, the target-derived signal (BMP ligand, light blue circle) is converted into a signal transducer (pMad) in the Thoracic ventral (Tv) neuron. This signal transducer translocates into the nucleus, where it integrates with a combinatorial TF code (Dimmed (Dim), Ap (Ap), Squeeze (Sqz), Zfh1 and Dachshund (Dac) to turn on the neuropeptide *FMRFa*. Details of pMad formation and translocation in described in Figure 1.1.

1.3. Rationale/Hypothesis/Objectives

FMRFa is the best characterized target-dependent TDG in any model system. Thus, it is most ready for an in-depth *cis*-regulatory analysis to understand how TDGs are controlled by target-derived signalling and intrinsic transcription factors. Since BMP signalling and *FMRFa* intrinsic transcription factor codes primarily regulate gene expression at the transcriptional level, I propose that they integrate at the *cis*-regulatory regions of *FMRFa*. Notably, I propose that the homeodomain transcription factor Apterous is a key integrator with BMP signalling due to its central role in *FMRFa* activation and its ability to directly bind the Tv enhancer. The Tv enhancer provides a convenient tool to test our ideas because it narrows down *FMRFa* regulation in the Tv neurons to a 446 bp DNA fragment.

In this thesis, I hypothesize that BMP signalling acts directly with Apterous at the Tv enhancer to cell-specifically regulate the expression of *FMRFa*.

My project aimed to:

1.) Determine whether BMP signalling regulates the Tv enhancer.

2.) If so, identify the BMP-responsive elements in the Tv enhancer and their relationship to Apterous-responsive elements.

3.) Examine the ability of the BMP-activated molecules to associate with the Tv enhancer, and with Apterous.

2. Materials and methods

2.1. Fly Genetics

The following fly stocks were used: w^{1118}; **apterous**: ap^{md544} (a P element P{GawB} insertion 5bp upstream of the *apterous* promoter, a strong hypomorphic allele referred to here as ap^{GAL4} (O'Keefe et al., 1998)), ap^{P44} (amorphic *apterous* allele (Bourgouin et al., 1992)), *UAS-apterous* (Allan et al., 2003); **wishful thinking**: wit^{A12} (null allele, nonsense point mutation before transmembrane domain (Marques et al., 2002)); wit^{B11} (point mutant that acts as a genetic null (Marques et al., 2002)); **dachshund**: dac^{3} (amorphic allele due to 25kb insertion (Tavsanli et al., 2004)), *Df(2L) Exel7086* (a *dac* deficiency, here referred to as dac^{Df}), *UAS dac* (Miguel-Aliaga et al., 2004); **eyes absent**: eya^{E1} (nonsense mutation that acts as strong hypomorphic allele (Bui et al., 2000), $eya^{Cli-IID}$ (nonsense mutation that acts as a genetic null (Bui et al., 2000); **zfh1**: zfh^{00865} (a P element P{PZ} insertion 52bp upstream of the promoter of the longest transcript that is known to be a protein null in Tv neurons (Justice et al., 1995; Vogler and Urban, 2008)); **Enhancer trap line $OK6^{GAL4}$** that expresses GAL4 in most if not all motorneurons (Aberle et al., 2002); **Tv-enhancer $FMRFa$-$lacZ$** transgenic insert pWF-E17 (Benveniste et al., 1998); **Tv-$nEYFP$** wildtype and mutant reporter flies were injected into flies bearing an attP acceptor site at P2 (Chr 3) by Genetics Services Inc, Cambridge, MA; (www.geneticservices.com). attP2 was found to provide optimal reporter activity over two other sites, attP16 and attP40 (see Appendix A). The integrase lines were maintained as homozygous stocks.

Mutants were kept over *CyO, Actin-GFP* or *TM3, Ser, Actin-GFP* balancer chromosomes. w^{1118} was typically used as a control. All crosses were maintained at 25^{0}C on standard cornmeal food.

2.2. Molecular Biology/Transgene Construction

2.2.1. ***pHS mCherry-nls attB*** was constructed in the following way: The loxP-UAS-MCS-SV40polyA cassette was excised from the pUAST attB vector (Bischof et al., 2007) by NheI and SpeI sites, and the backbone overhangs were filled-in using Klenow fragment. The blunt-ended vector was then ligated with an EcoRV-loxP-MCS-hsp70TATA-nuclear mCherry-early SV40

polyA-attB-SwaI cassette, made as described below. The MCS-hsp70TATA-nuclear mCherry-early SV40 polyA portion of the pH mCherry vector (C.Y.J.T, Allan lab; generated from pHStinger (Barolo et al., 2004)) was fused to the attB portion of pUAST attB using splicing by overlap extension (Horton et al., 1990) Amplification of the fused product was done using a 5' primer carrying an EcoRV-loxP-MluI overhang, and a 3' primer containing a SwaI-ZraI overhang. Several improvements were made to the loxP/attB backbone to facilitate a more diverse range of cloning tasks. First, three unique restriction sites were introduced: 1.) a MluI site was placed between the loxP and MCS region. 2.) an AvrII site was engineered between the SV40 early polyadenylation signal and attB sequence. 3.) a ZraI site was placed immediately 3' of the attB sequence. Second, three commonly used restriction sites, BamHI, NheI and SpeI were turned into unique sites in this new vector. Notably, SpeI is now positioned immediately 5' of the SV40 early polyadenylation signal, allowing for convenient exchange of reporters when used in concert with the unique AgeI site, located immediately 5' of the mCherry gene.

Primers for generating EcoRV-loxP-MCS-hsp70TATA-nuclear mCherry-early SV40 polyA-attB-SwaI cassette are as follows:

1st segment: 5'-GTGTGTGTGCTAGCATAACTTCGTATAATGTATGCTATACGAAGT
TATACGCGTGCATGCTGCAGCAGATCTGGTCT - 3'
5' – ATACATTGATGAGTTTGGACAAACC -3'

2nd segment: 5' – GTCCAAACTCATCAATGTATCCTAGGGTCGACGATGTAGGTCAC
GG-3'
5' – CCCCATTTAAATGACGTCGTCGACATGCCCGCCGTGAC – 3'

Fused product: 5' – CGCCGGGGATATCATAACTTCGTATAATGTATG – 3'
5' – CCCCATTTAAATGACGTCGTCGACATGCCCGCCGTGAC

2.2.2. *Tv^{wt}-nEYFP* and *Tv^{mutX}-nEYFP attB*. The 446 bp Tv *cis*-regulatory fragment was amplified from genomic DNA (Oregon R) and engineered to have XbaI and EcoRI restriction sites on the 5' and 3' ends, respectively. This fragment was cloned into XbaI and EcoRI digested pHS mCherry-nls attB, generating pHS Tv-mCherry.nls attB. However, this reporter was found to be weakly expressed in transgenic flies. Thus, mCherry.nls was excised from the pHS-Tv-mCherry-nls-attB vector through AgeI and SpeI restriction sites, and the remaining backbone was ligated to an nEYFP.nls fragment carrying 5'-AgeI and 3'-SpeI overhangs, generating pHS-Tv-nEYFP-nls-attB, here referred to as *Tv^{wt}-nEYFP*. Subsequent mutant

constructs, Tv^{mutX}-*nEYFP,* were all cloned into this construct via replacement of the wildtype Tv sequence through XbaI and EcoRI restriction sites. All Tv mutant fragments were generated by PCR-based mutagenesis. $Tv^{mMadA/B/C}$ was the first mutant generated and conveniently served as the template for many single- and double- mutant combinations. Primers used to generate all mutants are listed in Appendix C, Table C.2.

2.2.3. *Insulated Tv-nEYFP.nls attB.* Using pStinger (Barolo et al., 2004) as a PCR template, the 5' insulator element was amplified and engineered to carry 5'-MluI and 3'BglII restriction sites while the 3' insulator element was amplified and engineered to carry 5'-AvrII and 3'-SpeI restriction sites. The MluI/BglII digested 5' insulator PCR product was cloned into the MluI/BglII restriction sites in pHS-Tv-nEYFP-nls-attB. The resultant vector was then digested with AvrII and ligated with the AvrII/SpeI digested 3' insulator PCR product. Both insulator elements were inserted in the same orientation relative to the MCS-reporter cassette as that in pStinger.

Primers for insulator elements

5' insulator:	5' – TTTTACGCGTATGCATCACGTAATAAGTGTGCG – 3'
	5' – TTTTCCAGATCTGCTGCAGCATGC – 3'
3' insulator:	5' – TTTTCCTAGGCACGTAATAAGTGTGCGT – 3'
	5' – GGGGACTAGTAATTGATCGGCTAAATGG – 3'

2.2.4. Tv-mCherry in pGL3 for S2 cell transfection – To construct the Tv-mCherry.nls reporter, the ~1.1 kb Tv-hsp70TATA-mCherry.nls fragment was amplified from pHS-Tv-mCherry (unpublished CYJT, Allan Lab) and engineered to be flanked by HindIII/XbaI restriction sites. This fragment was used to replace the HindIII-*luc*-XbaI fragment in pGL3, generating pGTv-nmC. To generate the Tv mutant mCherry constructs, mutant fragments were excised from the pHS Tv mutant nEYFP-attB vector via XbaI/Xho I restriction digest while the wildtype Tv fragment was excised from the pGTv-mC via NheI/XhoI restriction digest. The XbaI-Tv mutant-XhoI fragment is then ligated to the NheI-pHS-NEYFP-attB-XhoI backbone, producing the corresponding Tv mutant mCherry vectors. Wildtype and mutant Tv *cis*-regulatory fragments were excised from their respective nEYFP-attB vectors by digestion with XbaI/XhoI, and cloned into the pG-mCherry.nls.

2.2.5. HA-Ap in pAct5.1 – The HA-tagged Apterous cDNA construct for expression in S2 cells was made from cloning a KpnI/XbaI fragment excised from pCDNA3M-HA-Ap (Linda Jurata) into the KpnI/XbaI restriction sites in the pAct vector (a gift from Eric Jan, UBC).

2.2.6. GST LIMless Ap in bacterial vector – In order to minimize interference from the LIM domain on Ap DNA binding in vitro, a truncated, LIM-less Ap coding sequence (Benveniste et al., 1998) flanked by EcoRI/BamHI sites was amplified by PCR from pCDNA3M-HA-Ap (Linda Jurata) and cloned into pGEX-2TK via the EcoRI/BamHI restriction sites.

2.3. EMSA/Protein Synthesis

GST-MadN, GST-LIMless Ap and MBP-MedN were expressed in Rossetta bacteria cells, purified under non-denaturing conditions and dialyzed into storage buffer of 20mM HEPES pH 7.8 (25.6 degrees), 50mM KCl, 1mM DTT, and 10% glycerol. Aliquots were stored in -80 degrees until use.

EMSA probes were made from annealing of complementary single stranded oligonucleotides, followed by end-labelling with gamma P32 ATP (Amersham Biosciences) using T4 Kinase. Probes were purified using Centrispin-20 separation columns (Princeton Separations) and counted for radioactivity. Rabiolabelled probes (50,000 cpm) were incubated with purified proteins in a solution containing 20mM HEPES pH 7.8, 50mM KCl, 1mM DTT, 1mg/ml BSA, 0.25mM EDTA. As an exception, 50ng poly dIdC-dIdC was included in EMSA reactions involving HA-LIMLess Ap to reduce non-specific binding. After 30 minute incubation at room temperature, samples were loaded onto a pre-run 5% polyacrylamide gel at 100V, and then electrophoresed at 200V at room temperature for approximately 25 minutes in 0.5X TBE buffer. Gels were exposed to phosphor screens (Molecular Dynamics) and imaged on phosphoimager (Typhoon). All probe sequences used are listed in the appendix.

2.4. Immunohistochemistry/Confocal Imaging
2.4.1. Immunohistochemistry protocol

The following antibodies were used: anti-FMRFa (1:500) (Peninsula Laboratories Inc.), anti-β-Galactosidase clone 40-1a (1:100), anti-Eyes absent (1:100) (both from the

Developmental Studies Hybridoma Bank). Immunolabeling was carried out using standard protocols as previously described (Allan et al., 2003).

2.4.2. Confocal imaging and analysis procedure

All images were acquired on an Olympus FV1000 confocal microscope as multiple TIFF files representing individual Z-stacks. All images for comparison were taken from tissues that were processed simultaneously, mounted on the same slide, and imaged with identical confocal settings, using Nyquist optics and setting the high/low so as not to saturate the most intense fluorescent pixels. To count numbers of Tv neurons expressing Tv-nEYFP, we only included VNC's in which all six Tv neurons were identified (using anti-FMRFa immunoreactivity). Of these, we analyzed Z-stacked images and counted the number of nEYFP-expressing Tv neurons by eye. For measurement of fluorescence intensity, we imaged every identified Tv neuron (using anti-FMRFa immunoreactivity) in each CNS collected. Raw files were imported into Image J (US National Institutes of Health) for analysis. For each Tv neuron, we compressed all Z-slices spanning whole Tv neurons using the Z-projector function, set to sum the pixel intensities from each Z-slice. Each Tv neuron was outlined and the mean of the summed pixel intensity for each neuron was measured. Background fluorescence intensity was corrected for by subtracting a directly adjacent region of background (of equal size to the Tv neuron) from the same summed Z-stack for each Tv neuron. The resulting value for each Tv neuron was then incorporated as a single datum point towards the mean fluorescence intensity for each experiment. To normalize data across multiple time points and genotypes, we further expressed each datum point as a percentage of the mean of the Tv^{wt}-$nEYFP$ control for that experiment.

2.5. Co-immunoprecipitation

One and a half million S2 cells were transiently transfected with 0.3ug per plasmid type of pAct5.1-HA-Ap, pAct5.1-FLAG-Mad, pAct5.1-myc-Medea and pAc5.1-TkvQD (generous gifts from Konrad Basler) in a 6 well plate using Effectene (Invitrogen). Forty-eight hours after transfection, cells were lysed on ice for 30 minutes in 50mM Tris-HCl pH 8.0, 150mM NaCl, 1% NP40, 2mM EDTA plus protease inhibitors and Phosstop phosphatase inhibitors (Roche). After pelleting of cell debris, the supernatants were isolated and pre-cleared with protein A/G sepharose beads for 1 hour at 4 degrees Celsius, 1000ug of cell lysate was then used for

immunoprecipitation of HA-Apterous with 0ug or 2ug of rabbit anti-HA (ab9110, AbCam) overnight at 4 degrees Celsius. On the second day, protein A/G sepharose beads were added to immunoprecipitated antibody-antigen complexes for 2 hours at 4^0C. This was followed by washes in cell lysis buffer and then boiling of beads to elute immunoprecipitated products in 2x SDS loading buffer. Samples were then loaded for western blot analysis, using 1:500 rat anti-HA (3F10, Roche), 1:1000 mouse anti-FLAG (M2, Sigma), 1:500 mouse anti-myc (9E10, Sigma).

2.6. Bioinformatics

The Tv-enhancer sequence was derived from Release r5.18 (May 29, 2009) of the *D.melanogaster* genome (downloaded from FlyBase) and is situated at Chromosome 2R: 5792874..5793318. Evoprinter was utilized to rapidly identify the top three non-overlapping BLAT alignments of the *D.melanogaster* Tv-enhancer sequence in the genomes of all 12 sequenced *Drosophilid* species (Odenwald et al., 2005; Yavatkar et al., 2008). For each eBLAT that aligned to the *D.melanogaster* Tv enhancer piece, we performed both a BLAST search (in Flybase) and a BLAT search (in UCSC Browser) of that sequence on the pertinent species' genome. We then examined whether that sequence was near the FMRFa gene for that species. For every species, only the 1st BLAT was in the vicinity of the FMRFa for each genome, and in all cases was found 5' of the FMRFa gene. These are all shown in Appendix G. To account for potential annotation failures of additional FMRFa genes, that may lie within the region of eBLAT2 or 3 sequences, we performed TBLASTN 2.2.21 (NCBI), using the *D.melanogaster* FMRFa full length amino acid sequence as a query, to uncover additional FMRFa genes. The repetitive nature of this polypeptide (with multiple FMRFGR/K motifs) makes it simple to verify true versus false low identity "hits". In every case, only one FMRFa gene was found in each genome and BLAT2/3 sequences were not found to be in the proximity of that FMRFa gene. These data, utilizing the sequenced genomes currently available for each *Drosophila* species, suggest that the Tv enhancer has not been subjected to rearrangement or duplication throughout the evolution of *Drosophila*, spanning *D.melanogaster* to *D.grimshawi*. These data further establish that the Tv enhancer is upstream of the FMRFa gene in all species.

UCSC Browser (http://genome.ucsc.edu/cgi-bin/hgGateway), Galaxy (http://main.g2.bx.psu.edu/), and a stand-alone downloaded version of the ClustalX graphical

interface (Chenna et al., 2003; Larkin et al., 2007) were utilized for sequence alignments of the Tv-enhancer from the twelve *Drosophila* species. The *D.melanogaster* Tv-enhancer coordinates (Chr 2R: 5792874..5793318) was input into the UCSC Browser genome browser search. From the genome browser, we extracted sequences via the Table Browser using the following criteria: Region: Position Chr2R: 5792874-5793318 Group: Comparative Genomics; Track: Conservation; Table: Multiz15way, with the output set to MAF and sent directly to Galaxy. In Galaxy, the MAF format was converted into FASTA. The Tv-enhancer comprises two distinct MAF blocks. These were concatenated using the "FASTA manipulation>Concatenate FASTA alignment by species" command. The output generated was saved as a .txt file and uploaded into ClustalX to view aligned sequences.

Alignments shown in Fig.4.3.were 50bp fragments of the Tv enhancer downloaded from the Genome Browser using the PDF/PS Tab and downloaded as a .pdf file. These were then coloured in Preview to highlight specific sequence features and selected for upload into this document using Grab.

3. Results: Basic Characterization of the Tv enhancer

3.1. An integrase-based Tv^{wt}-nEYFP reporter to study the cis-regulation of FMRFa

In order to dissect the *cis*-regulation of *FMRFa* in the Tv neurons, I turned to a detailed analysis of the 446 bp Tv-enhancer (Benveniste et al., 1998). I utilized the φC31 integrase-based transgenesis system for my transgenic analysis (Bischof et al., 2007). This methodology targets injected transgenes (carrying the attB sequence) to a specific genomic location (containing the attP sequence), for stable germline transmission. Since all wildtype and mutant reporters are integrated into the same genomic locus, position effects are controlled for, and allow for reliable comparison between wildtype and mutant reporter activities. To utilize this system, I generated a transformation plasmid containing an attB sequence for site-specific integration into genomic attP sites, and a nuclear-localized EYFP (nEYFP) reporter placed downstream of the hsp70 minimal promoter (Fig. 3.1 A). All wildtype and mutant Tv-enhancer sequences (See Appendix C for sequences) were sub-cloned into this plasmid as EcoRI/XbaI fragments and placed upstream of the minimal promoter. The wildtype Tv reporter will be referred to as Tv^{wt}-nEYFP, while mutant Tv reporters will be referred to as Tv^{mutX}-nEYFP, with ^{mutX} denoting the nature of the mutation. I screened numerous well-characterized attP integration sites (see Methods and Materials) and found that attP2 (Groth et al., 2004) provided optimal reporter expression with minimal background, as previously reported (Markstein et al., 2008) (see Appendix A, Figure A.1). Accordingly, all subsequent experiments were conducted with nEYFP reporters inserted into the attP2 locus (Fig. 3.1 B). Contrary to reports that gypsy insulator elements have a positively acting effect on reporter expression levels (Markstein et al., 2008), an insulated Tv^{wt}-nEYFP construct (see Materials and Methods) did not show significant difference in activity from the non-insulated version, suggesting that the effects reported may be cell-specific or transgene-specific (data not shown). This issue was not pursued any further.

3.1.1. Expression of Tv^{wt}-nEYFP

The expression of Tv^{wt}-nEYFP in early larval stage 1 (L1) larvae was highly restricted to Tv neurons, as defined by co-labelling for anti-FMRFa, and was observed in 5.8±0.1 Tv neurons per VNC (n= 72 VNC's) (Fig. 3.2). Like the endogenous peptide, Tv^{wt}-nEYFP expression persists throughout the larval stages and is also detectable in adults (Eade and Allan, 2009) (see

Appendix A, Figure A.2). Taken together, these results indicate that *Tv-nEYFP* is a faithful reporter of *FMRFa* expression in the Tv neurons.
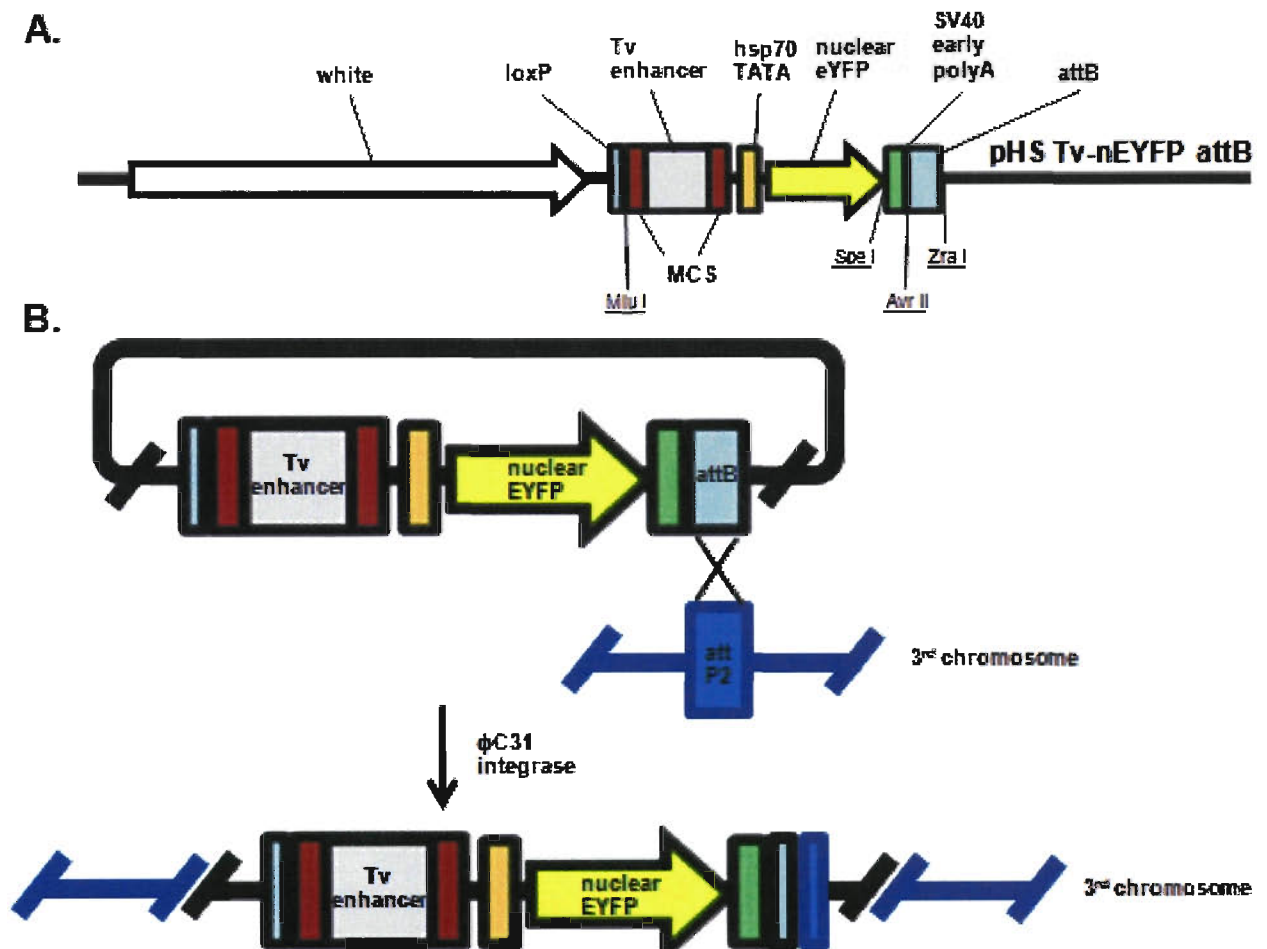
Figure 3.1 Schematic representation of pHS Tv-nEYFP attB and the integrase system.
A. Map of pHS Tv-nEYFP attB showing the main features in the following order, starting from 5' to 3'. Unique restriction sites introduced to this vector are underlined.
B. Site-specific integration of pHS Tv-nEYFP attB into genomic attP site (attP2). Slanted lines represent points where flanking sequences are omitted from the Figure.
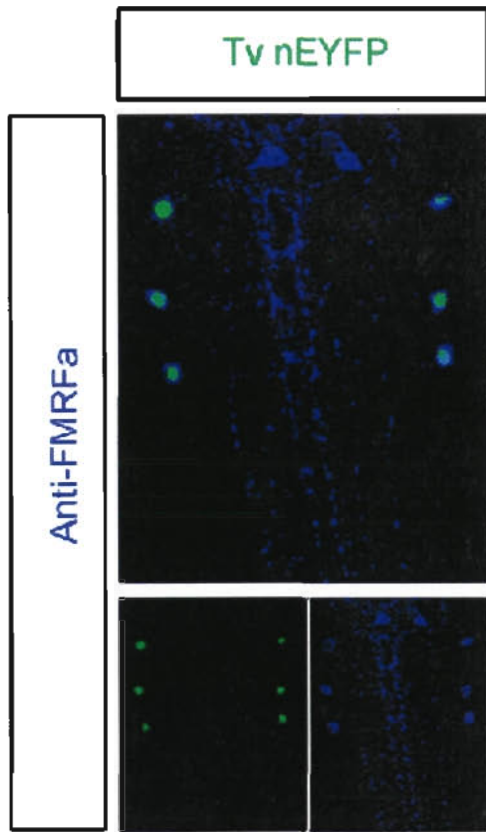
Figure 3.2. *Tv-nEYFP* is a faithful reporter of *FMRFa* expression.
Projected confocal stack through the Drosophila L1 ventral nerve cord (VNC) expressing nEYFP driven from the Tv-enhancer (*Tv-nEYFP;* green) and immunostained with anti-FMRFa (blue).
(Top panel) Overlaid image of Tv-nEYFP and anti-FMRFa shows perfect overlap of both signals in the Tv neurons, indicating that the *Tv-nEYFP* reporter is a faithful reporter of *FMRFa* expression.
(Bottom panels) Split fluorophore images, showing individual fluorescent image for *Tv-nEYFP* (left panel) and FMRFa peptide (right panel). Note that anti-FMRFa also detects a set of medial neurons termed the SE neurons. SE neurons are not BMP-dependent nor are they regulated by the Tv-specific combinatorial code.

3.2. The Tv enhancer is BMP-dependent and responsive to combinatorial transcriptional regulation.

In order to determine whether the Tv enhancer is a suitable tool for studying the *cis*-regulatory interplay between target-derived BMP signalling and Tv neuron-specific transcription factor codes, we tested the responsiveness of Tv^{wt}-*nEYFP* to regulators previously found to regulate endogenous *FMRFa* expression. In all cases, I utilized anti-Eyes absent or anti-FMRFa immunoreactivity to visualize the Tv-cluster in each thoracic hemisegment (Miguel-Aliaga et al., 2004). Data summary and statistical analysis of all experiments in this section can be found in Appendix B, Figure B.1.

3.2.1. BMP pathway

Wishful thinking is a BMP type II receptor shown to be essential for *FMRFa* expression (see Introduction). In *wishful thinking* null mutants (wit^{A12}/wit^{B11}), Tv^{wt}-*nEYFP* expression was eliminated, being observed in 0.0±0.0 Tv neurons per VNC (n= 10 VNC's, P=4.9 x 10^{-23} compared to +/+ control) (Figure 3.3 B and C). I obtained similar results with a *Tv-lacZ* reporter (Benveniste et al., 1998) (see Appendix A, Figure A.3). This indicates that, similar to the endogenous peptide, *Tv-nEYFP* expression is dependent on target-derived BMP signalling.

3.2.2. Apterous

In a strong *apterous* mutant background (ap^{GAL4}/ap^{P44}), Tv^{wt}-*nEYFP* was severely affected (Fig. 3.3 E), being observed in 0.7±0.3 Tv neurons per VNC (n= 10 VNC's, P=5.2 x 10^{-15} compared to +/+ controls). Interestingly, Tv^{wt}-*nEYFP* was expressed in 3.0±0.4 Tv neurons per VNC (n= 10 VNC's, P=1.9 x 10^{-7} compared to +/+ control) in a heterozygous *apterous* mutant background $(ap^{GAL4}/+)$, indicating that *apterous* is haploinsufficient for *Tv-nEYFP* expression (Fig. 3.3 D). Haploinsufficiency for *apterous* was not observed with endogenous *FMRFa* expression and a previously-studied *Tv-lacZ* reporter (Allan et al., 2003; Benveniste et al., 1998), suggesting that the heterozygous Tv^{wt}-*nEYFP* reporter provides a sensitive readout for screening regulators of *FMRFa* in the Tv neurons.

20

3.2.3. Dachshund

Dachshund (Dac) is a *FMRFa* regulator that can trigger ectopic expression of the peptide when mis-expressed with *apterous* in post-mitotic neurons of the Drosophila nervous system (Miguel-Aliaga et al., 2004). In null *dac* mutants (*Df(2L)Exel7086/dac³*), Tv^{wt}-*nEYFP* expression was slightly reduced, being observed in 4.8±0.2 Tv neurons per VNC (n= 14 VNC's, P=0.0012 compared to controls) (Fig. 3.3 G). This is reminiscent of the slight down-regulation of FMRFa peptide in *dac* mutants (Miguel-Aliaga et al., 2004). However, as was the case for *apterous* mutants, Tv^{wt}-*nEYFP* was more sensitive than the endogenous peptide was to *dac* mutation (Miguel-Aliaga et al., 2004). Furthermore, Tv^{wt}-*nEYFP*-expressing neurons displayed visibly weaker reporter activity in *dac* mutant background than in wildtype background (+/+ or *Df(2L)Exel7086/+*) (Fig. 3.3 F, G). Taken together, these results suggest that Dac is a moderate regulator of the Tv enhancer.

3.2.4. Zfh 1

Zfh1 was previously shown to regulate the expression of *FMRFa* and the *Tv-lacZ* reporter (Vogler and Urban, 2008). I verified that Tv enhancer expression is dependent on Zfh1, as Tv^{wt}-*nEYFP* is expressed in 0.7±0.6 Tv neurons per VNC (n=9 VNC's, P=5x10^{-11} compared to controls) in a *zfh1* hypomorphic mutant background (*zfh^{00865}/zfh^{00865}*) (Fig. 3.3 I).

3.2.5. Eyes absent

Next, I examined whether Eyes absent regulates the Tv enhancer. Here, I tested the *eya* mutant combination (*eya^{E1}/eya$^{Cli-IID}$*) and found that embryos failed to reach late 17/larval stages, the critical time window for analyzing *FMRFa* activation. Given the hypersensitivity of Tv^{wt}-*nEYFP*, I analyzed Tv^{wt}-*nEYFP* expression in a heterozygote *eya* mutant background (*eyaE1/+*) and found no significant difference compared to wildtype (data not shown). Further experimentation will be needed to determine the effect of this TF on Tv expression.

3.2.6. Apterous and Dachshund can induce ectopic expression of Tv^{wt}-*nEYFP*

Next, I tested whether *apterous* and/or *dachshund* is sufficient for Tv enhancer activation by mis-expressing combinations of *apterous* and/or *dachshund* in post-mitotic motor neurons (Miguel-Aliaga et al., 2004) using the *OK6-GAL4* driver (Aberle et al., 2002). I found that

whereas mis-expression of *apterous* alone failed to trigger any ectopic neurons (0.0±0.0 neurons per VNC (n= 2 VNC's), mis-expression of *apterous* and *dachshund,* or *dachshund* alone in motorneurons triggered significant ectopic expression of *Tv-nEYFP*, being observed in 74.3±3.8 neurons per VNC (n= 3 VNC's, P=0006 compared to controls) and 51.0±6.2 Tv neurons per VNC (n= 3 VNC's, P=0.008 compared to controls), respectively (Fig. 3.4 A-D). Mis-expression of *apterous* and *dachshund* is more potent than mis-expression of *dachshund* alone, as the combination triggered higher number of ectopic cells (P=0.03 between *OK6 GAL4 x UAS dac* and *OK6^{GAL4} x UAS ap, UAS dac*) and visibly stronger ectopic nEYFP signals (Fig. 3.4 C, D). We also detected similar ectopic *Tv-lacZ* expression from mis-expression of Apterous and Dachshund in all post-mitotic neurons, using the pan-neuronal GAL4 driver, *elav$^{GAL4-C155}$* (see Appendix A, Figure A.3.). Since mis-expression of *dachshund* alone could not ectopically activate an 8.0 kb *FMRFa* enhancer that includes the Tv enhancer (Miguel-Aliaga et al., 2004), my result could indicate that the Tv enhancer is hypersensitive to the positive regulatory effects of *dachshund.* Lastly, I also tested the ability of another *FMRFa* regulator, *sqz,* to activate ectopic *Tv-nEYFP* expression when misexpressed in combination with *apterous,* but failed to detect any ectopic effect (data not shown).

3.3. Summary

In summary, these results indicate that the Tv enhancer behaves in a similar manner to the endogenous FMRFa peptide in numerous key regards, and contains sequences required for both the gain-of-function and loss-of-function phenotypes conferred by manipulation of the BMP pathway and the *FMRFa* combinatorial code (Allan et al., 2005; Allan et al., 2003; Miguel-Aliaga et al., 2004). It should be noted that in all mutants for all genes tested above, except *eya,* Tv neurons survive to the point of analysis and they properly innervate the neurohemal organ (Allan et al., 2003; Benveniste et al., 1998; Miguel-Aliaga et al., 2004; Vogler and Urban, 2008), suggesting that the decrease in reporter activity is caused by a gene regulatory defect rather than a cell-wide problem.
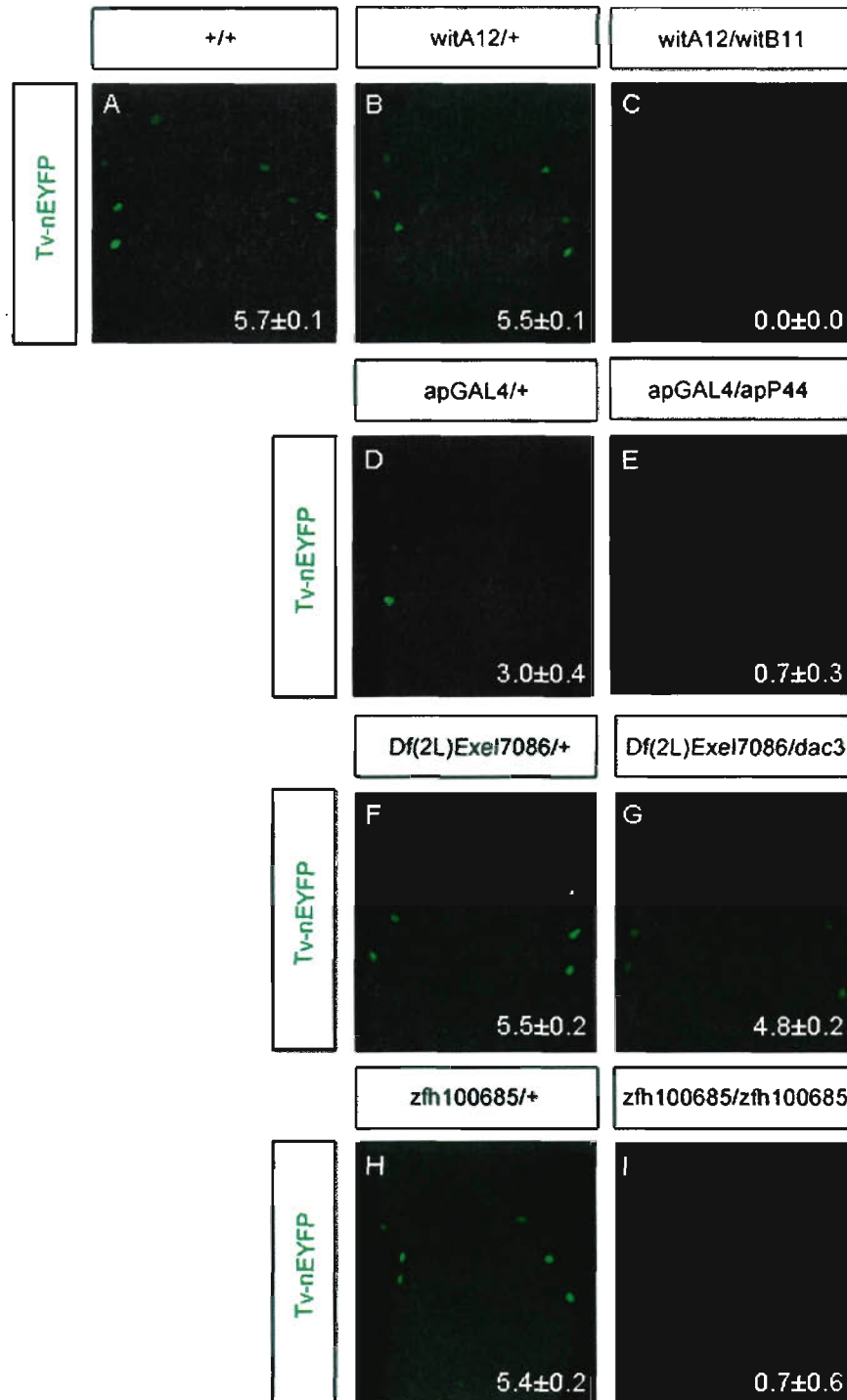
Figure 3.3. The Tv enhancer is regulated by BMP signalling and the *FMRFa* transcription factor code. Stacked projection of L1 larval VNC expressing one copy of $Tv^{wt}\text{-}nEYFP$ (green) in wildtype (A) and various mutant backgrounds (B-I). *Tv-nEYFP* expression is dependent on BMP signalling, as shown in *wit* mutants (compare B with C). Tv-nEYFP is also dependent upon *apterous* (compare A with D and E), *dachshund* (compare F with G), and *zfh1* (compare H with I). White font letters at bottom right corner of images represents number of nEYFP-positive neurons/Tv cluster ± S.E.M, as identified using either anti-Eya or anti-FMRFa.

23

w1118	UAS Ap	UAS Dac	UAS Ap, UAS Dac

OK6 GAL4; Tv-nEYFP

A

B

C — 51.0±6.2

D — 74.3±3.8
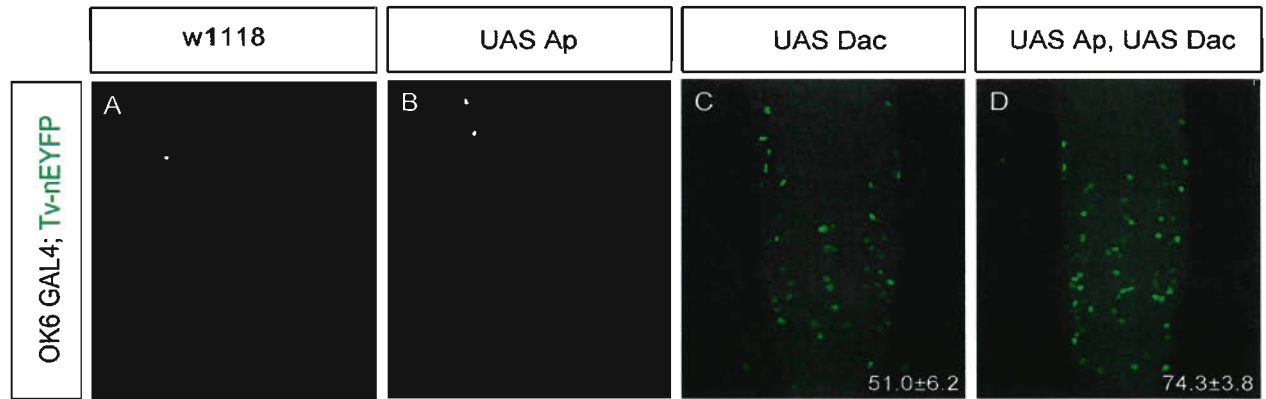
Figure 3.4. Misexpression of Apterous and Dachshund induces ectopic Tv-nEYFP expression. Stacked projection of L1 larval CNS displaying *Tv-nEYFP* (green) expression pattern after misexpression of no transgene (A), *UAS-ap* (B), *UAS-dac* (C), or *UAS-ap, UAS-dac* (D), driven by *OK6^{GAL4}*. White-font numbers at bottom right corner of images indicate mean number of ectopic nEYFP-positive neurons counted in the VNC. S.E.M is also indicated.

4. Results: Identification of a putative BMP-responsive element.

4.1. Conservation of sequences that are putative homeodomain-Smad integration sites

Since the Tv enhancer is BMP-dependent, it is possible that BMP signalling directly controls *FMRFa* transcription in the Tv neurons through Mad and/or Medea binding to this fragment (see Introduction). If this is true, then one would expect to find sequences corresponding to consensus Mad and/or Medea binding sites within the Tv enhancer. We further expected functional Mad/Medea binding sites to be evolutionarily conserved. Thus, we attempted to identify potential BMP-responsive elements through conservation analysis of the Tv enhancer sequence (Fig. 4.1). *Cis*-regulatory elements important for *FMRFa* expression in the Tv neurons are expected to be conserved throughout the ~50 million years of Drosophilid evolution; I confirmed FMRFa expression (using immunostaining for FMRFa) in the stereotypical Tv neurons (identified using anti-Eya immunostaining) in the CNS of the Drosophilid species *D. ananassae, D. willistoni* and *D. virilis* (see Appendix A, Figure A.4) (Fig. 4.2 A) (Clark et al., 2007; Stark et al., 2007).

We utilized many sequence alignment tools to analyze the Tv enhancer, but found two to be most useful for our analysis. These are the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway) (Kuhn et al., 2007) and Evoprinter (http://evoprinter.ninds.nih.gov/evoprintprogramHD/evphd.html) (Odenwald et al., 2005; Yavatkar et al., 2008). The UCSC genome browser provides a rapid display of best-match, pair-wise alignments of the query sequence (eg. *D.melanogaster* Tv enhancer) to all other *Drosophila* genomes, using blastz (Chiaromonte et al., 2002)). These are displayed as multiple sequence alignments of all twelve *Drosophila* species using multiz (Blanchette et al., 2004), and assessed for conservation (imaged as "mountains" that delineate degree of conservation) using phastCons (Siepel et al., 2005). Evoprinter performs three non-overlapping BLAT alignments (Kent, 2002) with a series of modifications termed enhanced BLAT (eBLAT) (Yavatkar et al., 2008), with the output being the query sequence (eg. *D.melanogaster* Tv enhancer) colour-coded to highlight which bases are conserved across all species selected in a search (Odenwald et al., 2005). One advantage of Evoprinter is the ability to 'test' the conservation of any combination of *Drosophilid* species, which assists in filtering out atypical non-alignment of sequences of

individual species, and to identify highly conserved, duplicated or rearranged sequences between species.

Here, I show the results of our analysis utilizing UCSC genome browser (Figs. 4.1, 4.3) and Evoprinter (Fig. 4.2) alignments, with the 446 bp Tv-enhancer as the input query for both. Notably, the three Apterous-binding HD motifs were found to be absolutely conserved throughout evolution, except for one base pair mismatch between *Drosophila melanogaster* and *Drosophila willistoni* at position 6 of the third HD motif.

Importantly, we found that each of the three HD motifs is juxtaposed to a conserved sequence that matches the approximate consensus GSCGNC of a Mad-binding motif. Only one of those sites, HD-A, also has an adjacent Medea-binding site of the consensus GTCT (Gao and Laughon, 2007; von Bubnoff et al., 2005). This Medea-binding site is also conserved throughout *Drosophila* evolution, except for a single base-pair mismatch in *D. ananassae*. Together with the HD sequence, we denote these three sites as HD/BRE-A, B and C. Importantly, HD/BRE-A was the most conserved module, in all its constituent HD, Mad and Medea-binding sequences, as well as in their relative spacing. HD/BRE-B and HD/BRE-C were less well conserved in their putative Mad-binding sequences (Fig. 4.3).

A

>*D.melanogaster* Tv enhancer full sequence
CCATCTGCAGACGTGGTTTTCGAACGTATTTATATTGATTATGGGTGATCGTCA
ACAAGAGCAGTGGACACCCAATAAACCTGTCCAAAAACCCGACACATTTCTGC
CCAGTCATGCGTGGTGGACAATAGCCAAATGCCATTGATGAGACTCGTCTCCA
AAACTTTGGCCTTTTGCCGGGCCGTAATTACAGACTTCCGTCTTTTGAACAGTT
TTTTCAGCCCCACCCAAGAGTCGAGTCTTGAAAAGCTGGCTGGGATGGGGTGG
TTTCGGGTGCTGGACGAGATGCCAGAGGCGCCACAATGTATCCTGTTACAGGT
TACAGGGCCATAAAGCGCCATAAACGCCGCGACGGCAATGGCAATTAATAACG
CATACGGACACGTAGTCGATCCACTGGCTAGAAGGCTAATTGGACGTGCCCGG
CCAGGATGTCCCTGCTCAT
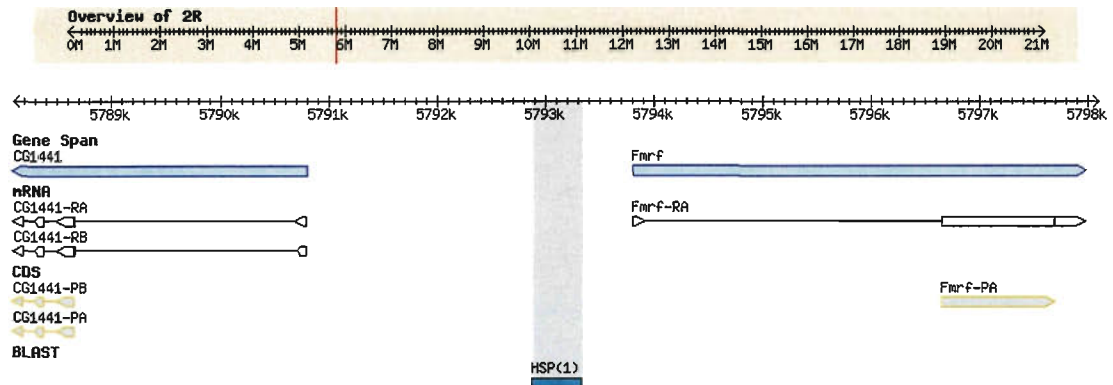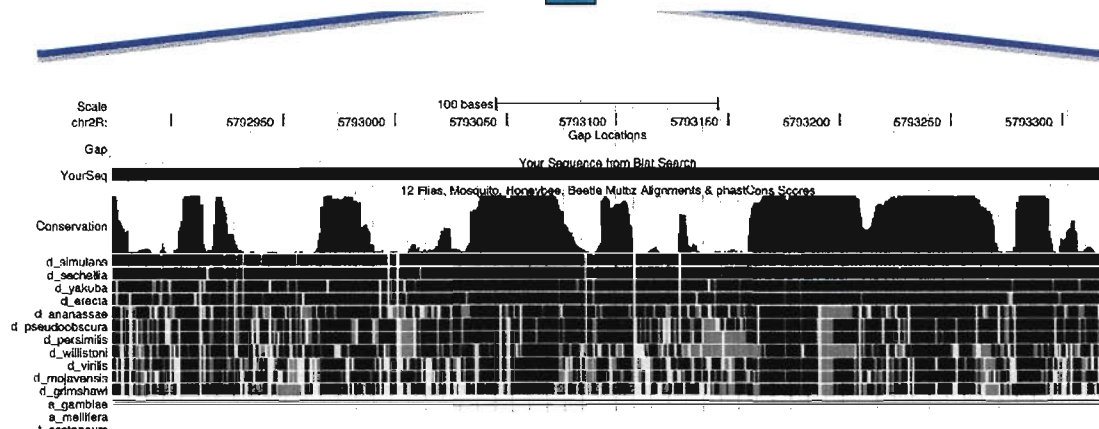
Figure 4.1: **Phylogenetic conservation of the Tv enhancer**: (**A**) Full sequence of Tv enhancer from *D.melanogaster* in FASTA format (**B**) Screenshot from Flybase Genome Browser (2R:5788096..5798095) showing BLAST hit for the Tv-enhancer (2R: 5792874..5793318; grey bar) of *D.melanogaster* Tv enhancer sequence, 5' of the FMRFa gene. (**C**) PDF output from UCSC Browser showing multiple species sequence conservation through the Tv enhancer (using phastCons).

A
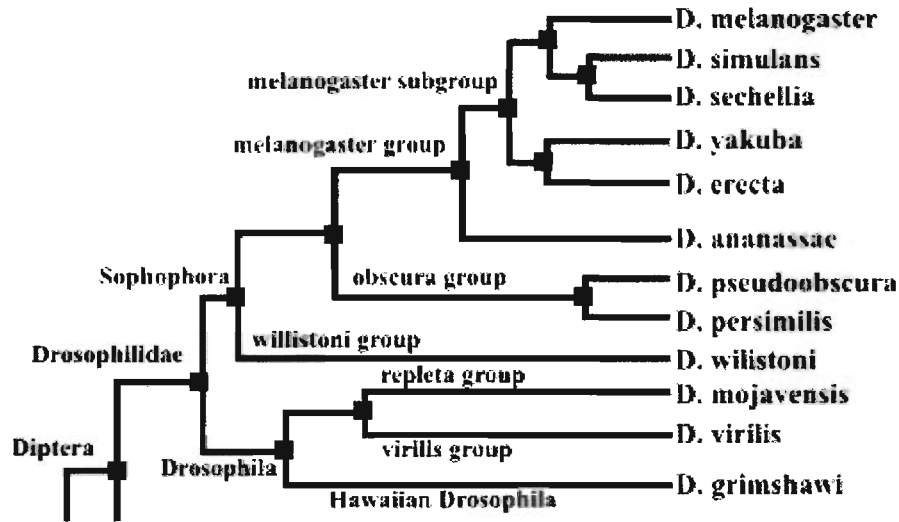
B

Tv enhancer sequence, showing EvoDiff profile.

ccatctgcagacgtggttttcgaacgtatttatattgattatgggtgatcgtcaacaagagcagtggacacccaataaacctgtccaaaaaccc

gacacatttctgcccagtcatgcgtggtggacaatagccaaatgccattgatgagactcgtctccaAAACTTTGGCCTTTtgcc

gGGCCGTAATTACAGACTTCCGtCtTttgaacagtttttcagccccacccaagagtcgagtcttgaaaagctggctg

ggatgcggtggtttcgggtgctgGAcGagaTGCCAgAGGCGCCACAAtGTATCCtgttacagGTTACAG

GGCCATAAAgCgCCATAAAcgccGCGACGgCAAtGgCAATTAATAaCGCATACGgACA

CGTAGtcgatccactggctagaaGGCTAATTGGACGTGCccGgCcAGGatgtccctgctcat

Figure 4.2: Sequence comparison of the Tv enhancer between 12 *Drosophila* species. **(A)** Phylogenetic tree showing evolutionary relationships of the sequenced Drosophilids. Screenshot from Flybase BLAST homepage (http://flybase.org/blast/) **(B)** Evoprint (in EvoDiff format) of the full length sequence of the *D.melanogaster* Tv-enhancer; Black capital letters are bases in the *D.melanogaster* reference sequence that are absolutely conserved in the *D.simulans, D.sechellia, D.erecta, D.yakuba, D.pseudoobscura, D.virilis* or *D.grimshawi* orthologous DNAs. Coloured bases are bases in the *D.melanogaster* sequence that are conserved in all species except the species coloured. The underlined sequences include HD/BRE A, B, and C in the order highlighted, and are shown in **Fig 4.3**.

A

Tv enhancer sequence, showing EvoDiff profile.

ccatctgcagacgtggttttcgaacgtatttatattgattatgggtgatcgtcaacaagagcagtggacacccaataaacctgtccaaaaacc
gacacatttctgcccagtcatgcgtggtggacaatagccaaatgccattgatgagactcgtctcca**AAAC**T**TTGGCCTTT**tgcc
g**GGCCGTAATTACAGACTT**CCG₁CtTttgaacagttttttcagccccacccaagagtcgagtcttgaaaagctggctg
ggatggggtggtttcgggtgctg**GA**cG**aga**TGCCA**g**AGGCGCCACAA**tGTATCC**tgttacag**GTTACAG**
GGCCATAAA**gC**gCCATAA**A**cgcc**GCGACG**g**CAA**tG**gCAATTAATA**a**CGCATACG**g**ACA**
CGTAGtcgatccactggctagaa**GGCTAATTGGACG**TGCccGgCcAGGatgtccctgct

B. double underline sequence (HD/BRE-A)
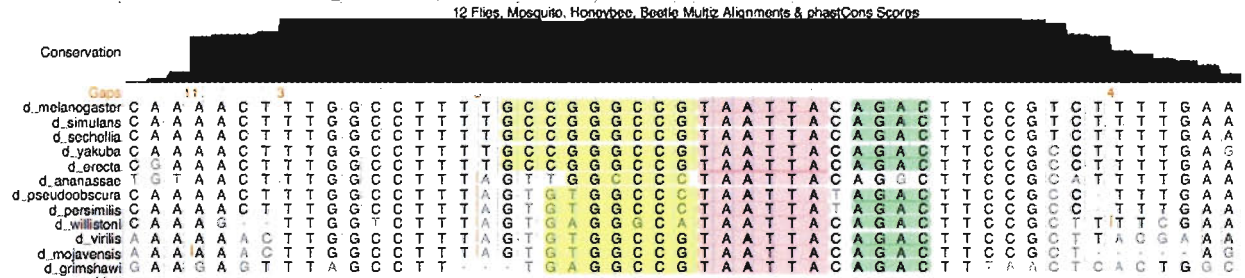
C. single underline sequence (HD/BRE-B)
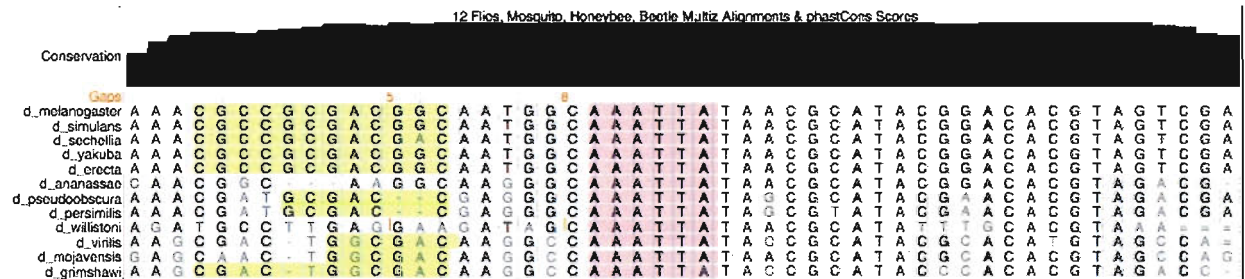
D. dotted underline sequence (HD/BRE-C)
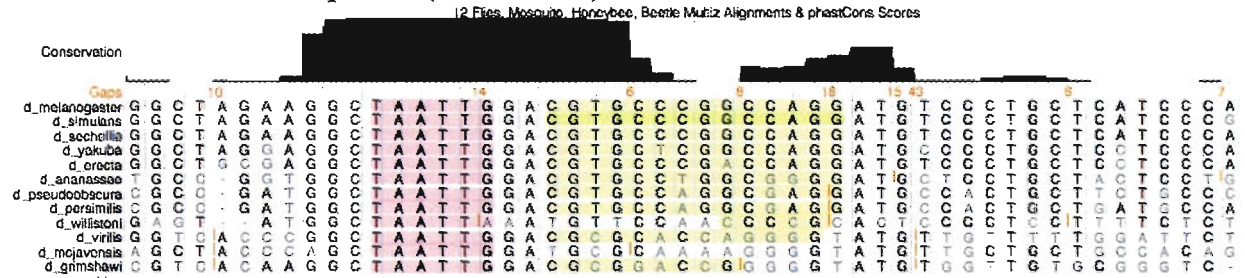
Figure 4.3. Detailed alignment of HD/BRE-A/B/C modules of the Tv enhancer. (A) Evoprint of the Tv-enhancer from Fig 4.2. **(B-D)** 12 species sequence alignment through underlined region from (A). Colour shading represents a consensus homeodomain-binding site (HD), a consensus Mad-binding site (Mad), and a consensus Medea-binding site (Med). Note the Gaps, denoted l (with the number of bases added for each species shown above {eg 10 or 14}). See Appendix H for full multiple alignment format file (MAF) of Tv enhancer to see these additional bases.

4.2. Functional analysis of predicted HD/BRE modules

The close association of predicted Mad/Medea-binding sites with HD-binding sites suggests a functional relationship between Apterous and BMP signalling. Although HD-C was previously shown to be important for Tv-enhancer expression (Benveniste et al., 1998), nothing is known regarding the functional role for each of the other HD sites, or the candidate Mad and Medea sites. To test the functional role of each candidate HD/BRE module, I compared reporter expression of wildtype versus mutant Tv enhancer sequences, using site-specifically integrated *Tv-nEYFP* reporters (as genetic heterozygotes). I examined *Tv-nEYFP* reporters carrying double mutations of adjacent Mad and HD binding sites from each HD/BRE module, and detected a decrease in reporter expression in all cases (Fig 4.4). Whereas Tv^{wt}-$nEYFP$ was expressed in 5.8±0.1 Tv neurons per VNC (n=72 VNCs), $Tv^{mMad-A/HD-A}$ was expressed in 0.1±0.1 Tv neurons per VNC (n=8 VNCs; P=4x10^{-8} compared to wildtype), indicating a near abolishment of reporter expression. $Tv^{mMad-B/HD-B}$ and $Tv^{mMad-C/HD-C}$ did not differ significantly from wildtype reporter in terms of cell count, expressing in 5.7±0.2 Tv neurons per VNC (n=6; P=0.6 compared to wildtype) and 5.5±0.2 Tv neurons per VNC (n=6; P=0.2 compared to wildtype), respectively. However, mutation of HD/BRE B or C did down-regulate reporter expression levels, with the respective reporters expressing at 26.8±2.7 and 49.8±7.6 % of wildtype fluorescent intensity (Fig. 4.4; Appendix B, Table B.3,4). These results indicate that all three HD/BRE regions contain sequences important for Tv expression, but that only HD/BRE-A is critical for Tv-enhancer activity.

4.3. Functional analysis of putative Mad binding sites

In order to test the contribution of individual Mad-binding sites to the expression of the Tv enhancer, I examined the effect of mutating individual candidate Mad-binding sites on *Tv-nEYFP* activity. Notably, only mutation of Mad-binding site A led to a reduction in reporter expression. Whereas Tv^{wt}-$nEYFP$ was expressed in 5.8±0.1 Tv neurons per VNC (n=72 VNCs), Tv^{mMad-A}-$nEYFP$ was expressed in 3.7±0.5 Tv neurons per VNC (n=7 VNCs; P=2x10^{-13}, compared to wildtype) (Fig. 4.4 C; Appendix B, Table 3, 4). Tv^{mMad-A}-$nEYFP$ also displayed a strong decrease in fluorescent intensity compared to wildtype (Fig. 4.4 B). In contrast, Tv^{mMad-B}-$nEYFP$ and Tv^{mMad-C}-$nEYFP$ displayed wildtype expression levels, being observed in 5.9±0.1 Tv

30

neurons per VNC (n=9 VNCs; No significant difference compared to wildtype) and 6.0±0.0 Tv neurons per VNC (n=12 VNCs; No significant difference compared to wildtype), respectively (Fig.4.4; Appendix B, Table 3). Furthermore, nEYFP intensity in these two mutants did not differ significantly from wildtype (Fig. 4.4 B and Appendix B, Table 4).

Next, I tested *Tv-nEYFP* reporters containing all possible combinations of mutant Mad-binding sites to determine whether Mad-B and C contribute combinatorially to the expression of the Tv-enhancer. Notably, combination of mutated Mad-A with mutated Mad-B and/or C did not result in any further reduction in reporter activity from that of mutant Mad-A alone, as $Tv^{mMad-A/B}$-*nEYFP*, $Tv^{mMad-A/C}$-*nEYFP* and $Tv^{mMad-A/B/C}$-*nEYFP* were expressed in 4.0±0.3 Tv neurons per VNC (n=7 VNCs; P=0.3 compared to Tv^{mMad-A}-*nEYFP*), 4.5±0.5 Tv neurons per VNC (n=4 VNCs; P=0.6 compared to Tv^{mMad-A}-*nEYFP*), and 4.0±0.0 Tv neurons per VNC (n=3 VNCs; P=0.7 compared to Tv^{mMad-A}-*nEYFP*), respectively (Fig. 4.4; Appendix B, Table 3). Moreover, double mutation of Mad-B and C failed to reduce Tv-nEYFP reporter expression relative to wildtype; $Tv^{mMad-B/C}$-*nEYFP* was expressed in 5.7±0.2 Tv neurons per VNC (n=6 VNCs; P=0.6 compared to wildtype). I also measured fluorescent intensity of *Tv-nEYFP* reporters for all mutant combinations above, and confirmed that Mad-B and C sequences play no significant role in Tv-enhancer activity (Fig. 4.4 B; Appendix B, Table 4).

4.4. Summary of bioinformatics and HD/BRE analysis

Taken together, these results indicate that only HD/BRE-A contains functional motifs predicted to bind the BMP second messenger/transcription factor Mad. Importantly, double mutation of Mad-A and HD-A caused a more severe down-regulation of reporter activity than that of Mad-A alone, suggesting that HD-A is also functional. Furthermore, since neither mutation of Mad-B nor C affected Tv expression, it is likely that mutation of HD-B and C contributed solely to the decrease in reporter activities of $Tv^{mMad-B/HD-B}$-*nEYFP* and $Tv^{mMad-C/HD-C}$-*nEYFP*, respectively. Thus, HD/BRE-A is the most likely candidate region for direct regulation by the BMP pathway. Given the juxtaposition of Mad, Med and HD binding motifs at HD/BRE-A, this module may also serve as the integration point between BMP signalling and a homeodomain transcription factor, which we postulate to be Apterous.
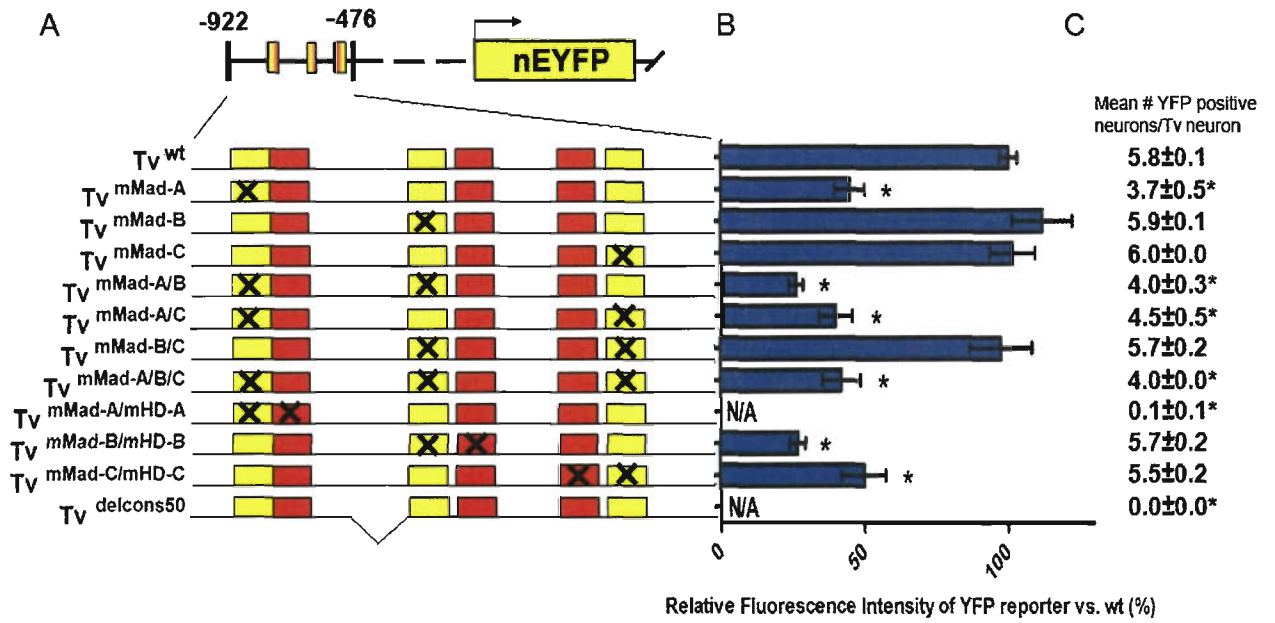
Figure 4.4. Expression analysis of heterozygote *Tv-nEYFP* reporters. (A). Schematic representation of *Tv-nEYFP* and numerous mutant versions of the reporter. Predicted Mad-binding sites (yellow) and known homeodomain-binding sites (red) are highlighted. Mutated sites are represented by X. Expression states of wildtype and variant reporter are measured as relative fluorescence intensity (B) and mean number of nEYFP-positive neurons/Tv neuron (C). **(B).** Relative fluorescence intensity is plotted directly next to the corresponding reporter shown in (A). Asterisks indicate a P value < 0.0001 when compared to *Tv-nEYFP*. $Tv^{mMad-A/mHD-A}$ and $Tv^{delcons50}$ are not shown for intensity, due to the low to zero number of nEYFP-positive Tv neurons, respectively. Error bars represent S.E.M. N/A indicates not available, as the specified samples have little to no YFP neurons available for quantification. **(C).** Mean number of nEYFP-positive neurons/Tv neuron is tabulated in the same order as the reporters shown in (A). Anti-FMRFa was used to visualize the Tv neurons in all cases. Asterisks represent P value <0.0001 when compared to *Tv-nEYFP*.

5. Results: Transgenic analysis of HD/BRE-A

5.1. Detailed analysis of HD/BRE-A.

Although both HD and candidate Mad-binding motifs of HD/BRE-A were shown to be essential for Tv-enhancer activity (Chapter 4), it was unclear whether the consensus Medea binding sequence, hereafter referred to as Med-A, is also functional. Furthermore, it was unclear whether each predicted motif in the module acts independently or acts synergistically with one another. In order to address these issues, I performed a more detailed analysis of HD/BRE-A. I used homozygous *Tv-nEYFP* reporters in all subsequent experiments because they display stronger fluorescent signals and improves one's ability to distinguish the intensity profile of weak-expressing Tv-enhancer mutants.

5.1.1. Mad-A and HD-A

To begin my analysis, I re-examined the effects of mutations of Mad-A and HD-A on homozygous reporter activities. First, I verified that Mad-A is required for Tv expression. Although homozygous Tv^{mMad-A}-$nEYFP$ was expressed in the normal number of Tv neurons, at 5.8±0.3 Tv neurons per VNC (n=4 VNCs, P=0.10 compared to wildtype), its expression levels was strongly reduced, at 30.7±3.8 % of wildtype nEYFP intensity (n=23 neurons, P=2.7 x 10^{-9} compared to wildtype) (Fig. 5.1; Appendix B, Table 5, 6). I also tested whether a more subtle mutation of Mad-A would disrupt reporter expression. In $Tv^{mMad-A-4bp}$, I mutated 4 bases out of the 9 bp sequence, as opposed to mutation of 6 bases in Tv^{mMad-A} (Fig. 5.1; Appendix C, Table C.1). I found that $Tv^{mMad-A-4bp}$-$nEYFP$ did not significantly differ from Tv^{mMad-A} in either cell count or expression level; $Tv^{mMad-A-4bp}$-$nEYFP$ was observed in 5.7±0.3 Tv neurons per VNC (n=3, P=0.046 compared to wildtype, P=0.84 compared to Tv^{mMad-A}-$nEYFP)$ and expressed at 20.5±3.5 % of wildtype nEYFP intensity (n=47 neurons, P=1.2x10^{-21} compared to wildtype, P=0.09 compared to Tv^{mMad-A}) (Appendix B, Table 5, 6).

Second, I found that HD-A alone is critical for Tv activation. Tv^{mHD-A}-$nEYFP$ was expressed in 0.4±0.2 Tv neurons per VNC (n=7 VNCs, P=1.4 x10^{-14} compared to wildtype) and at 9.7±7.0 % of wildtype nEYFP intensity (n=5 neurons, P=0.00027) (Fig 5.1). To determine

whether Mad-A and HD-A are independently-acting motifs, I examined reporters bearing a double mutation of Mad-A and HD-A site. Interestingly, I found that $Tv^{mMad-A/HD-A}$-nEYFP did not differ significantly from reporters bearing mutation of HD-A alone; $Tv^{mMad-A/HD-A}$-nEYFP was observed in 1.0±0.0 Tv neurons per VNC (n=2, P=2.5 x 10^{-33} compared to wildtype, P=0.19 compared to Tv^{mHD-A}-nEYFP) and expressed at 5.3±1.7 % of wildtype nEYFP intensity (n=3 neurons, P=0.0023 compared to wildtype, P=0.65 compared to Tv^{mHD-A}-nEYFP). While the sample size is small, these results suggest that HD-A is epistatic to Mad-A, and is contrary to the model that the two motifs act independently of each other. However, one issue with this interpretation is that the expression levels are so low in the Tv^{mHD-A}-nEYFP that it may be difficult to precisely ascertain any differences between the single HD-A mutant and double HD-A/Mad-A mutant. This may be resolved by adding increasing numbers of Tv-nEYFP reporters, and/or using anti-EYFP immunoreactivity (anti-GFP) to enhance the reporter signal. Regardless, co-mutation of Mad-A did not enhance the loss of expression of an HD-A mutation.

5.1.2. Med-A is a functional motif

An important feature that distinguishes HD/BRE-A from HD/BRE-B and C is the presence of an evolutionarily conserved, putative Medea binding site, or Med-A, which lies directly adjacent to HD-A (Figs. 4.2, 4.3). To test the functionality of this motif, I swapped two base pairs within the motif (5'-AGAC-3' to 5'-GAAC-3') that is predicted to make direct contact with the MH1 domain of Smad4/Medea (Gao and Laughon, 2007). Tv^{mMed-A}-nEYFP displayed a slight reduction in cell count, at 5.3±0.3 Tv neurons per VNC (n=4, P=2.8 x 10^{-6}) but was expressed at strongly reduced levels compared to wildtype, at 12.5±1.4 % of wildtype NEYFP intensity (n=60 neurons, P= 1.5x10^{-12} compared to wildtype) (Fig. 5.1). This indicates that Med-A is required for wildtype Tv expression. Note that this is a similar effect as mutation of Mad-A ($Tv^{mMad-A-4bp}$-nEYFP was observed in 5.7±0.3 Tv neurons per VNC and expressed at 20.5±3.5 % of wildtype nEYFP intensity). Overall, the above results showed that each predicted binding motif in HD/BRE-A is essential for Tv activation.

Besides Med-A, three other Medea consensus sequences were positioned within 50 bp of HD/BRE-A; two are located upstream of HD/BRE-A, while the other is located 4 bp downstream from Med-A.

tgagactcgtctccaAAACTTTGGCCTTTtgccgGGCCGTAATTACAGACTTCCGtCtTttga

None of these additional motifs were found to be conserved throughout *Drosophila* evolution (Fig 4.2). To determine whether these sequences contribute to the residual activity left in a Med-A mutant reporter, I generated multiple combinations of mutations in these predicted Medea binding site. One of these Tv mutants contains a mutation in Med-A along with mutations in the two closest predicted Medea sites (*Tv^{m3Med}-nEYFP*). Interestingly, while I found no significant difference in cell count, I found that reporter activity was significantly higher in *Tv^{m3Med}-nEYFP* than in *Tv^{mMed-A}-nEYFP*; *Tv^{m3Med}-nEYFP* was observed in 5.3±0.3 Tv neurons per VNC (n=6, P=0.86 compared to *Tv^{mMed-A}-nEYFP)* and expressed at 25.6±3.2 % of wildtype nEYFP intensity (n=35, P=5.2x10^{-9} compared to wildtype, P=9.0x10^{-5}compared to *Tv^{mMed-A}-nEYFP*) (Fig. 5.1). While this might indicate that one of the surrounding Medea binding sites is a repressive element, additional *Tv^{m3Med}-nEYFP* lines will have to be tested to rule out any variation in transgenic attP2 lines expression. Nevertheless, this preliminary result suggest that the surrounding Medea sites do not compensate for mutation in Med-A, and are not likely to be positively-acting elements in the Tv-enhancer.
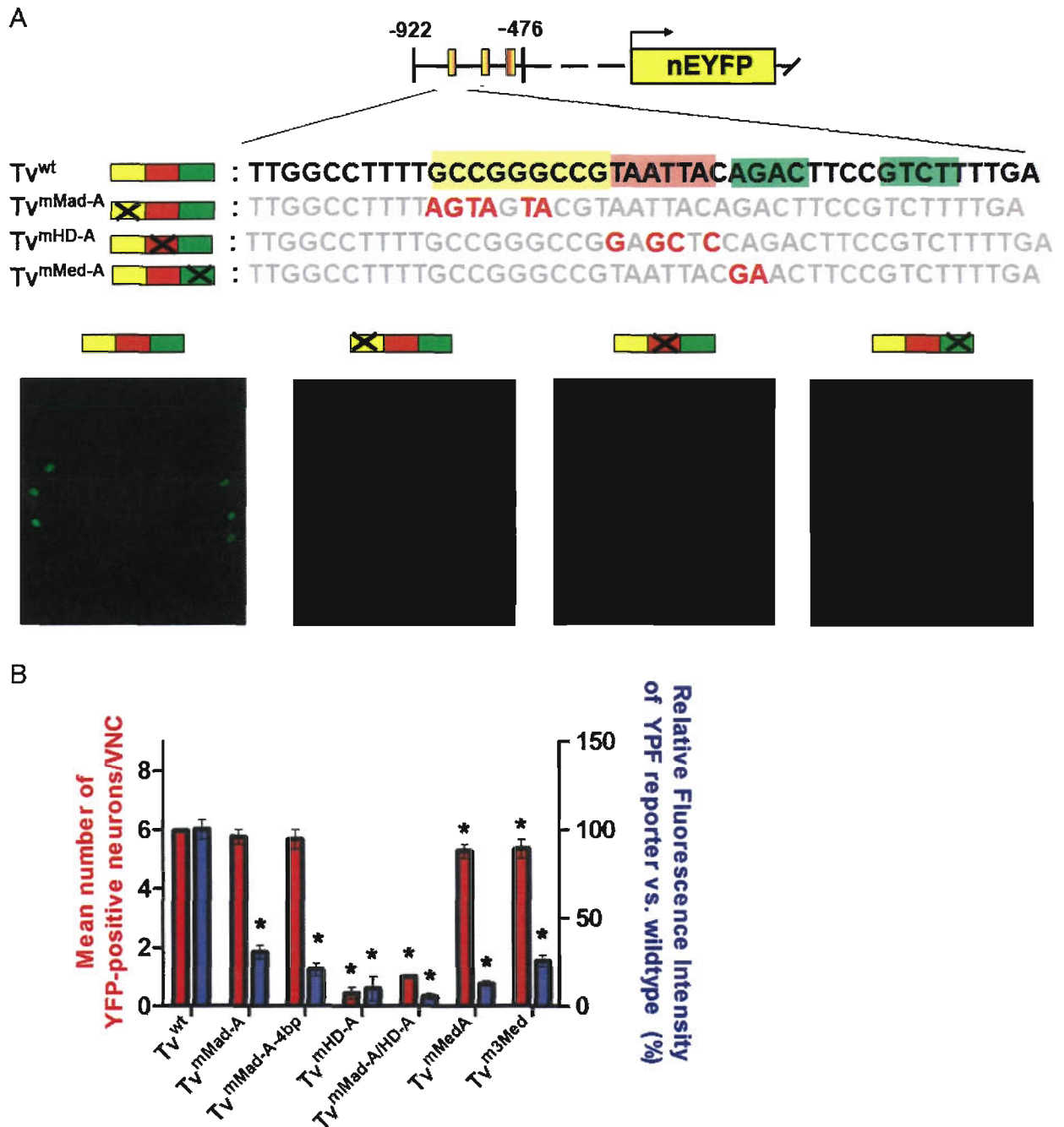
Figure 5.1. Every element in the identified HD/BRE-A is important for *Tv-nEYFP* expression. A. Schematic representation of *Tv-nEYFP* and HD/BRE-A region. Homozygous reporters were used for this experiment. HD/BRE-A is comprised of a predicted Mad (yellow), Apterous (red) and Medea (green) binding sites, and is represented by a simplified, color-coded rectangle box. In order to study the importance of these sites, mutagenesis was carried out and sequences in red font represent the mutant base pairs. X on one of the rectangle boxes indicates that the particular binding site has been mutated according to the displayed mutant sequences shown. Stacked projections of L1 larval CNS expressing *Tv-nEYFP* reporters with mutations in all three predicted binding sites are shown. **B.** Expression states of wildtype and mutant *Tv-nEYFP* reporters is represented as mean number of nEYFP-positive Tv neurons/VNC (red bars) as well as relative fluorescence intensity vs. wt (blue bars). Asterisks represent P value < 0.005 compared to wildtype. Error bars represent S.E.M.

36

5.1.3. Relative spacing between HD/BRE-A binding motifs is critical for wildtype Tv activation

Next, we tested for functional spatial requirements for the HD, Mad and Medea motifs in HD/BRE-A, commonly used to indicate molecular interactions between *trans*-acting factors (Fig. 10). Previous studies showed that there is a requirement for proper spacing between Mad and Medea motifs (Gao and Laughon, 2007) in BMP-responsiveness. In the presence of an additional transcription factor binding site, such as a homeodomain factor, certain studies have shown a requirement for native spacing between HD and BMP-binding sequences (Walsh and Carroll, 2007). In contrast, other studies have shown that a juxtaposed Medea site is not critical for BMP-responsiveness in the presence of a second juxtaposed motif for an additional transcription factor, including homeodomain transcription factors (Brugger et al., 2004).

In order to discriminate a spatial requirement for the HD motif in relation to the Mad and Medea motifs in HD/BRE-A, we created two mutant *Tv-nEYFP* constructs in which a five base-pair spacer was inserted between the Mad and HD motifs ($Tv^{ins5Mad-A/HD-A}$-*nEYFP*) and between the HD and Medea motifs ($Tv^{ins5HD-A/Med-A}$-*nEYFP*) (Fig. 5.2 A). A 5 bp spacer will presumably place maximal steric strain on molecular interactions by imposing a half-helical turn between DNA-binding motifs. In these two constructs, the spacing between Mad-A and Med-A sites is equivalent, but the relationship of the HD-A site to either the Mad-A or Med-A motifs was altered. Intriguingly, I observed a dramatic down-regulation of reporter expression when a spacer was inserted between HD-A and Med-A. $Tv^{ins5HD-A/Med-A}$-*nEYFP* was observed in fewer cells than wildtype, at 4.7±0.3 Tv neurons per VNC (n=3 VNCs; P=6.8 x10^{-12} compared to wildtype), and was expressed at a dramatically lower level, at 5.6±0.8 % of wildtype nEYFP intensity (n=24, P=1.1x10^{-13}) (Fig. 5.2 B). In contrast, $Tv^{ins5Mad-A/HD-A}$-*nEYFP* showed no significant reduction in cell count, at 5.7±0.3 Tv neurons per VNC (n=3 VNCs; P=0.046 compared to wildtype) but was expressed at a significantly lower level than wildtype, at 77.6±4.9 % of wildtype nEYFP intensity (n=53 neurons; P=0.0028 compared to wildtype) (Fig. 5.2 B). These data indicate that no critical spatial requirements exist between the Mad and Medea, or Mad and HD motifs - in the context of a 5bp spacer within the HD/BRE-A, but that there is a critical spacing requirement between the HD and Medea motifs. I am in the process of generating Tv enhancer mutants bearing different 5bp spacer insertions to rule out any sequence-specific effects in $Tv^{ins5HD-A/Med-A}$-*nEYFP* and $Tv^{ins5Mad-A/HD-A}$-*nEYFP*.

5.2. Summary

Overall, the above results indicate that each predicted DNA-binding element in the HD/BRE-A module is essential for activation of the Tv enhancer. Furthermore, the relative orientation of the predicted binding sequences is critical for proper Tv activation, suggesting a strict spatial requirement for proper binding between Apterous, Mad and Medea.
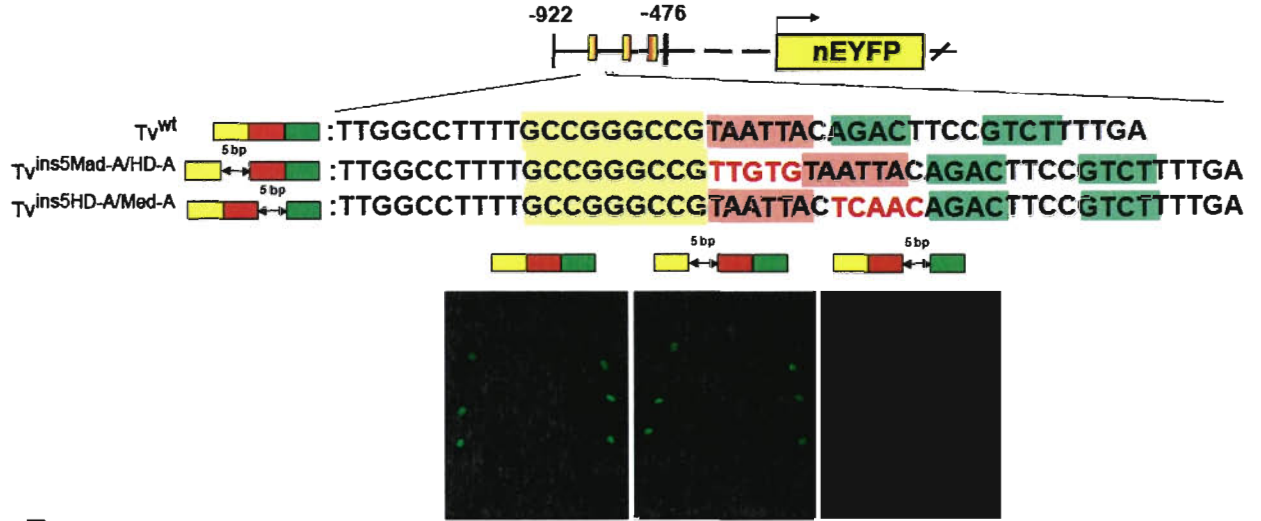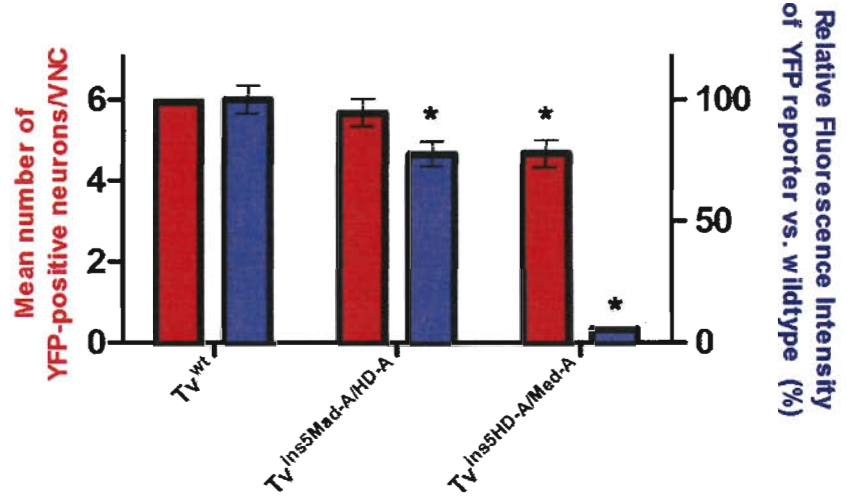
Figure 5.2. Spacing between predicted Mad, Apterous, and Medea binding sites is critical for normal Tv expression. A. Schematic representation of *Tv-nEYFP* and HD/BRE-A region, including sequence of spacer mutants. Homozygous reporters were used for this experiment. HD/BRE-A is comprised of a predicted Mad (yellow), Apterous (red) and Medea (green) binding sites, and is represented by a simplified, color-coded rectangle box. In order to study the spatial relationship between these sites, a 5 bp spacer (horizontal, double arrowhead) was inserted on either site of the Apterous binding site. Stacked projections of L1 larval CNS expressing *Tv-nEYFP* reporters bearing the insertion mutations are shown. **B.** Expression states of wildtype and mutant *Tv-nEYFP* reporters is represented as mean number of nEYFP-positive Tv neurons/VNC (red bars) as well as relative fluorescence intensity vs. wt (blue bars). Asterisks represent P value < 0.005 compared to wildtype. Error bars represent S.E.M.

39

6. Results: Biochemical characterization of HD/BRE modules

6.1. Apterous, Mad and Medea bind appropriate HD/BRE sequences in *vitro*.

6.1.1. Overview

Thus far, transgenic analysis suggest that HD/BRE-A is the most likely site of integration between BMP signalling and Apterous in the regulation of the Tv enhancer. This model would be strengthened if the proposed transcription factors can physically associate with HD/BRE-A at the appropriate motifs. To test whether Apterous, Mad and Medea can bind HD/BRE-A at putative HD, Mad and Medea motifs, respectively, I performed DNA-protein interaction studies using electrophoretic mobility shift assays (EMSA).

6.1.2. Apterous

First, I examined Apterous binding to the Tv enhancer, using a N-terminal, GST fusion to the homeodomain of Apterous (GST-LIMless Ap) (for SDS-PAGE analysis of purified proteins see Appendix E, Figure E.1.). In agreement with a previous report (Benveniste et al., 1998), I found that Apterous can bind specifically to all three HD motifs in the Tv enhancer (HD-A/B/C; Fig. 6.1; see Appendix E, Figure E.2). Strikingly, whereas HD/BRE-A oligonucleotides carrying wildtype HD-A motif or mutated Mad-A motif were able to compete for Apterous binding to radiolabelled, wildtype HD/BRE-A, oligonucleotides bearing mutation in HD-A motif failed to show any competition even at 1000 fold excess level (Fig. 6.2 A). This result shows that Apterous binding is highly specific to HD-A.

6.1.3. Mad

Next, I tested whether Mad can bind to the Tv enhancer, using a previously characterized N-terminal fusion of GST to the MH1 domain of Mad (GST-MadN) (Kim et al., 1998; Walsh and Carroll, 2007). I found that GST-MadN can associate specifically with all three wildtype HD/BRE sequences (Mad-A/B/C) in *vitro* (Fig. 6.1; see Appendix E, Figure E.2.). Since only Mad-A is essential for Tv expression in *vivo,* I will focus my analysis here on this motif. Mad and Medea both possess evolutionarily divergent MH1 DNA binding domains, which have

evolved to bind different consensus sequences (Walsh and Carroll, 2007). However, truncated Mad MH1 domain was previously found to display residual binding activity for consensus Medea sequences (Walsh and Carroll, 2007). Since HD/BRE-A also contains a Medea consensus sequence, I examined the ability of GST-MadN to bind to oligonucleotides carrying only Mad-A and HD-A motifs (HD/BRE-A-Madonly). I found that GST-MadN caused a single band shift when incubated with HD/BRE-A-Madonly oligonucleotides, suggesting that MadN binds to this sequence as a single species (Fig. 6.1 B). GST-MadN binding affinity for HD/BRE-A-Madonly was reduced by mutating the Mad-A motif (Fig. 6.1 B). Furthermore, wildtype HD/BRE-A-Madonly sequence competed more efficiently for MadN binding than mutant HD/BRE-A-Madonly sequences where Mad-A is mutated (Fig. 6.2 B). Together, these results indicates that MadN can associate specifically with Mad-A.

6.1.4. Medea

Finally, I tested whether Med can bind to HD/BRE-A, using an N-terminal fusion of MBP to the MH1 domain of Med (MBP-MedN). Besides Med-A, an additional consensus Medea binding site is present on HD/BRE-A oligonucleotides. Thus, both sites were always mutated together in the following EMSA experiments. Upon incubation with oligonucleotides bearing the HD/BRE-A sequence, MBP-MedN caused two very weak band shifts (Fig. 6.1 B). The top band shift can be eliminated by mutating both consensus Medea binding sites in the HD/BRE-A oligonucleotide (Fig. 6.2 B) suggesting specific binding by MedN. The bottom band shift may represent non-specific binding by incompletely translated, or degraded MBP-MedN. In accordance with this idea, non-specific binding was also observed when MBP-tag alone was incubated with the HD/BRE-A oligonucleotide (Fig. 6.2 B). As noted above, binding by MBP-MedN to HD/BRE-A is weak and may indicate that majority of the purified proteins are inactive. In support of this, significant aggregation of radiolabelled HD/BRE-A oligonucleotides was observed in the wells, suggesting formation of insoluble protein-DNA complexes. To obtain better results, purification of MBP-MedN may have to be optimized to enhance solubility. Alternatively, GST-MedN has been shown to provide robust binding to consensus Medea sequences (Walsh and Carroll, 2007), and may be the more suitable construct to use for future experiments.

6.1.5. Summary

To summarize, I found that the DNA-binding domains of Apterous, Mad and Medea can specifically associate with the HD/BRE-A sequence *in vitro*. Coincidentally, the same mutations that reduced reporter activity in *vivo* also disrupted DNA binding of each of the factors. Taken together, these results indicate that Apterous, Mad and Medea can all associate specifically with each of their cognate sequence motifs in the HD/BRE module and strongly support the idea that these three proteins also bind to HD/BRE-A in *vivo* to regulate the expression of *FMRFa*.
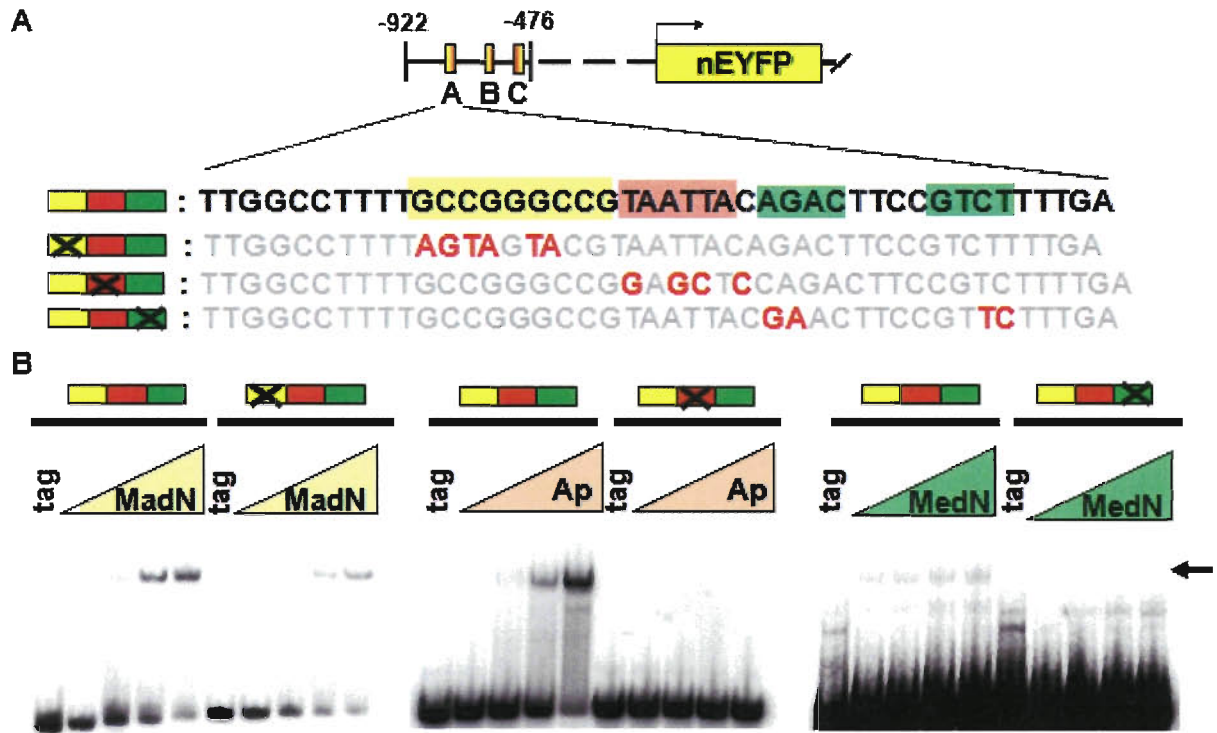
Figure 6.1. Apterous, Mad and Medea can all associate specifically with HD/BRE-A. **A.** Schematic representation of *Tv-nEFYP* and a zoomed-in sequence showing the HD/BRE-A region. Predicted Mad, Apterous and Medea binding sites are highlighted in yellow, red and green, respectively. Simplified representation of HD/BRE-A is displayed as colour-coded boxes indicating the corresponding binding sites. X in box indicates that the corresponding binding site has been mutated. **B.** GST-MadN (MadN), HA-LIMless Ap (Ap), and GST-MedN (MedN) can all bind to radiolabelled oligonucleotides bearing the wildtype HD/BRE-A sequence shown in (A). Binding activity is either lost or reduced to oligonucleotides bearing mutation of predicted Mad, Apterous or Medea bindings sites as indicated in (A). Tag control for MadN and Ap binding experiments are purified GST. Tag control for MedN binding experiment is purified MBP. Arrow indicates band shift which is eliminated upon mutation of consensus Medea sites.
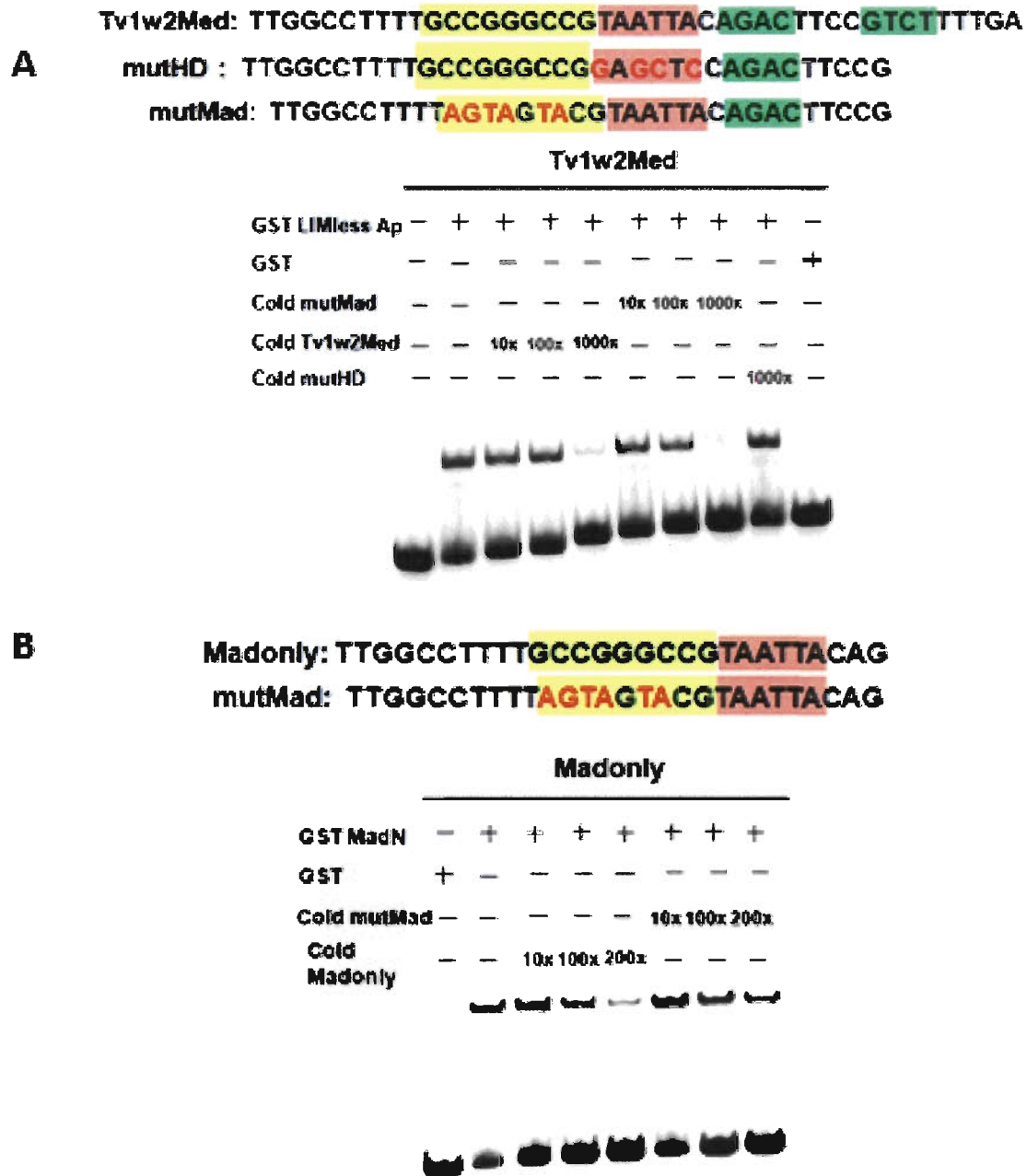
Figure 6.2. Competition assay showing sequence specificity in Apterous and Mad binding.
Oligonucleotide sequences representing predicted Mad, Apterous and Medea binding sites are highlighted in yellow, red and green colours, respectively. Red font letters indicate mutated sequences. **A.** EMSA experiment showing band shifts caused by binding of GST-LIMless Ap to radiolabelled Tv1w2Med oligonucleotides. Band shift signals are sharply reduced from competition with cold, Tv1w2Med or mutMad oligonucleotides at 1000 fold excess of radiolabelled probes. However, cold oligonucleotides bearing mutation in the predicted homeodomain binding site (mutHD) fail to compete away the signal even at 1000 fold excess of radiolabelled probes. **B.** EMSA experiment showing band shifts caused by binding of GST-MadN to radiolabelled Madonly oligonucleotides. Band shift signals are more efficiently competed away by addition of cold mutMad oligonucleotides over addition of cold, wildtype Madonly oligonucleotides.

44

6.2. Apterous fails to co-immunoprecipitate with Mad or Medea in vitro

The ability of Ap/Mad/Medea to associate specifically with their corresponding binding sites in *vitro*, combined with transgenic data indicating a requirement for proper spacing between their corresponding sites in *vivo,* raises the possibility that Ap/Mad/Med activate *FMRFa* expression through the formation of a physical complex. To test this idea, I examined whether an N-terminally tagged HA-Apterous can co-immunoprecipitate FLAG-Mad or myc-Medea from S2 cell lysate in a constitutively active BMP signalling background, induced by constitutively-activated BMP-type I receptor, Thickveins (TkvQD-FLAG). Using anti-HA for immunoprecipitation of HA-Ap, I detected expression of all transfected products in the input and flow-through fractions (Fig. 6.3). However, while I detected HA-Ap in the IP fraction, I could not detect FLAG-Mad or myc-Medea in the same fractions (Fig. 6.3). The same result was obtained following co-immunoprecipitation experiments carried out in low (75 mM NaCl) and high (300 mM NaCl) salt conditions (data not shown). Together, these results fail to provide support for the idea that Apterous can associate with Mad and/or Medea. However, proving a negative result can be difficult, and further tests must be performed to optimise these tests. A positive control to perform would be to show that tagged Mad can co-immunoprecipitate tagged Medea, and vice versa. Moreover, tagged Mad or Medea will have to be tested for pull-down of tagged Apterous.

6.3. Summary

In summary, I found that Apterous, Mad and Medea can all interact specifically with their predicted binding motifs in HD/BRE-A in *vitro.* However, it is still unclear whether these factors can associate with HD/BRE-A in *vivo.* Furthermore, no evidence has been found to date for physical association between Apterous and Mad or Medea.
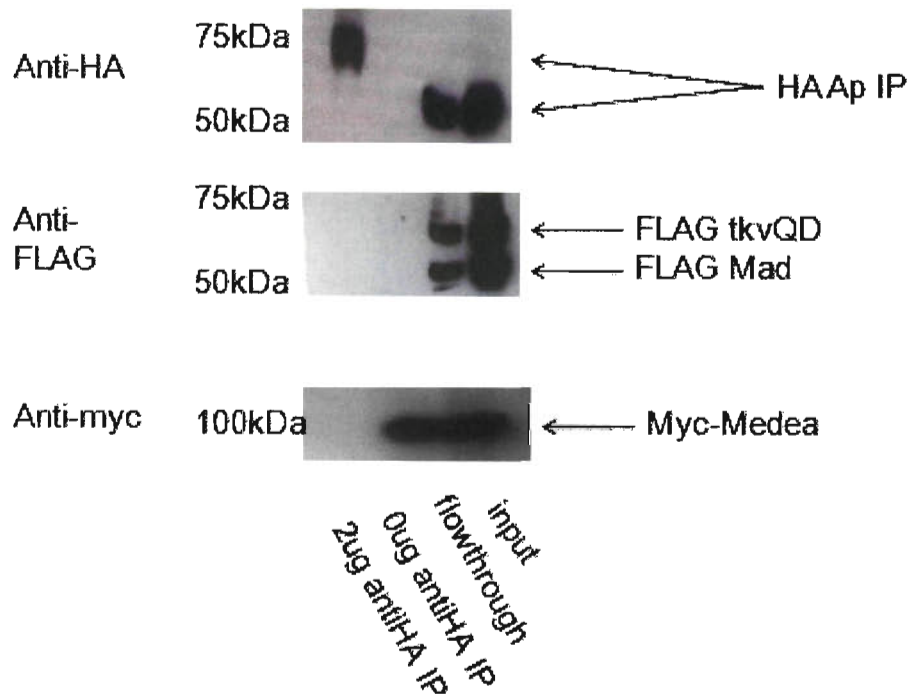
Figure 6.3. Apterous failed to co-immunoprecipitate Mad or Medea in BMP-active, S2 cells. Images showing western blot results from immunoprecipitation of HA-Apterous using anti-HA. S2 cells transfected with FLAG-TkvQD, HA-Apterous, FLAG-Mad, and myc-Medea and co-immunoprecipitation was carried out . While all transfected products were found in input and flow-through, only HA-Apterous was detected in the IP fraction. Migration of HA-Apterous was distorted on the gel, and the 2ug anti-HA IP sample was found near the 75kDa ladder band.

7. Discussion

7.1. *Cis*-regulatory integration of BMP signalling and Apterous at the Tv enhancer.

The above results support the model that target-derived BMP signalling and the homeodomain transcription factor Apterous directly activates *FMRFa* expression in the Tv neurons. The two factors integrate at the Tv enhancer through direct binding of both Mad/Medea and Apterous to the HD/BRE-A module. Several lines of evidence support this idea. First, expression of the Tv enhancer is dependent on Apterous and BMP signalling. Second, the Tv-enhancer includes a module that consists of highly conserved transcription factor binding sites for Mad, Medea and Apterous, collectively called HD/BRE-A. Third, all three binding sites are necessary for normal Tv expression. Fourth, Mad, Apterous and Medea can all specifically associate with their predicted binding sites in the HD/BRE-A module in *vitro*. Fifth, proper topology between Mad, Apterous and Medea binding sites is required for wildtype Tv expression, suggesting a spatial requirement for binding between these transcription factors. Altogether, this study demonstrates the first mechanistic link between target-derived signalling and homeodomain transcription factors in the activation of TDGs in post-mitotic neurons.

Is the *cis*-regulatory integration of BMP signalling and homeodomain transcription factors a common mechanism for specifying neuronal identity? There are reasons to believe this is the case. First, BMP/homeodomain-coupled *cis*-regulation of genes have been demonstrated outside the nervous system, in vertebrates and invertebrates (Brugger et al., 2004; Walsh and Carroll, 2007). This suggests that homeodomain transcription factors can act with Mad/Medea in different contexts. Second, recent findings in our laboratory suggest that HD/BRE modules are prevalent in the *cis*-regulatory regions of BMP-responsive TDGs. For instance, we have found that *proctolin* is a BMP-dependent neuropeptide expressed in a small subset of Drosophila neurons termed the posterior cluster neurons (Cattaneo et al.). Its expression pattern is faithfully reproduced by a BMP-dependent enhancer that is ~500bp in length. Similar to *FMRFa*, both *proctoring* and its ~500bp enhancer are also regulated by a homeodomain transcription factor, HB9 (see Appendix F). This suggests that BMP signalling and HB9 may integrate at HD/BRE modules to activate *proctoring* expression. Several predicted HD/BREs exist in the *proctoring* enhancer, containing HD-, Mad- and Medea-binding motifs (see Appendix F). These predicted

modules differ from HD/BRE-A in the Tv enhancer in terms of positioning and exact sequences. However, it remains unclear whether these modules are functional, and whether they can be recognized by HB9, Mad and Medea.

Cis-regulatory analysis of another BMP-dependent neuropeptide, dilp7, offers a contrasting view on the action of BMP signalling and homeodomain transcription factors in regulating TDGs. Dilp7's expression pattern can be recapitulated by a ~1.5 kb enhancer (see Appendix G). However, individual deletion of two highly conserved, predicted HD/BRE modules failed to disrupt expression of the 1.0kb enhancer (data not shown). Furthermore, while both the dilp7 transcript and the 1.0 kb enhancer are regulated by HB9, only the transcript responds to BMP regulation (Miguel-Aliaga et al., 2008). These studies suggest that the action of homeodomain transcription factors and BMP signalling may be separable in some cases. Thus, further studies are needed to determine whether the Tv enhancer can serve as a paradigm for understanding how target-derived signalling integrate with intrinsic transcription factor codes in the nervous system.

7.2. Mechanism of *cis*-regulatory integration at the Tv enhancer

The FMRFa neuropeptide is expressed in 17 diverse neuronal types and its expression can be recapitulated by an 8.0 kb enhancer, which includes the 446 bp Tv enhancer (Benveniste et al., 1998). In this study I verified that Apterous and Zfh1 regulate the Tv enhancer, and discovered two additional regulators of this sequence: BMP signalling and Dachshund. It is striking that at least 4 of the 7 known *FMRFa* regulators in the Tv neurons act through a 446 base pair DNA sequence. The integration of multiple regulatory inputs into such a tight genomic region further illustrates the modular nature of transcriptional regulation.

7.2.1. Apterous and BMP signalling

Here, I provide evidence that Apterous and the BMP transcriptional regulators, Mad and Medea, can directly regulate Tv expression through association with the HD/BRE-A module. Interestingly, I observed that the HD-A is epistatic to Mad-A, supporting the notion that Apterous and Mad act synergistically through HD/BRE-A to regulate Tv expression. It is unlikely that the two predicted motifs are actually bound by one factor. Whereas mutation of each motif caused drastic down-regulation of the reporter, insertion of a 5-bp spacer between

Mad-A and HD-A only slightly reduced reporter activity. This result would not be expected if an unknown regulator binds onto an overlapping sequence between both predicted motifs. Further experiments would be needed to determine whether HD-A is also epistatic to Med-A.

Integration of BMP signalling and Apterous at HD/BRE-A alone probably does not sufficiently specify gene expression in the Tv neurons; a 3x tandem repeat of HD/BRE-A minimal sequence failed to direct reporter activity in the Tv neurons (data not shown). This is in spite of the fact that pMad and Apterous expression patterns overlap uniquely in the Tv neurons. It is possible that the 3xHD/BRE-A sequence does induce reporter expression in Tv neurons, but at levels below detection. It is also possible that BMP signalling and Apterous require additional factors for proper integration at HD/BRE-A. This is supported by the fact that directly upstream of HD/BRE-A lies a highly conserved stretch of sequence that our bioinformatics analysis predicts may bind nuclear hormone factors (Fig. 6). To rule out both possibilities, I constructed synthetic sequences carrying 6 and 12 tandem repeats of HD/BRE-A plus the predicted nuclear hormone factor binding site, and examined whether they can drive reporter expression in Tv neurons. Shortly after the completion of this thesis, I found that such a sequence indeed drove YFP expression in the Tv neurons (Appendix A, Fig. A.5.). Ectopic expression of this reporter was also detected in the brain lobes and in other neurons of the VNC, suggesting that additional flanking elements may be required to refine the spatial expression of *FMRFa* (Appendix A, Fig. A.5.). Furthermore, I found that the predicted nuclear hormone factor binding site is necessary for Tv enhancer expression (Appendix A, Fig. A.6.). These preliminary results now point to a model where Apterous, BMP signalling and an unknown factor act directly at the HD/BRE-A region to specify gene expression in the Tv neurons (Fig. 7.1).

Apterous likely contributes to *FMRFa* regulation through additional binding to the HD-B and C motifs. Although it is unclear whether Apterous acts alone at those motifs, the identification of a functional 50 bp region, directly adjacent to HD-B, plus the conservation of sequences surrounding the HD-B and C motifs, suggest that additional factors might be involved. BMP signalling is unlikely to be one of these factors. Although Mad-binding sequences surround HD -B and C, and can be bound by recombinant Mad in *vitro,* the same sequences were not necessary for Tv enhancer activation in *vivo*. This might explain the weak conservation of Mad-B and C relative to Mad-A. Several reasons could account for the observation that Mad-B and C are not functional in *vivo*. First, the spacing between

homeodomain-binding motifs and Mad-binding motifs in HD/BRE-B and C are different from that in HD/BRE-A. This could alter the topology of Mad binding relative to Apterous binding, and prevent the proper activator complexes from forming. Second, neither HD/BRE-B nor C has a nearby consensus Medea sequence, which is critical for formation of a pMad/Med complex. It would be interesting to test whether Mad-B and C can be manipulated into functional, BMP-responsive elements, by adding juxtaposed Medea motifs, when HD/BRE-A is mutated. However, as noted above, blocks of highly conserved sequences surround HD-B and C, and are likely to be functionally important. Alterations in spacing or insertion of a Medea binding motif in this region might disrupt the ability of unknown factors to interact with Apterous, and further confound experimental results. Instead, one might test the hypothesis that Mad-B and C are true Mad-binding motifs by determining whether either of them can functionally replace Mad-A at the HD/BRE-A module. Since $Tv^{Ins5Mad-A/HD-A}$-$nEYFP$ expresses at near-wildtype activity, changes in spacing from insertion of Mad-B and C might be tolerated by the Tv enhancer.

7.2.2. Dachshund and Zfh1

Currently, we do not know how Dachshund and Zfh1 regulate *FMRFa.*. However, there are reasons to believe that the two also directly act with BMP-signalling and Apterous at the Tv enhancer.

Dachshund is a transcriptional co-activator whose homologs are known to bind Smad complexes to regulate gene expression in *Drosophila* and in vertebrates (Kida et al., 2004; Takaesu et al., 2006; Wu et al., 2003). Although Dachshund only moderately regulates *Tv-nEYFP* expression, it can synergistically act with Apterous to trigger ectopic *Tv-nEYFP* expression in the nervous system. Furthermore, Dachshund and Apterous depends on BMP signalling for their ectopic activation of the 8.0 kb *FMRFa* enhancer (Miguel-Aliaga et al., 2004). Taken together, these observations suggest that Dachshund may be a modulator of the BMP pathway, and may directly interacts with Mad/Medea and Apterous at the HD/BRE-A to regulate *FMRFa* levels. Genetic interaction studies could help reveal functional relationships between Dachshund, Apterous and BMP-signalling. Preliminarily, I tested whether transheterozygous *ap* and *dac* mutants would show genetic interactions by affecting expression of the sensitized *Tv-nEYFP* reporter, but did not find any effect (data not shown).

It is curious to note that, using similar GAL4 drivers, misexpression of Dachshund alone induced ectopic expression of the Tv enhancer in many neurons, but only weakly activated the 8.0 kb *FMRFa* enhancer in several neurons (see Fig. 3.4.) (Miguel-Aliaga et al., 2004). On the other hand, Apterous cannot induce any ectopic *Tv-nEYFP* by itself. Thus, it is possible that within the *FMRFa* 8.0 kb enhancer, repressive elements suppress indiscriminate, BMP-dependent activation of the reporter in non-appropriate neurons. While over-expression of Dachshund may prime the pMad/Medea complex for indiscriminate activation of the 8.0 kb enhancer, further over-expression of Apterous is needed to overcome the repressive elements. When the repressive elements are absent, as is the case for *Tv-nEYFP,* over-expression of Dachshund can sufficiently induce BMP-dependent activation of the reporter, and Apterous further strengthens the response.

It is unclear how Zfh1 regulates *FMRFa* expression. Since the epistatic relationship between Zfh1 and other *FMRFa* regulators has not been determined, it is hard to evaluate whether Zfh1 directly regulates *FMRFa* through the Tv enhancer, or through regulation of another transcription factor. However, there are reasons to believe that Zfh1 can act directly with BMP signalling. SIP-1, the mammalian homolog of Zfh1, has been shown to bind to the MH2 domain of Smads (Verschueren et al., 1999). Although not yet tested, Zfh1's presumed ability to bind both Smads and homeodomain motifs suggests that it may also act with BMP signalling at HD/BRE-A. DNA binding experiments will be needed to test whether Zfh1 can associate with HD-A. A recent report suggests that Zfh1 can directly regulate Tv expression through binding to the HD-C motif (Vogler and Urban, 2008). However, it is unlikely that Zfh1 acts solely through HD-C binding, because *Tv-nEYFP* showed weaker expression in a *zfh1* mutant background than $Tv^{mMad\text{-}C/HD\text{-}C}\text{-}nEYFP$ in a wildtype background (Mad-C is not a functional element.)

7.3 A collaborative mechanism of *cis*-regulatory integration between BMP signalling and Apterous?

Although mutant analysis and enhancer mutagenesis experiments show a synergistic relationship between BMP signalling and Apterous in the activation of the Tv enhancer, it is unclear how the two factors might act together at the molecular level. A model of co-operative regulation through direct binding between transcriptional partners is attractive, given the strict,

evolutionarily conserved topology between Mad-A, HD-A and Med-A. However, co-immunoprecipitation experiments failed to lend support for the model that Apterous can interact with either Mad or Medea. On the other hand, my results are more in agreement with a collaborative regulatory mechanism where two transcription factors are required to bind in close proximity to each other, but do not directly interact. Binding of Apterous and the pMad/Medea complex may occur simultaneously on the HD/BRE-A module, with the two factors binding on opposite sides of the double helix. Insertion of a half-helical turn between Mad-A and Med-A binding sites may be tolerated, given the reported flexibility between Mad/Medea interactions (Gao and Laughon, 2007). However, such flexibility probably depends on the absence of additional interfering factors. By moving the HD-A motif onto the same helical side as that of Med-A, Apterous would be binding directly next to Medea, and this may have disrupted physical interactions between Medea and Mad and/or other factors, or vice versa, resulting in the dramatic down-regulation of Tv expression seen in $Tv\text{-}^{Ins5HD\text{-}A/Med\text{-}A}\text{-}nEYFP$.

7.4. Future questions

Despite the progress that my work has made, several issues have to be addressed. First, while mutations of Mad-A or Med-A led to reduction in reporter activities, the results failed to phenocopy the abolishment of reporter signal seen in a *wit* mutant background. This may be because the mutations used only partially disrupted Mad and Medea binding to HD/BRE-A. In support of this, the same Mad-A mutation used to disrupt reporter activity reduced, but not completely abolish MadN binding affinity for HD/BRE-A in *vitro*. If the Mad-A and Med-A mutations are indeed DNA-binding hypomorphs, a double-mutation of Mad-A and Med-A should further decrease reporter activity. Alternatively, wildtype BMP signalling may prevent the expression of the repressor molecule Brinker, which is known to repress BMP-dependent target genes in the absence of BMP signalling. The repression of Brinker in the wildtype backgrounds of $Tv^{mMad\text{-}A}\text{-}nEYFP$ and $Tv^{mMed\text{-}A}\text{-}nEYFP$ reporters might allow for leakiness of the Tv enhancer.

Second, it is unclear whether HD/BRE-A really responds to BMP signalling. I recently discovered that a 47 bp sequence including HD/BRE-A is sufficient for reporter expression in the Tv neurons (Appendix A, Fig. A.5.). Our model would be strengthened if it was shown that this sequence fails to drive reporter expression in a BMP mutant background. Another approach

would be to show Mad and Med binding to HDBRE-A in BMP-active conditions, but not when BMP signalling is suppressed. This would be best tested by EMSA in S2 cells.

Third, it is unknown whether Apterous, Mad and Medea can associate with the Tv enhancer in *vivo*. In a preliminary ChIP experiment, I found that a HA-tagged Apterous specifically associated with the Tv enhancer. However, mutation of all three HD-motifs failed to abolish the signal, suggesting saturation of DNA binding by Apterous. Thus, a more robust ChIP assay needs to be developed before this issue can be examined.

7.5 Conclusion

Taken together, my results provide the first mechanistic link between target-derived BMP signalling and homeodomain transcription factors in the regulation of TDGs. It remains to be seen whether this mechanism applies to other target-dependent TDGs, but important tools have now been generated to undergo further *cis*-regulatory studies (see Appendix F, G).
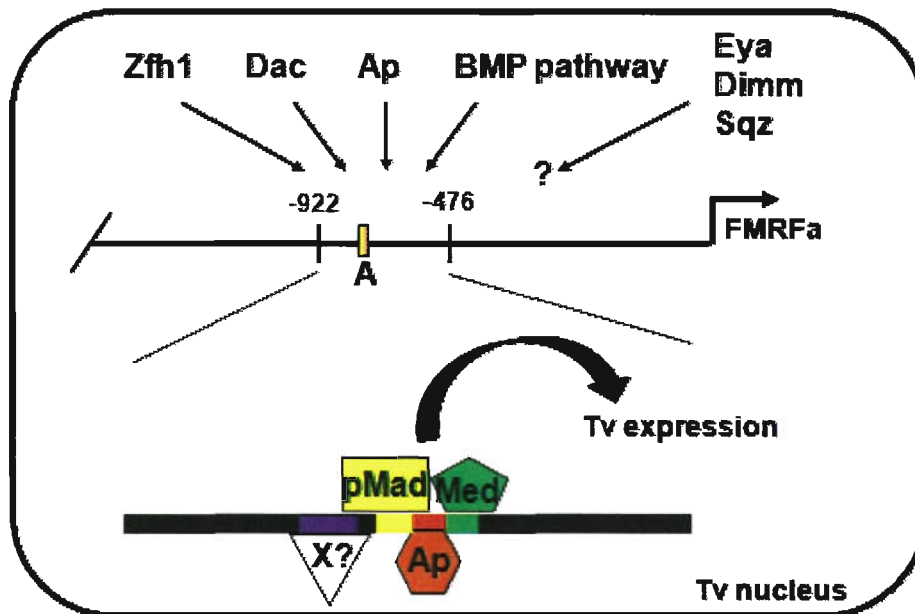
Figure 7.1. Model: Intersection of BMP signalling with Apterous at the Tv enhancer. Within the Tv nucleus, Apterous (Ap), Zfh1, Dachshund (Dac), and the BMP pathway is now know to regulate the expression of the Tv enhancer. It is still unclear whether Eyes absent (Eya), Dimmed (Dimm), or Squeeze (Sqz) also control Tv enhancer expression. Within the Tv enhancer, Apterous, Mad (pMad) and Medea is thought to bind directly to HD/BRE-A (yellow/red/green) and along with an unknown factor (X?) binding directly upstream of HD/BRE-A (purple), they form an activating complex that can specify gene expression in the Tv neurons.

53

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.

Affolter, M., Marty, T., Vigano, M.A., and Jazwinska, A. (2001). Nuclear interpretation of Dpp signaling in Drosophila. EMBO J *20*, 3298-3305.

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.

Arendt, D. (2005). Genes and homology in nervous system evolution: comparing gene functions, expression patterns, and cell type molecular fingerprints. Theory Biosci *124*, 185-197.

Arendt, D., and Nubler-Jung, K. (1996). Common ground plans in early brain development in mice and flies. Bioessays *18*, 255-259.

Arendt, D., and Nubler-Jung, K. (1999). Comparison of early nerve cord development in insects and vertebrates. Development *126*, 2309-2325.

Attisano, L., and Wrana, J.L. (2000). Smads as transcriptional co-modulators. Curr Opin Cell Biol *12*, 235-243.

Barolo, S., Castro, B., and Posakony, J.W. (2004). New Drosophila transgenic reporters: insulated P-element vectors expressing fast-maturing RFP. Biotechniques *36*, 436-440, 442.

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.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Pena-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., *et al.* (2008). Variation in homeodomain DNA binding revealed by high-resolution analysis of sequence preferences. Cell *133*, 1266-1276.

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.

Blanchette, M., Kent, W.J., Riemer, C., Elnitski, L., Smit, A.F., Roskin, K.M., Baertsch, R., Rosenbloom, K., Clawson, H., Green, E.D., *et al.* (2004). Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res *14*, 708-715.

Boules, M., Shaw, A., Fredrickson, P., and Richelson, E. (2007). Neurotensin agonists: potential in the treatment of schizophrenia. CNS Drugs *21*, 13-23.

Bourgouin, C., Lundgren, S.E., and Thomas, J.B. (1992). Apterous is a Drosophila LIM domain gene required for the development of a subset of embryonic muscles. Neuron *9*, 549-561.

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.

Briscoe, J., and Novitch, B.G. (2008). Regulatory pathways linking progenitor patterning, cell fates and neurogenesis in the ventral neural tube. Philos Trans R Soc Lond B Biol Sci *363*, 57-70.

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.

Bui, Q.T., Zimmerman, J.E., Liu, H., and Bonini, N.M. (2000). Molecular analysis of Drosophila eyes absent mutants reveals features of the conserved Eya domain. Genetics *155*, 709-720.

Cattaneo, A., Capsoni, S., Margotti, E., Righi, M., Kontsekova, E., Pavlik, P., Filipcik, P., and Novak, M. (1999). Functional blockade of tyrosine kinase A in the rat basal forebrain by a novel antagonistic anti-receptor monoclonal antibody. J Neurosci *19*, 9687-9697.

Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D. (2003). Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res *31*, 3497-3500.

Chiaromonte, F., Yap, V.B., and Miller, W. (2002). Scoring pairwise genomic sequence alignments. Pac Symp Biocomput, 115-126.

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.

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.

Coulombe, J.N., and Nishi, R. (1991). Stimulation of somatostatin expression in developing ciliary ganglion neurons by cells of the choroid layer. J Neurosci *11*, 553-562.

Dasen, J.S., Tice, B.C., Brenner-Morton, S., and Jessell, T.M. (2005). A Hox regulatory network establishes motor neuron pool identity and target-muscle connectivity. Cell *123*, 477-491.

Dietzl, G., Chen, D., Schnorrer, F., Su, K.C., Barinova, Y., Fellner, M., Gasser, B., Kinsey, K., Oppel, S., Scheiblauer, S., *et al.* (2007). A genome-wide transgenic RNAi library for conditional gene inactivation in Drosophila. Nature *448*, 151-156.

Duffy, J.B. (2002). GAL4 system in Drosophila: a fly geneticist's Swiss army knife. Genesis *34*, 1-15.

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.

Eaton, B.A., and Davis, G.W. (2005). LIM Kinase1 Controls Synaptic Stability Downstream of the Type II BMP Receptor. Neuron *47*, 695-708.

Edlund, T., and Jessell, T.M. (1999). Progression from extrinsic to intrinsic signaling in cell fate specification: a view from the nervous system. Cell *96*, 211-224.

Ernsberger, U., and Rohrer, H. (1999). Development of the cholinergic neurotransmitter phenotype in postganglionic sympathetic neurons. Cell Tissue Res *297*, 339-361.

Etchberger, J.F., Lorch, A., Sleumer, M.C., Zapf, R., Jones, S.J., Marra, M.A., Holt, R.A., Moerman, D.G., and Hobert, O. (2007). The molecular signature and cis-regulatory architecture of a C. elegans gustatory neuron. Genes Dev *21*, 1653-1674.

Faresse, N., Colland, F., Ferrand, N., Prunier, C., Bourgeade, M.F., and Atfi, A. (2008). Identification of PCTA, a TGIF antagonist that promotes PML function in TGF-beta signalling. EMBO J *27*, 1804-1815.

Gao, S., and Laughon, A. (2007). Flexible interaction of Drosophila Smad complexes with bipartite binding sites. Biochim Biophys Acta *1769*, 484-496.

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.

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.

Hall, A.K., Burke, R.M., Anand, M., and Dinsio, K.J. (2002). Activin and bone morphogenetic proteins are present in perinatal sensory neuron target tissues that induce neuropeptides. J Neurobiol *52*, 52-60.

Harrison, K.A., Thaler, J., Pfaff, S.L., Gu, H., and Kehrl, J.H. (1999). Pancreas dorsal lobe agenesis and abnormal islets of Langerhans in Hlxb9-deficient mice. Nat Genet *23*, 71-75.

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., and Westphal, H. (2000). Functions of LIM-homeobox genes. Trends Genet *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.

Horton, R.M., Cai, Z.L., Ho, S.N., and Pease, L.R. (1990). Gene splicing by overlap extension: tailor-made genes using the polymerase chain reaction. Biotechniques *8*, 528-535.

Hunter, C.S., and Rhodes, S.J. (2005). LIM-homeodomain genes in mammalian development and human disease. Mol Biol Rep *32*, 67-77.

Jessell, T.M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nat Rev Genet *1*, 20-29.

Jorgensen, J.S., Quirk, C.C., and Nilson, J.H. (2004). Multiple and overlapping combinatorial codes orchestrate hormonal responsiveness and dictate cell-specific expression of the genes encoding luteinizing hormone. Endocr Rev *25*, 521-542.

Justice, R.W., Zilian, O., Woods, D.F., Noll, M., and Bryant, P.J. (1995). The Drosophila tumor suppressor gene warts encodes a homolog of human myotonic dystrophy kinase and is required for the control of cell shape and proliferation. Genes Dev *9*, 534-546.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome Res *12*, 656-664.

Kida, Y., Maeda, Y., Shiraishi, T., Suzuki, T., and Ogura, T. (2004). Chick Dach1 interacts with the Smad complex and Sin3a to control AER formation and limb development along the proximodistal axis. Development *131*, 4179-4187.

Kim, J., Johnson, K., Chen, H.J., Carroll, S., and Laughon, A. (1998). *Drosophila* Mad binds to DNA and directly mediates activation of *vestigial* by Decapentaplegic. Nature *388*, 304-308.

Koo, S.J., and Pfaff, S.L. (2002). Fine-tuning motor neuron properties: signaling from the periphery. Neuron *35*, 823-826.

Kuhn, R.M., Karolchik, D., Zweig, A.S., Trumbower, H., Thomas, D.J., Thakkapallayil, A., Sugnet, C.W., Stanke, M., Smith, K.E., Siepel, A., *et al.* (2007). The UCSC genome browser database: update 2007. Nucleic Acids Res *35*, D668-673.

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.

Landgraf, M., and Thor, S. (2006). Development of Drosophila motoneurons: specification and morphology. Semin Cell Dev Biol *17*, 3-11.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., *et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Lechleider, R.J., Ryan, J.L., Garrett, L., Eng, C., Deng, C., Wynshaw-Boris, A., and Roberts, A.B. (2001). Targeted mutagenesis of Smad1 reveals an essential role in chorioallantoic fusion. Dev Biol *240*, 157-167.

Lee-Hoeflich, S.T., Causing, C.G., Podkowa, M., Zhao, X., Wrana, J.L., and Attisano, L. (2004). Activation of LIMK1 by binding to the BMP receptor, BMPRII, regulates BMP-dependent dendritogenesis. Embo J *23*, 4792-4801.

Lewis, D.A., and Levitt, P. (2002). Schizophrenia as a disorder of neurodevelopment. Annu Rev Neurosci *25*, 409-432.

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

Lovoulos, C., Tittle, S., Goldstein, L., Austin, D.J., Singh, S., Rocco, E., Keane, J., Tang, P., Kopf, G.S., and Elefteriades, J.A. (2004). Right ventricle-sparing heart transplantation effective against iatrogenic pulmonary hypertension. J Heart Lung Transplant *23*, 236-241.

Markstein, M., Pitsouli, C., Villalta, C., Celniker, S.E., and Perrimon, N. (2008). Exploiting position effects and the gypsy retrovirus insulator to engineer precisely expressed transgenes. Nat Genet *40*, 476-483.

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.

Massague, J., and Gomis, R.R. (2006). The logic of TGFbeta signaling. FEBS Lett *580*, 2811-2820.

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

Massague, J., and Wotton, D. (2000). Transcriptional control by the TGF-beta/Smad signaling system. EMBO J *19*, 1745-1754.

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.

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.

Monuki, E.S., Porter, F.D., and Walsh, C.A. (2001). Patterning of the dorsal telencephalon and cerebral cortex by a roof plate-Lhx2 pathway. Neuron *32*, 591-604.

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'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.

Odenwald, W.F., Rasband, W., Kuzin, A., and Brody, T. (2005). EVOPRINTER, a multigenomic comparative tool for rapid identification of functionally important DNA. Proc Natl Acad Sci U S A *102*, 14700-14705.

Oppenheim, R.W. (1989). The neurotrophic theory and naturally occurring motoneuron death. Trends Neurosci *12*, 252-255.

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.

Pfaff, S.L., Mendelsohn, M., Stewart, C.L., Edlund, T., and Jessell, T.M. (1996). Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. Cell *84*, 309-320.

Rawson, J.M., Lee, M., Kennedy, E.L., and Selleck, S.B. (2003). Drosophila neuromuscular synapse assembly and function require the TGF-beta type I receptor saxophone and the transcription factor Mad. J Neurobiol *55*, 134-150.

Saito, T., Iwata, N., Tsubuki, S., Takaki, Y., Takano, J., Huang, S.M., Suemoto, T., Higuchi, M., and

Saido, T.C. (2005). Somatostatin regulates brain amyloid beta peptide Abeta42 through modulation of proteolytic degradation. Nat Med *11*, 434-439.

Saito, T., Takaki, Y., Iwata, N., Trojanowski, J., and Saido, T.C. (2003). Alzheimer's disease, neuropeptides, neuropeptidase, and amyloid-beta peptide metabolism. Sci Aging Knowledge Environ *2003*, PE1.

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, 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.

Shen, W., Finnegan, S., Lein, P., Sullivan, S., Slaughter, M., and Higgins, D. (2004). Bone morphogenetic proteins regulate ionotropic glutamate receptors in human retina. Eur J Neurosci *20*, 2031-2037.

Sheng, H.Z., Zhadanov, A.B., Mosinger, B., Jr., Fujii, T., Bertuzzi, S., Grinberg, A., Lee, E.J., Huang, S.P., Mahon, K.A., and Westphal, H. (1996). Specification of pituitary cell lineages by the LIM homeobox gene Lhx3. Science *272*, 1004-1007.

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

Shirasaki, R., and Pfaff, S.L. (2002). Transcriptional codes and the control of neuronal identity. Annu Rev Neurosci *25*, 251-281.

Siepel, A., Bejerano, G., Pedersen, J.S., Hinrichs, A.S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L.W., Richards, S., *et al.* (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res *15*, 1034-1050.

Sirard, C., de la Pompa, J.L., Elia, A., Itie, A., Mirtsos, C., Cheung, A., Hahn, S., Wakeham, A., Schwartz, L., Kern, S.E., *et al.* (1998). The tumor suppressor gene Smad4/Dpc4 is required for gastrulation and later for anterior development of the mouse embryo. Genes Dev *12*, 107-119.

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

Sofroniew, M.V., Howe, C.L., and Mobley, W.C. (2001). Nerve growth factor signaling, neuroprotection, and neural repair. Annu Rev Neurosci *24*, 1217-1281.

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.

Stephan, K.E., Baldeweg, T., and Friston, K.J. (2006). Synaptic plasticity and dysconnection in schizophrenia. Biol Psychiatry *59*, 929-939.

Sun, M., Thomas, M.J., Herder, R., Bofenkamp, M.L., Selleck, S.B., and O'Connor, M.B. (2007). Presynaptic contributions of chordin to hippocampal plasticity and spatial learning. J Neurosci *27*, 7740-7750.

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.

Svingen, T., and Tonissen, K.F. (2006). Hox transcription factors and their elusive mammalian gene targets. Heredity *97*, 88-96.

Takaesu, N.T., Hyman-Walsh, C., Ye, Y., Wisotzkey, R.G., Stinchfield, M.J., O'Connor, M.B., Wotton, D., and Newfeld, S.J. (2006). dSno Facilitates Baboon Signaling in the Drosophila Brain by Switching the Affinity of Medea Away From Mad and Toward dSmad2. Genetics *174*, 1299-1313.

Tavsanli, B.C., Ostrin, E.J., Burgess, H.K., Middlebrooks, B.W., Pham, T.A., and Mardon, G. (2004). Structure-function analysis of the Drosophila retinal determination protein Dachshund. Dev Biol *272*, 231-247.

Thaler, J., Harrison, K., Sharma, K., Lettieri, K., Kehrl, J., and Pfaff, S.L. (1999). Active suppression of interneuron programs within developing motor neurons revealed by analysis of homeodomain factor HB9. Neuron *23*, 675-687.

Thaler, J.P., Koo, S.J., Kania, A., Lettieri, K., Andrews, S., Cox, C., Jessell, T.M., and Pfaff, S.L. (2004). A postmitotic role for Isl-class LIM homeodomain proteins in the assignment of visceral spinal motor neuron identity. Neuron *41*, 337-350.

Thor, S., and Thomas, J. (2002). Motor neuron specification in worms, flies and mice: conserved and 'lost' mechanisms. Curr Opin Genet Dev *12*, 558-564.

Tremblay, K.D., Dunn, N.R., and Robertson, E.J. (2001). Mouse embryos lacking Smad1 signals display defects in extra-embryonic tissues and germ cell formation. Development *128*, 3609-3621.

Verschueren, K., Remacle, J.E., Collart, C., Kraft, H., Baker, B.S., Tylzanowski, P., Nelles, L., Wuytens, G., Su, M.-T., Bodmer, R., *et al.* (1999). SIP1, a Novel Zinc Finger/Homeodomain Repressor, Interacts with Smad Proteins and Binds to 5'-CACCT Sequences in Candidate Target Genes. J. Biol. Chem. *274*, 20489-20498.

Vogler, G., and Urban, J. (2008). The transcription factor Zfh1 is involved in the regulation of neuropeptide expression and growth of larval neuromuscular junctions in Drosophila melanogaster. Dev Biol *319*, 78-85.

von Bubnoff, A., Peiffer, D.A., Blitz, I.L., Hayata, T., Ogata, S., Zeng, Q., Trunnell, M., and Cho, K.W. (2005). Phylogenetic footprinting and genome scanning identify vertebrate BMP response elements and new target genes. Dev Biol *281*, 210-226.

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

Wenick, A.S., and Hobert, O. (2004). Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in C. elegans. Dev Cell *6*, 757-770.

Withers, G.S., Higgins, D., Charette, M., and Banker, G. (2000). Bone morphogenetic protein-7 enhances dendritic growth and receptivity to innervation in cultured hippocampal neurons. Eur J Neurosci *12*, 106-116.

Wotton, D., and Massague, J. (2001). Smad transcriptional corepressors in TGF beta family signaling. Curr Top Microbiol Immunol *254*, 145-164.

Wu, K., Yang, Y., Wang, C., Davoli, M.A., D'Amico, M., Li, A., Cveklova, K., Kozmik, Z., Lisanti, M.P., and Russell, R.G.e.a. (2003). DACH1 inhibits transforming growth factor-beta signaling through binding Smad4. J. Biol. Chem. *278*, 51673-51684.

Xiao, Z., Liu, X., and Lodish, H.F. (2000). Importin beta mediates nuclear translocation of Smad 3. J Biol Chem *275*, 23425-23428.

Xu, P., and Hall, A.K. (2006). The role of activin in neuropeptide induction and pain sensation. Dev Biol *299*, 303-309.

Yang, X., Li, C., Xu, X., and Deng, C. (1998). The tumor suppressor SMAD4/DPC4 is essential for epiblast proliferation and mesoderm induction in mice. Proc Natl Acad Sci U S A *95*, 3667-3672.

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.

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.

Appendices

Appendix A – *FMRFa*/Tv enhancer-related images
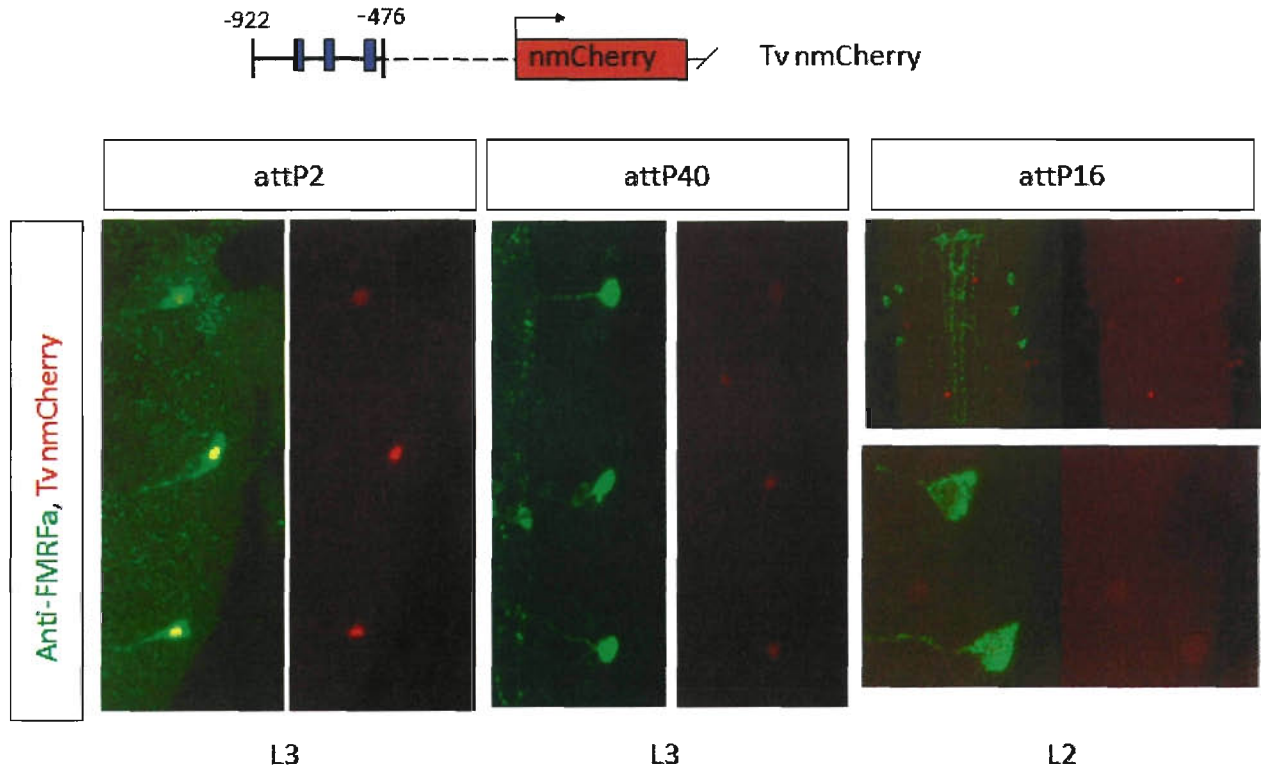
Figure A.1. Screen for optimal attP loci for Tv-nmCherry expression. Schematic representation of Tv-nmCherry, Stacked projection of larval CNS showing anti-FMRFa (green) and Tv-nmCherry (red) expression in the Tv neurons. Tv-nmCherry was integrated into similar vector backbone as Tv-nEYFP. Similar results were obtained for Tv-nEYFP integrated into attP2 and attP40. attP2 was found to be optimal attP site out of the three candidates, as judged by eye.

adult

Tv YFP, antiFMRF
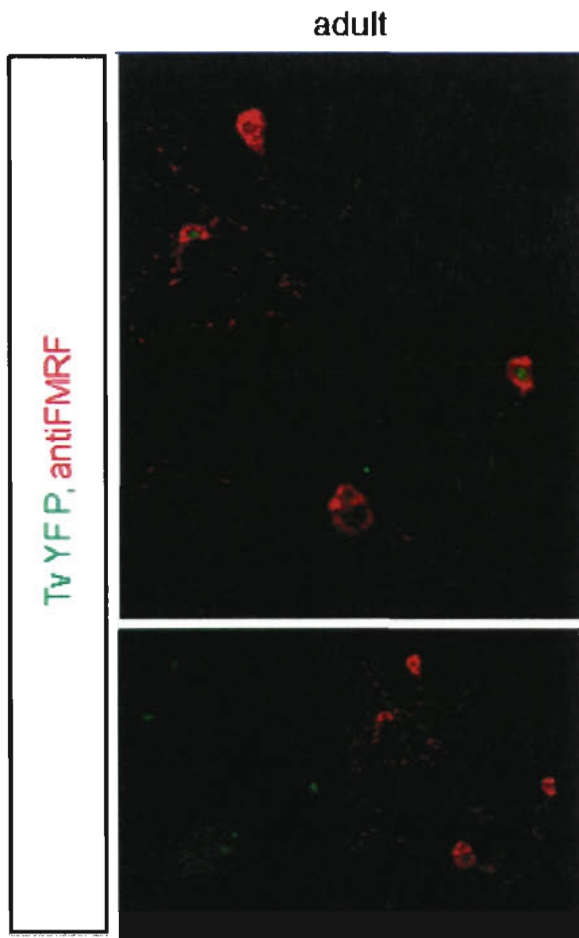
Figure A.2. *Tv-nEYFP* **is expressed in adult Tv neurons.** Stacked projection of adult CNS showing anti-FMRFa staining (red) and Tv-nEYFP expression (green).
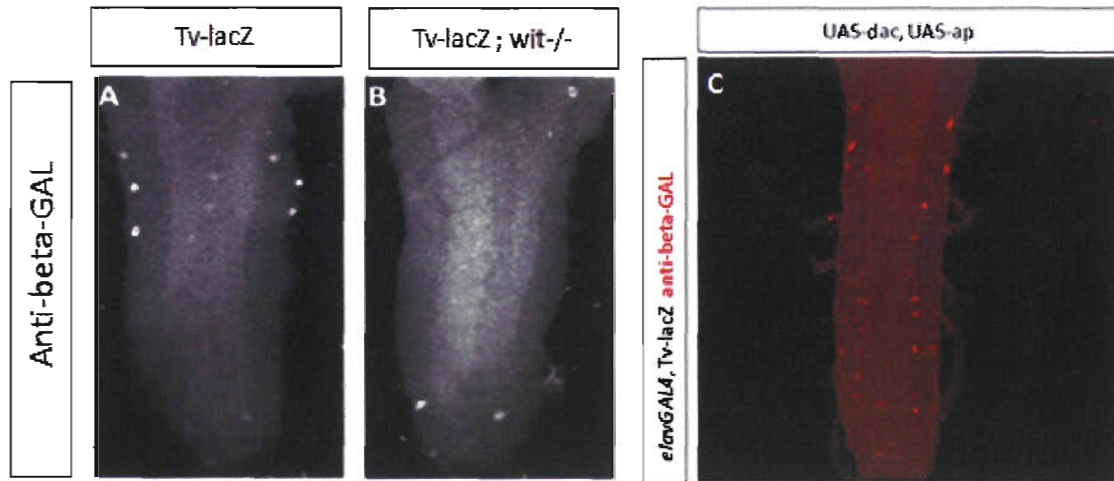
Figure A.3. Tv-lacZ is regulated by BMP signalling and can respond to Ap and Dac. Stacked projection of larval (A,B) and embryonic (C) CNS showing anti-beta-galactosidase staining for *Tv-lacZ* expression. **A, B.** *Tv-lacZ* expression is eliminated in *witA12/B11* mutant background (compare A (witA12/+) with B (witA12/witB11)). **C.** *Tv-lacZ* can also be induced by misexpression of Apterous and Dachshund in post-mitotic neurons by the pan-neuronal GAL4 driver, *elavGAL4.*
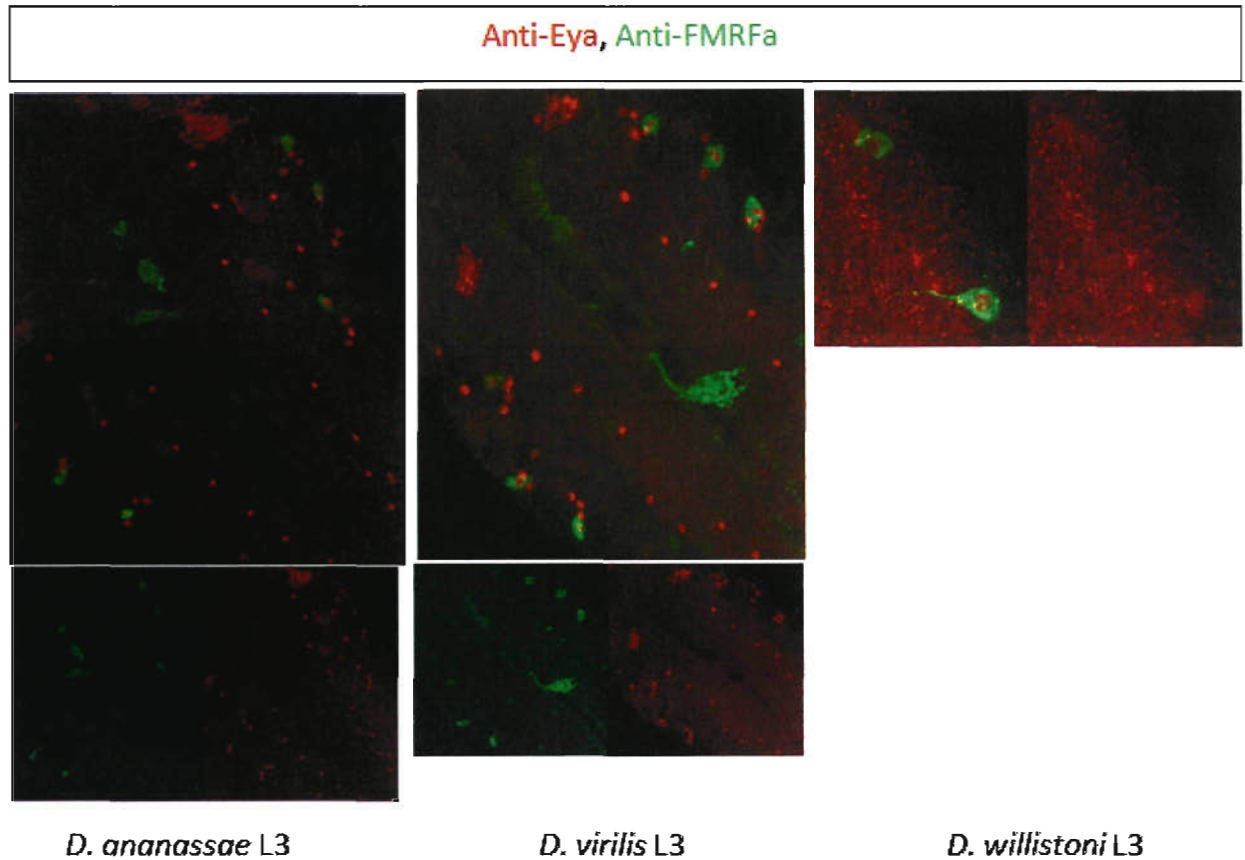
Figure A.4. Immunostaining of three divergent _Drosophila_ species reveal the conserved expression of _FMRFa_ in the stereotypical Tv neurons. Stacked projection of L3 larval CNS from _D. ananassae, D. virilise and D. willistoni._ FMRFa peptide was detected by anti-FMRFa (green) in all three species. Overlaid and individual channel images are shown, except in _D.willistoni._ Expression of the peptide is highly restricted to the stereotypical Tv cluster, as marked by anti-Eya (red).
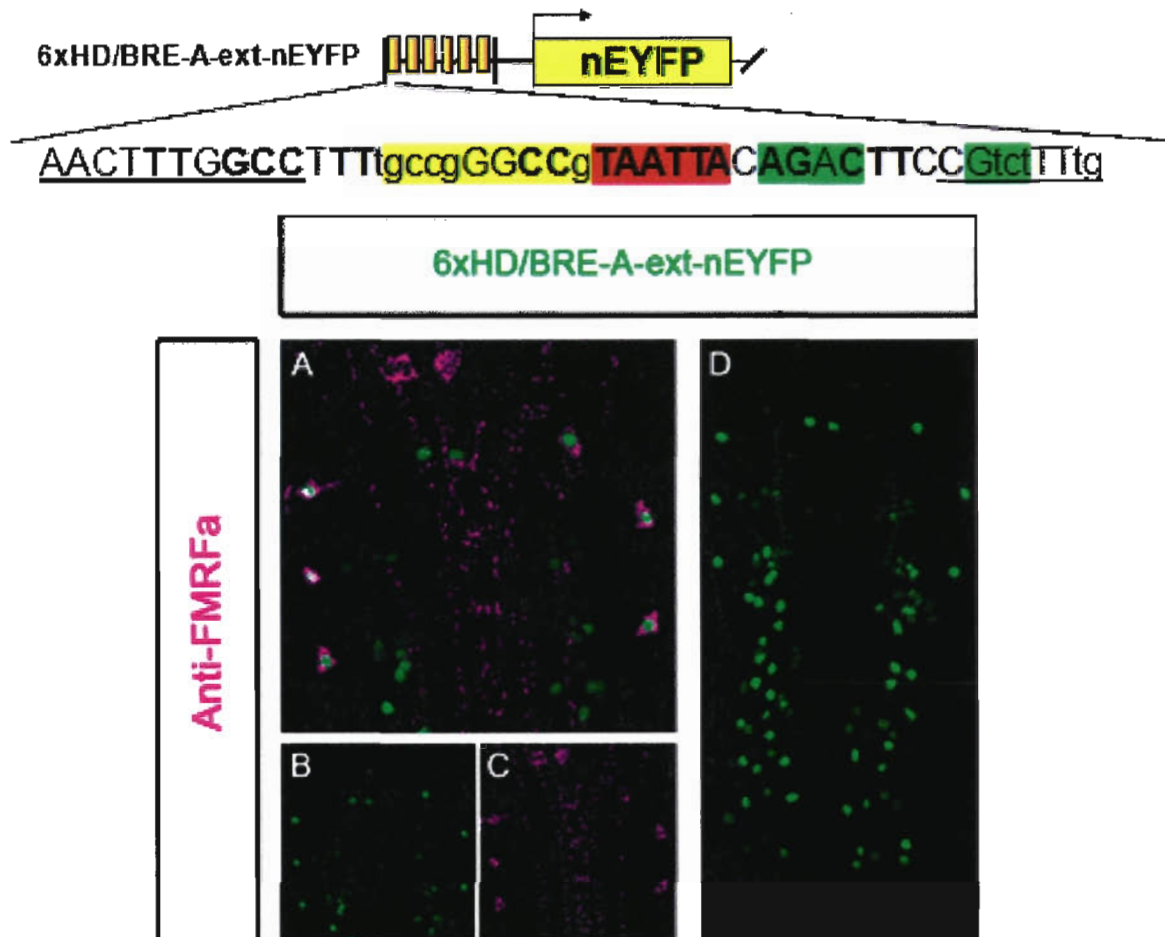
Figure A.5. A 47 bp sequence including HD/BRE-A can drive reporter expression in the Tv neurons. Schematic representation of 6xHD/BRE-A-ext-nEYFP. The exact sequence used for the repeats is shown. Capitalized letters are highly conserved throughout *Drosophila* evolution. Underlined letters indicate additional sequences added to a minimal HD/BRE-A sequence that failed to drive reporter expression in the Tv neurons (data not shown). Stacked projection of L2 larval CNS expressing nEYFP (green in box A, B, D) driven by 6xHD/BRE-A-ext sequence. FMRFa peptide was detected by anti-FMRFa (magenta in box A, C). Box D shows the extent of ectopic YFP expression along the VNC.
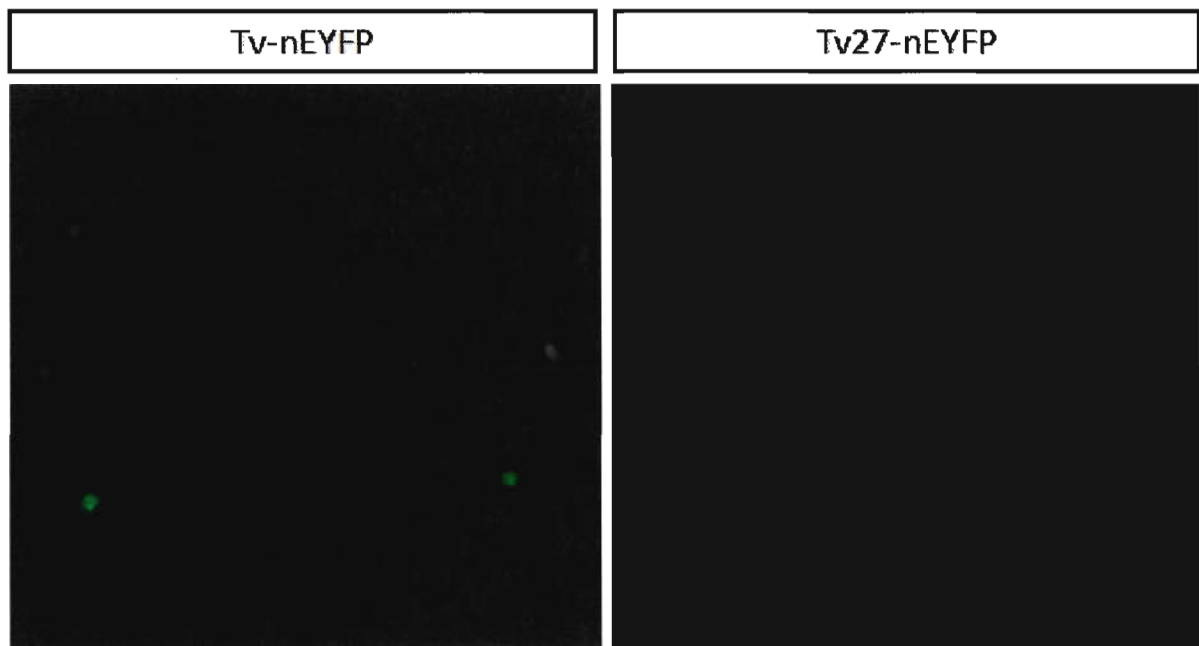
Figure A.6. An additional sequence upstream of HD/BRE-A is required for reporter expression in the Tv neurons. Exact sequences showing region surrounding HD/BRE-A (yellow/red/green). Tv wt sequence was mutated at a conserved stretch of sequence just 5' of HD/BRE-A (Tv27). Capitalized letters are highly conserved throughout *Drosophila* evolution. Underlined letters indicate additional sequences added to a minimal HD/BRE-A sequence that failed to drive reporter expression in the Tv neurons (data not shown). Stacked projection of L3 larval CNS expressing nEYFP (green) driven by Tv wt or Tv27.

Appendix B – *Tv-nEYFP* data analysis tables

Table B.1. Single-copy Tv-nEYFP expression in transcription factor mutant background

Gene	Genotype	mean # YFP neurons/ CNS	standard error	Number of CNS	ttest vs. *Tv-nEYFP*	ttest vs. *Tv-nEYFP*, mutant/+
	Tv-nEYFP/+	5.7	0.1	18		
wit	*Tv-nEYFP, witA12/+*	5.5	0.2	11	4.3E-01	
wit	*Tv-nEYFP, witA12/witB11*	0.0	0.0	12	6.6E-24	2.1E-16
dac	*Df(2L)Exel7086/+ ; Tv-nEYFP/+*	5.5	0.2	11	4.3E-01	
dac	*Df(2L)Exel7086/dac3 ; Tv-nEYFP/+*	4.8	0.2	14	1.2E-03	5.2E-02
ap	*apGAL4/+ ; Tv-nEYFP/+*	3.0	0.4	10	8.3E-08	
ap	*apGAL4/apP44 ; Tv-nEYFP/+*	0.7	0.3	10	1.2E-15	3.1E-04
zfh1	*Tv-nEYFP, zfh100865/+*	5.4	0.2	11	2.2E-01	
zfh1	*Tv-nEYFP, zfh100865/zfh100865*	0.7	0.6	9	1.6E-11	1.5E-06

Table B.2. Single-copy *Tv-nEYFP* expression after misexpression of transgenes by *OK6-GAL4*

OK6 GAL4 x	Average # of ectopic YFP neurons/CNS	Standard Error	Number of CNS	ttest vs. UAS Ap	ttest vs. UAS Dac
w1118	0.0	0.0	1		
UAS Ap	0.0	0.0	2		
UAS Dac	51.0	6.2	3	8.0E-03	
UAS Ap, UAS Dac	74.3	3.8	3		3.3E-02

Table B.3. Cell count analysis of heterozygous *Tv-nEYFP* reporters

Reporter	Mean #YFP Tv neurons	Standard Error	Number of CNS	ttest relative to Tv^{wt}-nEYFP	ttest relative to Tv^{mMad-A}-nEYFP	ttest relative to Tv^{mMad-B}-nEYFP	ttest relative to Tv^{mMad-C}-nEYFP
Tv^{wt}-nEYFP	5.8	0.1	72				
Tv^{mMad-A}-nEYFP	3.7	0.5	7	2.5E-13			
Tv^{mMad-B}-nEYFP	5.9	0.1	9	4.3E-01			
Tv^{mMad-C}-nEYFP	6.0	0.0	12	8.0E-02			
$Tv^{mMad-A/B}$-nEYFP	4.0	0.3	7	1.4E-13	3.5E-01	1.8E-05	
$Tv^{mMad-A/C}$-nEYFP	4.5	0.5	4	3.9E-06	6.5E-01		6.4E-05
$Tv^{mMad-B/C}$-nEYFP	5.7	0.2	6	6.2E-01		3.3E-01	3.5E-02
$Tv^{mMad-A/B/C}$-nEYFP	4.0	0.0	3	5.5E-09	7.4E-01	2.5E-06	N/A
$Tv^{mMad-A/HD-A}$-nEYFP	0.1	0.1	8	4.1E-48	7.7E-06		
$Tv^{mMad-B/HD-B}$-nEYFP	5.7	0.2	6	6.2E-01		3.3E-01	
$Tv^{mMad-C/HD-C}$-nEYFP	5.5	0.2	6	1.9E-01			4.9E-03
$Tv^{del\ cons.\ 50bp}$-nEYFP	0.0	0.0	8	4.3E-51			

Table B.4. Intensity measurement of heterozygous *Tv-nEYFP* reporters

Reporter	Relative fluorescence intensity of YFP-positive neuron relative to wildtype (%)	Standard Error	# neurons quantified	ttest vs. Tv^{wt}-nEYFP	ttest vs. Tv^{mMad-A}-nEYFP	ttest vs. Tv^{mMad-B}-nEYFP	ttest vs. Tv^{mMad-C}-nEYFP
Tv^{wt}-nEYFP	100.0	3.0	524				
Tv^{mMad-A}-nEYFP	44.8	5.1	35	3.0E-06			
Tv^{mMad-B}-nEYFP	111.9	10.4	48	2.5E-01			
Tv^{mMad-C}-nEYFP	101.5	7.8	61	8.7E-01			
Tv$^{mMad-A/B}$-nEYFP	26.3	2.4	36	2.8E-10	1.4E-03	1.1E-09	
Tv$^{mMad-A/C}$-nEYFP	39.9	5.6	27	7.4E-06	5.2E-01		3.1E-06
Tv$^{mMad-B/C}$-nEYFP	97.5	10.7	41	8.2E-01		3.4E-01	7.6E-01
Tv$^{mMad-A/B/C}$-nEYFP	42.0	6.5	33	9.9E-07	7.3E-01	1.2E-06	1.1E-06
Tv$^{mMad-A/HD-A}$-nEYFP	7.3	N/A	1	N/A	N/A		
Tv$^{mMad-B/HD-B}$-nEYFP	26.8	2.7	46	1.8E-12		9.1E-12	
Tv$^{mMad-C/HD-C}$-nEYFP	49.8	7.6	41	5.4E-06			1.6E-05

Table B.5. Cell count analysis of homozygous *Tv-nEYFP* reporters

Reporter	Mean # YFP neurons/cord	Standard Error	Number of CNS	ttest vs. Tv-nEYFP	ttest vs. Tv-mMad-A/HD-A-nEYFP	ttest vs. Tv mMed-A-nEYFP	ttest vs. Tv mMad-A-nEYFP
Tv^{wt}-nEYFP	6.0	0.0	45				
Tv^{mMad-A}-nEYFP	5.8	0.3	4	1.0E-01			
$Tv^{mMad-A-4bp}$-nEYFP	5.7	0.3	3	4.6E-02			8.5E-01
Tv^{mHD-A}-nEYFP	0.4	0.2	7	1.4E-44	1.9E-01		
Tv-$^{mMad-A/HD-A}$nEYFP	1.0	0.0	2	2.5E-33			
Tv^{mMed-A}-nEYFP	5.3	0.3	4	2.8E-06			
Tv^{m3Med}-nEYFP	5.3	0.3	6	1.5E-04		8.6E-01	
$Tv^{Ins5Mad-A/HD-A}$-nEYFP	5.7	0.3	3	4.6E-02			
$Tv^{Ins5HD-A/Med-A}$-nEYFP	4.7	0.3	3	6.8E-12			
$Tv^{delcons50}$-nEYFP	0.0	0.0	5	6.5E-48			

Table B.6. Intensity measurements of homozygous *Tv-nEYFP* reporters

Reporter	Relative fluorescence intensity of YFP-positive neuron vs. wildtype (%)	Standard Error	#neurons quantified	ttest vs. Tv^{wt}-nEYFP	ttest vs. Tv^{mMad-A}-nEYFP	ttest vs. Tv^{mMed-A}-nEYFP	ttest vs Tv^{mHD-A}-nEYFP
Tv^{wt}-nEYFP	100.0	5.7	277				
Tv^{mMad-A}-nEYFP	30.7	3.8	23	2.7E-09			
$Tv^{mMad-A-4bp}$-nEYFP	20.5	3.7	47	1.2E-21	9.2E-02		
Tv^{mHD-A}-nEYFP	9.7	7.0	5	2.7E-04			
$Tv\text{-}^{mMad-A/HD-A}$-nEYFP	5.3	1.7	3	2.3E-03			6.5E-01
Tv^{mMed-A}-nEYFP	12.5	1.4	60	1.5E-12			
Tv^{m3Med}-nEYFP	25.6	3.2	35	5.2E-09		9.0E-05	
$Tv^{Ins5Mad-A/HD-A}$-nEYFP	77.6	4.9	53	2.8E-03			
$Tv^{Ins5HD-A/Med-A}$-nEYFP	5.6	0.8	24	1.1E-13			

Appendix C – Tv mutant constructs primers

Table C.1. *Tv-nEYFP* reporters used in this study

Reporter	Region of mutation 5'-3'
Tv^{wt}-nEYFP	none
Tv^{mMad-A}-nEYFP	TTTTAGTAGTACGTAATTA
Tv^{mMad-B}-nEYFP	AAAGCGCCATAAAGTAGTAGTAGTAAATGGCAAATTATA
Tv^{mMad-C}-nEYFP	GGCTAATTGGAAGTAGTAGTAGTAGATGTCCCTGCT
Tv^{mMad-A/B}-nEYFP	combine mutations in mMadA and B
Tv^{mMad-A/C}-nEYFP	combine mutations in mMadA and C
Tv^{mMad-B/C}-nEYFP	combine mutations in mMadB and C
Tv^{mMad-A/B/C}-nEYFP	combine mutations in mMadA, B and C
Tv^{mMad-A/HD-A}-nEYFP	TTTTAGTAGTACGGAGCTCCAGAC
Tv^{mMad-B/HD-B}-nEYFP	AAAGCGCCATAAAGTAGTAGTAGTAAATGGCAAGAGCTC
Tv^{mMad-C/HD-C}-nEYFP	GGCGAGCTCGAAGTAGTAGTAGTAGATGTCCCTGCT
Tv^{del cons. 50bp}-nEYFP	deleted this region: TGCCAGAGGCGCCACAATGTATCCTGTTACAGGTTACAGGGCCATAAAGC
Tv^{mHD-A}-nEYFP	TTTTGCCGGGCCGGAGCTCCAGAC
Tv^{m3Med}-nEYFP	AGACTCGTTCCCAAAACTTT......CGTAATTACGAACTTCCGTTC
Tv^{mMed-A}-nEYFP	CGTAATTACGAACTTCCGTCT
Tv^{mMad-A-4bp}-nEYFP	GGCCTTTTGCTAGTACGTAATTA
Tv^{Ins5Mad-A/HD-A}-nEYFP	GCCGGGCCGTTGTGTAATTACAGAC
Tv^{Ins5HD-A/Med-A}-nEYFP	GCCGGGCCGTAATTACTCAACAGAC
3xHD/BRE-A-nEYFP	3x tandem repeat of : TTTGCCGGGCCGTAATTACAGACTTC

Note: Red font indicate mutated or inserted sequences

Table C.2. Primer sequences for generating Tv mutants

Tv mutant	Method/ Amplified Segment	Template	Forward 5'-3'	Reverse 5'-3'	Restriction site for coining
Tv^{wt}	PCR segment	genomic DNA	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	XbaI/EcoRI
Tv^{mMad-A}	SOE PCR segment 1	Tv$^{mMadA/B/C}$	CGGTCTAGAGCCATCTGCAGACGTGGT	GACGGAAGTCTGTAATTACG	
	SOE PCR segment 2	Tv^{wt}	CGTAATTACAGACTTCCGTCTTTTGAACA	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^{mMad-A}	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	XbaI/EcoRI
Tv^{mMad-B}	SOE PCR segment 1	Tv^{wt}	CGGTCTAGAGCCATCTGCAGACGTGGT	TTGCCATTTACTACTACTACTTTATGGCGCTTTATGGC	
	SOE PCR segment 2	Tv^{wt}	GTAGTAGTAGTAAATGGCAAATTATAACGCATACG	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^{mMad-B}	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	XbaI/EcoRI
Tv^{mMad-C}	PCR segment	Tv$^{mMadA/B/C}$	CGGTCTAGAGCCATCTGCAGACGTGGT	TAAGAATTCAATGAGCAGGGACATCTACTACTACTACTTCCAATTAGCCTTCTAGC	XbaI/EcoRI
Tv$^{mMad-A/B}$	SOE PCR segment 1	Tv$^{mMadA/B/C}$	CGGTCTAGAGCCATCTGCAGACGTGGT	TTGCCATTTACTACTACTACTTTATGGCGCTTTATGGC	
	SOE PCR segment 2	Tv^{wt}	GTAGTAGTAGTAAATGGCAAATTATAACGCATACG	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv$^{mMad-A/B}$	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	XbaI/EcoRI
Tv$^{mMad-A/C}$	SOE PCR segment 1	Tv$^{mMadA/B/C}$	CGGTCTAGAGCCATCTGCAGACGTGGT	GACGGAAGTCTGTAATTACG	
	SOE PCR segment 2	Tv^{wt}	CGTAATTACAGACTTCCGTCTTTTGAACA	TAAGAATTCAATGAGCAGGGACATCTACTACTACTACTTCCAATTAGCCTTCTAGC	
	SOE PCR final segment	Tv$^{mMadA/C}$	CGGTCTAGAGCCATCTGCAGACGTGGT	TAAGAATTCAATGAGCAGGGACATCTACTACTACTACTTCCAATTAGCCTTCTAGC	XbaI/EcoRI

Table C.2. (continued)

Tv mutant	Method/ Amplified Segment	Template	Forward 5'-3'	Reverse 5'-3'	Restriction site for coining
Tv^mMad-B/C	SOE PCR segment 1	Tv^wt	CGGTCTAGAGCCATCTGCAGACGTGGT	TTGCCATTTACTACTACTACTTTATGGCGCTTTATGGC	
	SOE PCR segment 2	Tv^mMadA/B/C	GTAGTAGTAGTAAATGGCAAATTATAACGCATACG	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^mMad-B/C	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	Xbal/EcoRI
Tv^mMad-A/B/C	SOE PCR segment 1	Tv^wt	CGGTCTAGAGCCATCTGCAGACGTGGT	GTCTGTAATTACGTACTACTAAAAGGCCAAAGTTTTGGAGACGAGTCT	
	SOE PCR segment 2	Tv^wt	AGTAGTACGTAATTACAGACTTCCGTCTTTTGAACAGTTTTTTCAGC	TTGCCATTTACTACTACTACTTTATGGCGCTTTATGGC	
	SOE PCR segment 3	Tv^wt	GTAGTAGTAGTAAATGGCAAATTATAACGCATACG	TAAGAATTCAATGAGCAGGGACATCTACTACTACTTCCAATTAGCCTTCTAGC	
	SOE PCR final segment	Tv^mMadA/B/C	CGGTCTAGAGCCATCTGCAGACGTGGT	TAAGAATTCAATGAGCAGGGACATCTACTACTACTTCCAATTAGCCTTCTAGC	Xbal/EcoRI
Tv^mMad-A/mHD-A	SOE PCR segment 1	Tv^wt	CGGTCTAGAGCCATCTGCAGACGTGGT	GTCTGGAGCTCCGTACTACTAAAAGGCCAAAGTTTTGG	
	SOE PCR segment 2	Tv^wt	AGTAGTACGGAGCTCCAGACTTCCGTCTTTTGAACA	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^mMad-A/mHD-A	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	Xbal/EcoRI
Tv^mMad-B/mHD-B	SOE PCR segment 1	Tv^wt	CGGTCTAGAGCCATCTGCAGACGTGGT	TTGCCATTTACTACTACTACTTTATGGCGCTTTATGGCCCTGTAACCTG	
	SOE PCR segment 2	Tv^wt	AATGGCAAATTATAACGCATACGGACACG	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^mMad-B/mHD-B	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	Xbal/EcoRI
Tv^mMad-C/mHD-C	PCR segment	Tv^wt	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATCTACTACTACTTCGAGCTCGCCTTCTAGCCAGTGGAT	

Table C.2. (continued)

Tv mutant	Method/ Amplified Segment	Template	Forward 5'-3'	Reverse 5'-3'	Restriction site for coining
Tv^{del cons 50bp}	SOE PCR segment 1	Tv^{wt}	CGGTCTAGAGCCATCTGCAGACGTGGT	ATTGCCGTCGCGGCGTTTATGGCTCTCGTCCAGCACCCGAA	
	SOE PCR segment 2	Tv^{wt}	GCCATAAACGCCGCG	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^{del cons 50bp}	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	Xbal/EcoRI
Tv^{mHD-A}-nEYFP	SOE PCR segment 1	Tv^{wt}	CGGTCTAGAGCCATCTGCAGACGTGGT	GGCTGAAAAAACTGTTCAAAAGACGGAAGTCTGGAGCTCCGGCCCGGCAAAAGG	
	SOE PCR segment 2	Tv^{wt}	TTTGAACAGTTTTTTCAGCCCCACCCA	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^{mHD-A}	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	Xbal/EcoRI
Tv^{m3Med}	SOE PCR segment 1	Tv^{wt}	CGGTCTAGAGCCATCTGCAGACGTGGT	GGCTGAAAAAACTGTTCAAAGAACGGAAGTTCGTAATTACGGCCCGGCAAAAGGCCAAAGTTTTGGGAACGAGTTC	
	SOE PCR segment 2	Tv^{wt}	TTTGAACAGTTTTTTCAGCCCCACCCA	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^{m3Med}	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	Xbal/EcoRI
Tv^{mMed-A}	SOE PCR segment 1	Tv^{wt}	CGGTCTAGAGCCATCTGCAGACGTGGT	GCTGAAAAAACTGTTCAAAAGACGGAAGTTCGTAATTACGGCCCGGC	
	SOE PCR segment 2	Tv^{wt}	TTTGAACAGTTTTTTCAGCCCCACCCA	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^{mMed-A}	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	Xbal/EcoRI

Table C.2. (continued)

Tv mutant	Method/ Amplified Segment	Template	Forward 5'-3'	Reverse 5'-3'	Restriction site for cloning
Tv^{mMad-A-4bp}	SOE PCR segment 1	Tv^{wt}	CGGTCTAGAGCCATCTGCAGACGTGGT	CTGAAAAAAACTGTTCAAAAGACGGAAGTCTGTAATTACGTACTAGCAAAAGGCCAAAGT	
	SOE PCR segment 2	Tv^{wt}	TTTGAACAGTTTTTTCAGCCCCACCCA	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^{mMad-A-4bp}	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	XbaI/EcoRI
Tv^{mMad-A-4bp}	SOE PCR segment 1	Tv^{wt}	CGGTCTAGAGCCATCTGCAGACGTGGT	CTGAAAAAAACTGTTCAAAAGACGGAAGTCTGTAATTACACAACGGCCCGGCAAAAGGCC	
	SOE PCR segment 2	Tv^{wt}	TTTGAACAGTTTTTTCAGCCCCACCCA	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^{mMad-A-4bp}	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	XbaI/EcoRI
Tv^{Ins5HD-A/Med-A}	SOE PCR segment 1	Tv^{wt}	CGGTCTAGAGCCATCTGCAGACGTGGT	CTGAAAAAAACTGTTCAAAAGACGGAAGTCTGTTGAGTAATTACGGCCCGGCAAAAGGCC	
	SOE PCR segment 2	Tv^{wt}	TTTGAACAGTTTTTTCAGCCCCACCCA	CCCGAATTCAATGAGCAGGGACATC	
	SOE PCR final segment	Tv^{Ins5HD-A/Med-A}	CGGTCTAGAGCCATCTGCAGACGTGGT	CCCGAATTCAATGAGCAGGGACATC	XbaI/EcoRI
3xHD/BRE-A	Anneal/Primer extension segment	None	CGGTCTAGATTTGCCGGGCCGTAATTACAGACTTCTTTGCCGGGCCGTAATTAC	CCCGAATTCGAAGTCTGTAATTACGGCCCGGCAAAGAAGTCTGTAATTACGGC	XbaI/EcoRI

80

Appendix D – EMSA oligonucleotides

Table D.1. List of oligonucleotides used for EMSA experiments

Oligo set	Ordered oligo name	Oligonucleotide sequence 5'-3'
Tv HDBREA with 2 Medea sites wt		
Tv HDBREA with 2 Medea sites wt		
	TvBRE1w2Med sense	TTG GCC TTT TGC CGG GCC GTA ATT ACA GAC TTC CGT CTT TTG A
	TvBRE1w2Med antisense	TCA AAA GAC GGA AGT CTG TAA TTA CGG CCC GGC AAA AGG CCA A
Tv HDBREA mut2Med		
	TvBRE1w2Med mut2Med sense	TTG GCC TTT TGC CGG GCC GTA ATT ACC AAC TTC CGT TCT TTG A
	TvBRE1w2Med mut2Med antisense	TCA AAG AAC GGA AGT TGG TAA TTA CGG CCC GGC AAA AGG CCA A
	Tv1w2Me m2MeB sen	TTG GCC TTT TGC CGG GCC GTA ATT ACA GTA TTC CAG TAT TTG A
	Tv1w2Me m2MeB antisen	TCA AAT ACT GGA ATA CTG TAA TTA CGG CCC GGC AAA AGG CCA A
Tv HDBREA mutMad+2Med		
	TvBRE1w2Med mutMad2Med sense	TTG GCC TTT TAG TAG TAC GTA ATT ACC AAC TTC CGT TCT TTG A
	TvBRE1w2Med mutMad2Med antisen	TCA AAG AAC GGA AGT TGG TAA TTA CGT ACT ACT AAA AGG CCA A
	TvBRE1w2Me mutMa2Me B sense	TTG GCC TTT TAG TAG TAC GTA ATT ACA GTA TTC CAG TAT TTG A
	TvBRE1w2Me mutMa2Me B antisen	TCA AAT ACT GGA ATA CTG TAA TTA CGT ACT ACT AAA AGG CCA A
Tv HDBREA mutate adjacent GGCC + Mad+2Med		
	Tv1w2Me mutGCMe A sense	TTA GTA TTT TAG TAG TAC GTA ATT ACC AAC TTC CGT TCT TTG A
	Tv1w2Me mutGCMe A antisense	TCA AAG AAC GGA AGT TGG TAA TTA CGT ACT ACT AAA ATA CTA A
	Tv1w2Me mutGCMe B sense	TTG AAG TTT TAG TAG TAC GTA ATT ACA GTA TTC CAG TAT TTG A
	Tv1w2Me mutGCMe B antisense	TCA AAT ACT GGA ATA CTG TAA TTA CGT ACT ACT AAA ACT TCA A

Note: Yellow highlighted oligonucleotides indicate those that were described in the thesis. Underlined text denotes start of a new set of EMSA oligonucleotides.

Table D.1. (continued)

Oligo set	Ordered oligo name	Oligonucleotide sequence 5'-3'
Tv HDBREA mutMad		
	Tv1w2Me mMad sense	TTG GCC TTT TAG TAG TAC GTA ATT ACA GAC TTC CGT CTT TTG A
	Tv1w2Me mMad antisen	TCA AAA GAC GGA AGT CTG TAA TTA CGT ACT ACT AAA AGG CCA A
TvHDBREA mutMad+Med1		
	Tv1w2Me mMaMe1B sen	TTG GCC TTT TAG TAG TAC GTA ATT ACA GTA TTC CGT CTT TTG A
	Tv1w2Me mMaMe1B antisen	TCA AAA GAC GGA ATA CTG TAA TTA CGT ACT ACT AAA AGG CCA A
Tv HDBREA mutMad+Med2		
	Tv1w2Me mMaMe2B sen	TTG GCC TTT TAG TAG TAC GTA ATT ACA GAC TTC CAG TAT TTG A
	Tv1w2Me mMaMe2B antisen	TCA AAT ACT GGA AGT CTG TAA TTA CGT ACT ACT AAA AGG CCA A
<u>Tv HDBREA noMed wt</u>		
Tv HDBREA noMed wt		
	TvBRE1noMed sense	TTGGCCTTTTGCCGGGCCGTAATTACAG
	TvBRE1noMed antisense	CTGTAATTACGGCCCGGCAAAAGGCCAA
Tv HDBREA noMed mutMad		
	TvBRE1noMed mutMad sense	TTGGCCTTTTAGTAGTACGTAATTACAG
	TvBRE1noMed mutMad antisen	CTGTAATTACGTACTACTAAAAGGCCAA
Tv HDBREA noMed mutate adjacent GGCC sequence		
	Tv1 noMe mutGC A sen	TTAGTATTTTAGTAGTACGTAATTACAG
	Tv1 noMe mutGC A antisen	CTGTAATTACGTACTACTAAAATACTAA
	Tv1 noMe mutGC B sen	TTGAAGTTTTAGTAGTACGTAATTACAG
	Tv1 noMe mutGC B antisen	CTGTAATTACGTACTACTAAAATACTAA

Note: Yellow highlighted oligonucleotides indicate those that were described in the thesis. Underlined text denotes start of a new set of EMSA oligonucleotides.

Table D.1. (continued)

Oligo set	Ordered oligo name	Oligonucleotide sequence 5'-3'
<u>Tv HDBREA with1 Medea site</u>		
Tv HDBREA with1 Medea site wt		
	TvBRE1w1Med sense	TTG GCC TTT TGC CGG GCC GTA ATT ACA GAC TTC CG
	TvBRE1w1Med antisense	CGG AAG TCT GTA ATT ACG GCC CGG CAA AAG GCC AA
Tv HDBREA mutMedea site		
	TvBRE1w1Med mutMad sense	TTG GCC TTT TAG TAG TAC GTA ATT ACA GAC TTC CG
	TvBRE1w1Med mutMad antisense	CGG AAG TCT GTA ATT ACG TAC TAC TAA AAG GCC AA
Tv HDBREA mutHD site		
	TvBRE1w1Med mutHD sense	TTG GCC TTT TGC CGG GCC GGA GCT CCA GAC TTC CG
	TvBRE1w1Med mutHD antisense	CGG AAG TCT GGA GCT CCG GCC CGG CAA AAG GCC AA
<u>Tv HDBREA + 2Medea sites without MadA</u>		
Tv HDBREA + 2Medea sites without MadA wt		
	TvBRE1 2Medonly sense	CGT AAT TAC AGA CTT CCG TCT TTT GA
	Tv1 2Medonly antisense	TCA AAA GAC GGA AGT CTG TAA TTA CG
Tv HDBREA mutated both Medea sites		
	Tv1 2Medonly mut2Me sense	CGT AAT TAC GAA CTT CCG TTC TTT GA
	Tv1 2Medonly mut2Me antisen	TCA AAG AAC GGA AGT TCG TAA TTA CG
<u>Tv HDBREB</u>		
Tv HDBREB wt		
	TvBRE2 sense	GCC ATA AAC GCC GCG ACG GCA ATG GCA AAT TAT AAC GCA TAC
	TvBRE2 antisense	GTA TGC GTT ATA ATT TGC CAT TGC CGT CGC GGC GTT TAT GGC

Note: Yellow highlighted oligonucleotides indicate those that were described in the thesis. Underlined text denotes start of a new set of EMSA oligonucleotides.

Table D.1. (continued)

Oligo set	Ordered oligo name	Oligonucleotide sequence 5'-3'
Tv HDBREB mutate Mad site		
	Tv2 mutMa sense	GCC ATA AAG TAG TAG TAG TAA ATG GCA AAT TAT AAC GCA TAC
	Tv2 mutMa anti	GTA TGC GTT ATA ATT TGC CAT TTA CTA CTA CTA CTT TAT GGC
Tv HDBREB mutate HD site		
	Tv2 mutHD sense	GCC ATA AAC GCC GCG ACG GCA ATG GCA AGA GCT CAC GCA TAC
	Tv2 mutHD antisen	GTA TGC GTG AGC TCT TGC CAT TGC CGT CGC GGC GTT TAT GGC
Tv HDBREC		
Tv HDBREC wt		
	TvBRE3 sense	GCT AGA AGG CTA ATT GGA CGT GCC CGG CCA GGA TGT CCC TGC
	TvBRE3 antisense	GCA GGG ACA TCC TGG CCG GGC ACG TCC AAT TAG CCT TCT AGC
Tv HDBREC mutate Mad site		
	Tv3 mutMa sense	GCT AGA AGG CTA ATT GGA AGT AGT AGT AGT AGA TGT CCC TGC
	Tv3 mutMa antisense	GCA GGG ACA TCT ACT ACT ACT ACT TCC AAT TAG CCT TCT AGC
Tv HDBREC mutate HD site		
	Tv3 mutHD sense	GCT AGA AGG CGA GCT CGA CGT GCC CGG CCA GGA TGT CCC TGC
	Tv3 mutHD anti	GCA GGG ACA TCC TGG CCG GGC ACG TCG AGC TCG CCT TCT AGC
Misc. oligos		
	BrkS EMSA sense	AAT TCG ACT GGC GAC ATT CTG TCT GTG GCG ATC GCG GCC
	BrkS EMSA antisen	GGC CGC GAT CGC CAC AGA CAG AAT GTC GCC AGT CGA ATT
	Q+ vg comp EMSA sense	TTT GTG CTT GGC TGC CGT CGC GAT TCG ACA ACT TTG G
	Q+ vg comp EMSA antisense	CCA AAG TTG TCG AAT CGC GAC GGC AGC CAA GCA CAA A
	Qm vg comp EMSA sense	TTT GTG CTT GAG ATC TAG ATC TAT TCG ACA ACT TTG G
	Qm vg comp EMSA antisense	CCA AAG TTG TCG AAT AGA TCT AGA TCT CAA GCA CAA A
	sal M1 EMSA sense	AAT CAT ATT AAG ACG GGC ACA TTA TAA A
	sal M1 EMSA antisense	TTT ATA ATG TGC CCG TCT TAA TAT GAT T
	sal pm808 EMSA sense	AAT CAT ATT AAA ACG GGC ACA TTA TAA A

Note: Yellow highlighted oligonucleotides indicate those that were described in the thesis. Underlined text denotes start of a new set of EMSA oligonucleotides.

Table D.1. (continued)

Oligo set	Ordered oligo name	Oligonucleotide sequence 5'-3'
	sal pm808 EMSA antisense	TTT ATA ATG TGC CCG TTT TAA TAT GAT T
	Vg EMSA sense	CTTGGCTGCCGTCGCGATTC
	Vg EMSA antisense	GAATCGCGACGGCAGCCAAG
Tv con50 region		
	Tv mutcon50 A sense	TGG ACG AGA TGC CAG AGG CGC CAC AAT GTA TCC TGC CGC AGG TTA CAG G
	Tv mutcon50 A antisense	CCT GTA ACC TGC GGC AGG ATA CAT TGT GGC GCC TCT GGC ATC TCG TCC A
	Tv mutcon50B sense	CAG GCC GCA GGG CCA TAA AGC GCC ATA AAC
	Tv mutcon50B antisense	GTT TAT GGC GCT TTA TGG CCC TGC GGC CTG
	Tv mutbef50 sense	TTG AAA AGC TGG CTG GGA TGG GGT GGC CCC GGG TGC TGG ACG AGA T
	Tv mutbef50 anti	ATC TCG TCC AGC ACC CGG GGC CAC CCC ATC CCA GCC AGC TTT TCA A

Note: Yellow highlighted oligonucleotides indicate those that were described in the thesis. Underlined text denotes start of a new set of EMSA oligonucleotides.
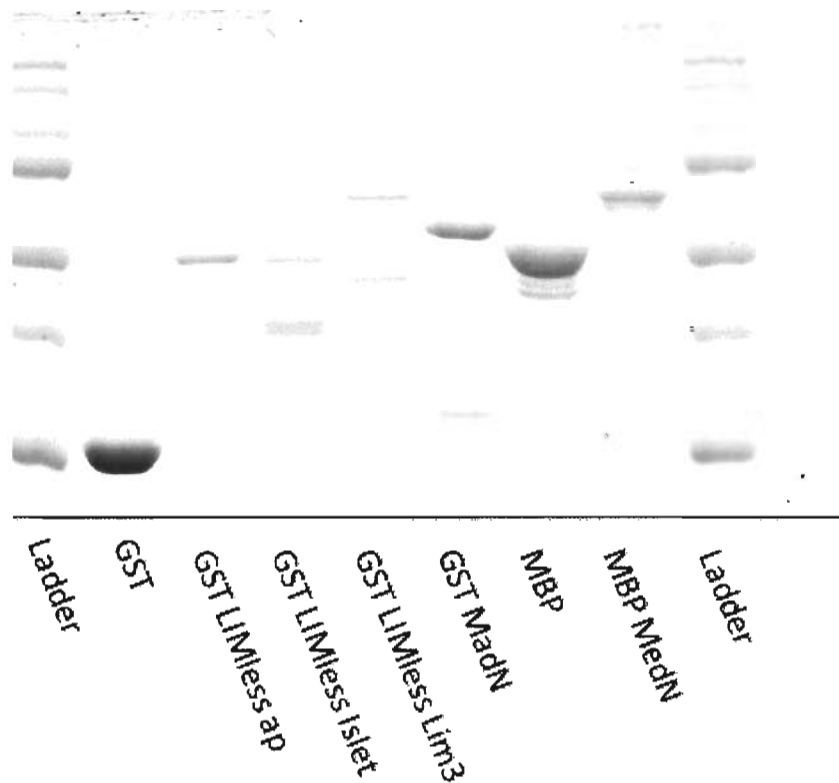
Appendix E – EMSA-related figures

Figure E.1.SDS-PAGE analysis of bacterially purified proteins used for EMSA. Bacterially expressed GST, GST-LIMless Ap, GST-LIMless Islet, GST-LIMless Lim3, GST-MadN, MBP and MBP-MedN were quantified by UV Spectroscopy at 280nm. 3ug of products were loaded onto a 10% gel, ran and stained with coomassie blue. Ladder is BioRad Precision Plus.
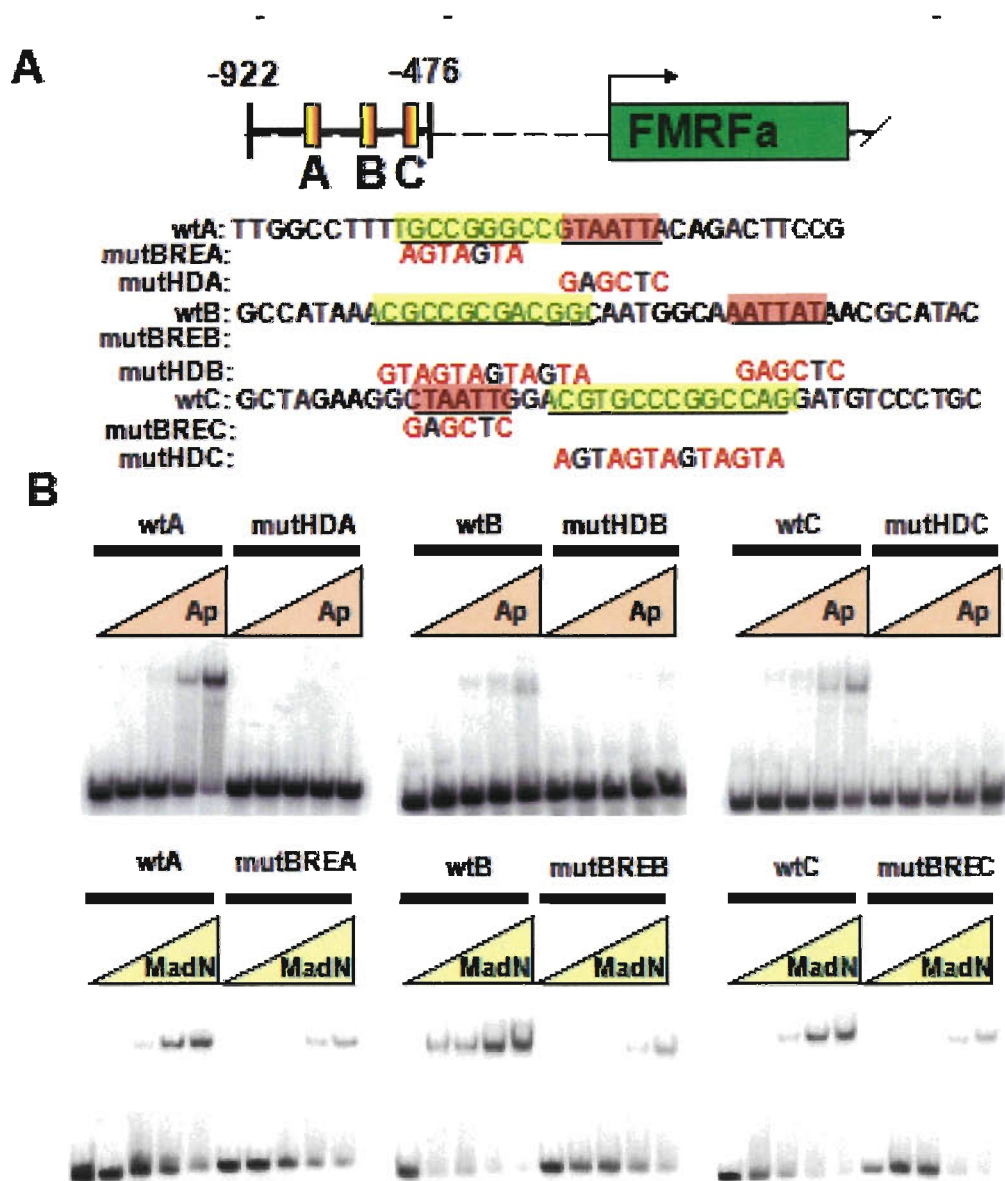
Figure E.2. Apterous and Mad can bind specifically with all three predicted HD/BRE sequences. A. Schematic representation of *FMRFa* locus, with Tv enhancer marked by its position relative to the *FMRFa* transcriptional start site. HDBRE-A ,B and C are marked in red/yellow box in the enhancer. Wildtype and mutant sequences of oligonucleotides containing homeodomain-binding sites (red) and predicted Mad-binding sites (yellow) are shown. Red font indicates mutant bases. Exact sequences of oligonucleotides used for EMSA in (B) are shown, except for wtA, which is missing the sequence 5'ACTTCCG-3' in the 3' end in this particular EMSA experiment. **B.** Bacterially-purified GST-LIMless Ap (top panel) and GST-MadN (bottom panel) can all interact specifically with each HDBRE in *vitro*. Binding activity is lost when each protein is tested for binding to oligonucleotides bearing mutations in their corresponding binding sites.
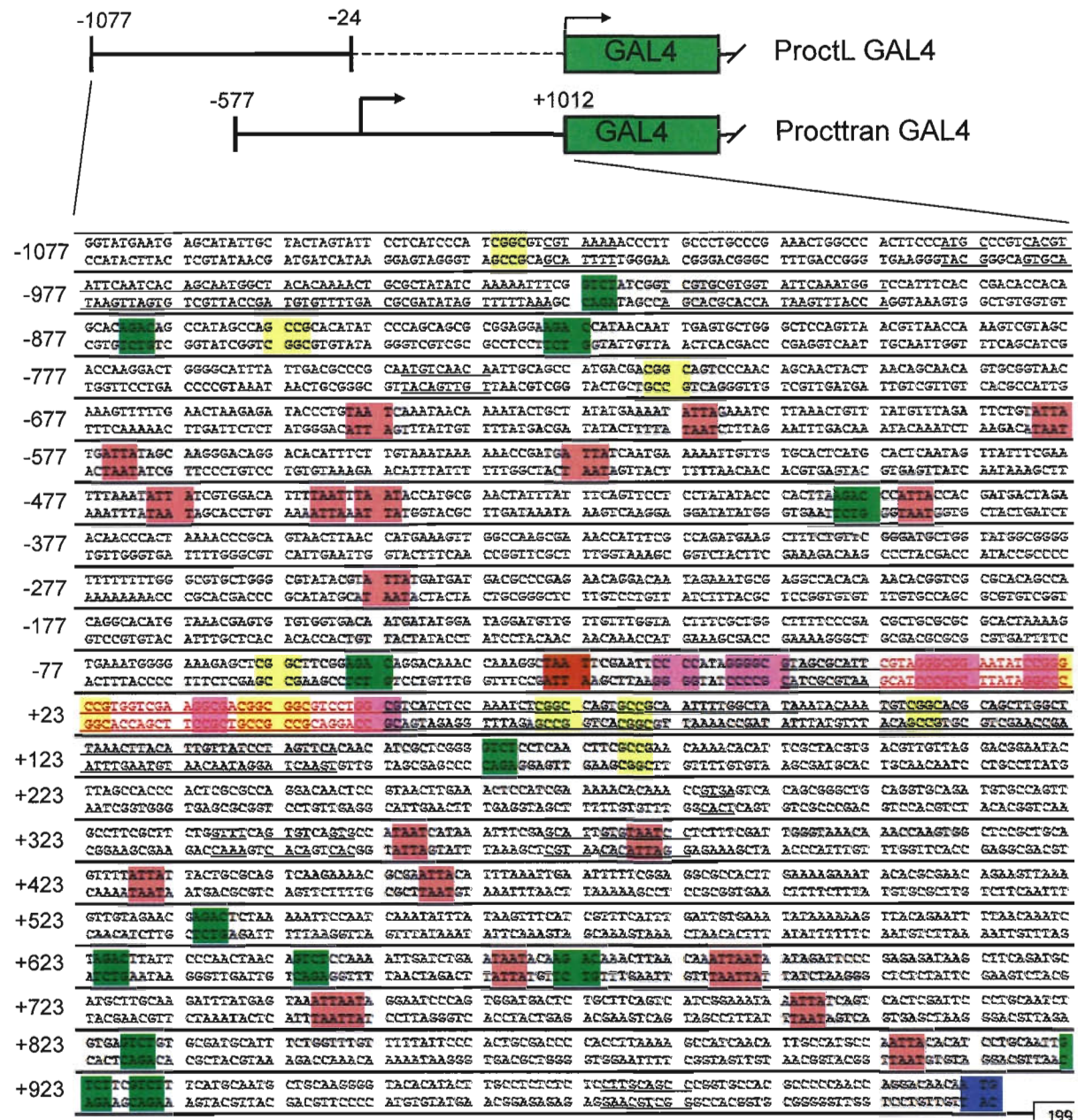
88

Appendix F – Proctolin

Figure F.1. Detailed analysis of *proctolin* cis-regulatory region included in the *proctolin* reporters. Evolutionarily conserved cluster of sequences across *Drosophila* species ((*D.melanogaster, D. erecta, D. persimilis, D. pseudoobscura, D.virilis,* and *D.grimshawi*) are underlined. The putative *proctolin* core promoter (red font) is found to be surrounded by a cluster of homeodomain consensus (highlight in red), Mad consensus or near consensus (highlight in yellow or purple)/Medea consensus (highlight in green) sequences. Start methionine of *GAL4(*and also *proctolin)* is highlighted in blue.
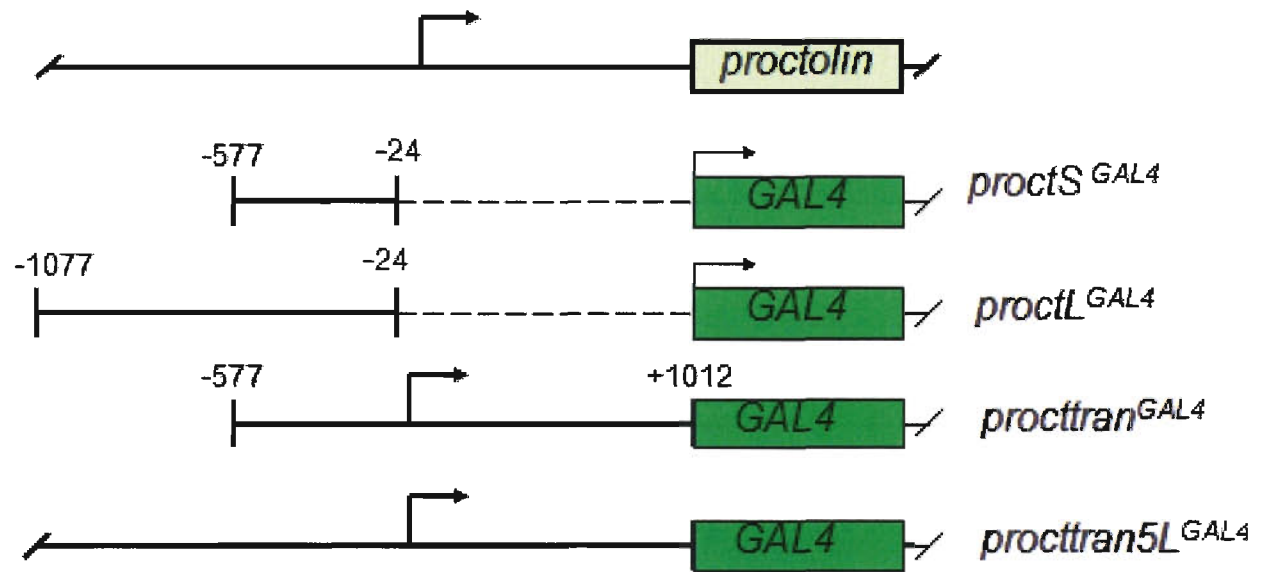
Figure F.2. Proctolin enhancer-GAL4 constructs. Schematic representation of *proctolin* enhancers linked to GAL4 drivers.
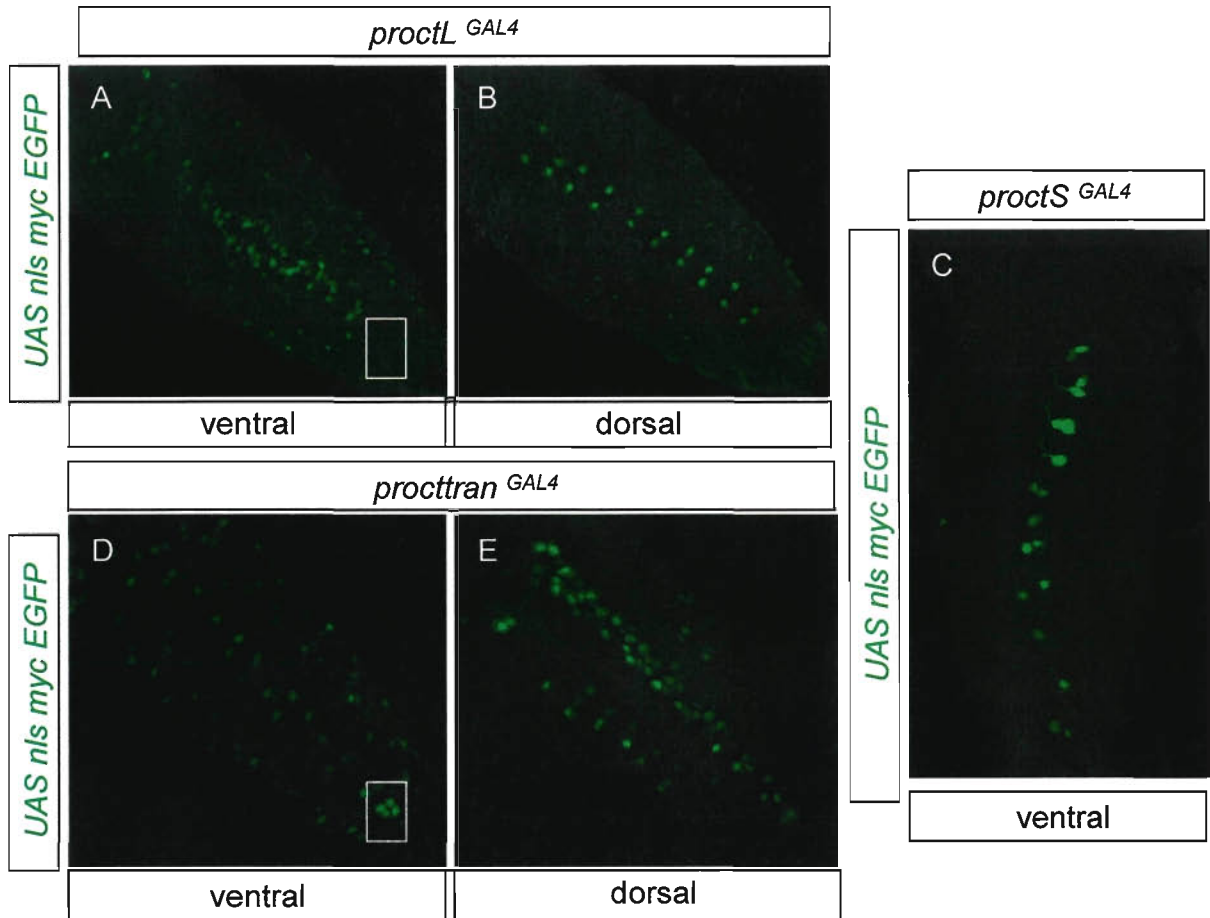
Figure F.3. Expression pattern of *proctolin* reporters in the VNC of the 3rd instar larvae of *Drosophila.*
Confocoal z-series of the ventral and dorsal sides of the *proctL^{GAL4}* (A and B), *proctS ^{GAL4}* (C) and *procttran^{GAL4}* (D and E) reporter expression in the dorsal and ventral regions of the larval CNS. *UAS-nls-myc-EGFP* is driven by the above reporters for detection of expression. No dorsal image was included for *proctS ^{GAL4}* because no *UAS-nls-myc-EGFP* expression was detected. *UAS-nls-myc-EGFP* expression was detected in the Pc neurons (white box) when driven by *procttran^{GAL4}* (D) but not by *proctL^{GAL4}* (A).
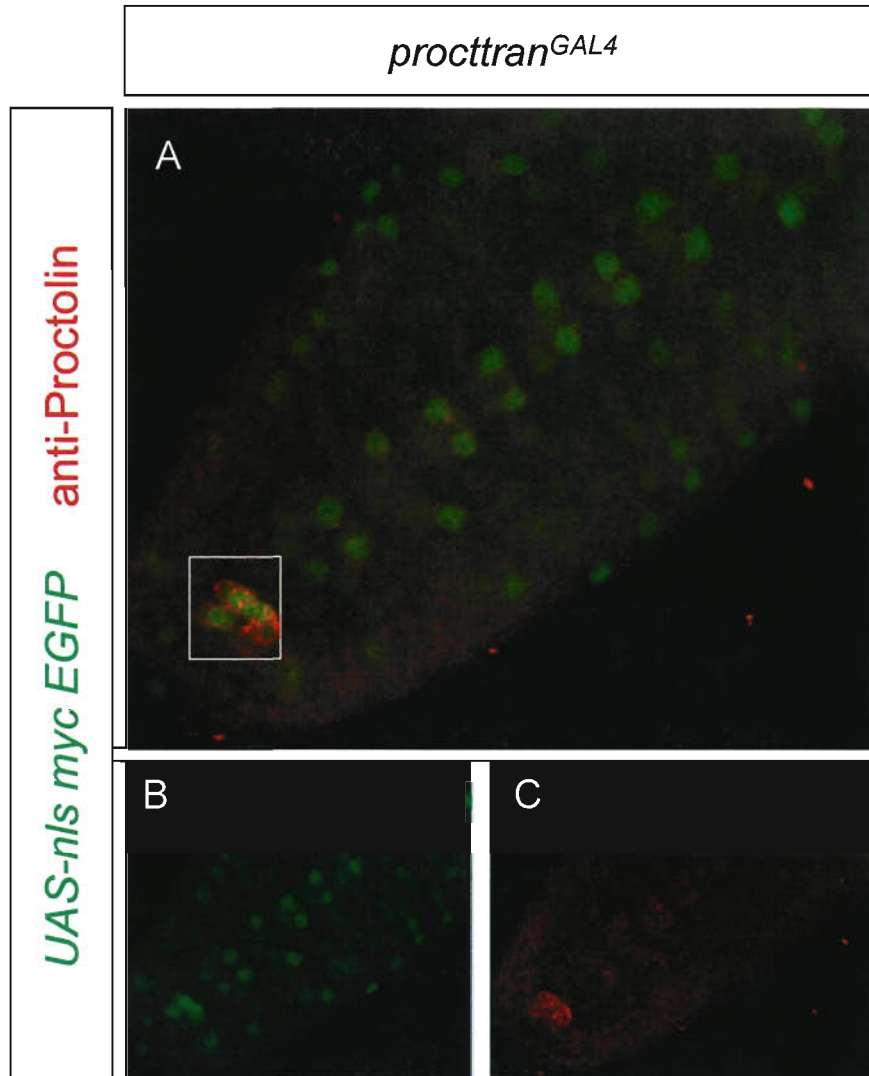
Figure F.4. The *procttran*^{GAL4} reporter produced similar expression pattern as Proctolin at first instar larval stage L1. (A) Merged confocal z-series of *proctL*^{GAL4}, *UAS-nls-myc-EGFP* expression pattern (B) and immunostaining of Proctolin (C) shows overlap along the midline clusters of neurons and the Pc neurons (A, white box).
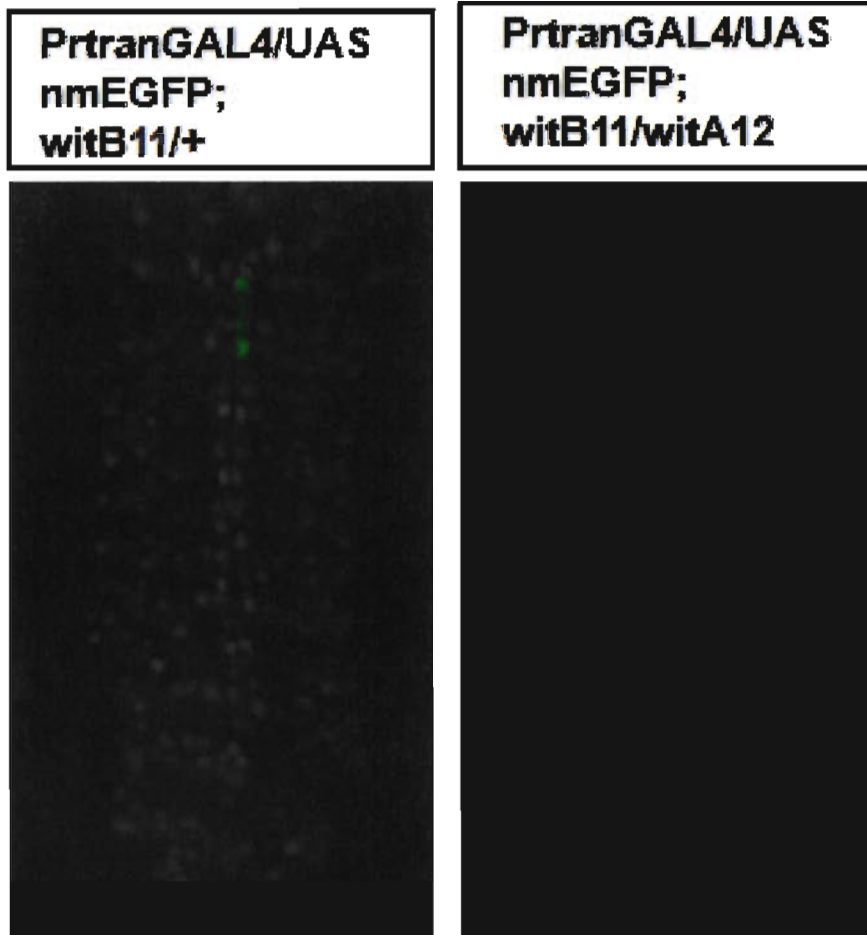
Figure F.5. Prtran GAL4 is regulated by BMP signalling. Stacked projection of larval L1CNS showing UAS nmEGFP driven by Prtran GAL4 in wildtype (witB11/+) and *wit* (witB11/witA12) background. Prtran is a ~1 kb translational fusion with GAL4 taken from just in front of the *proctolin* ATG codon.
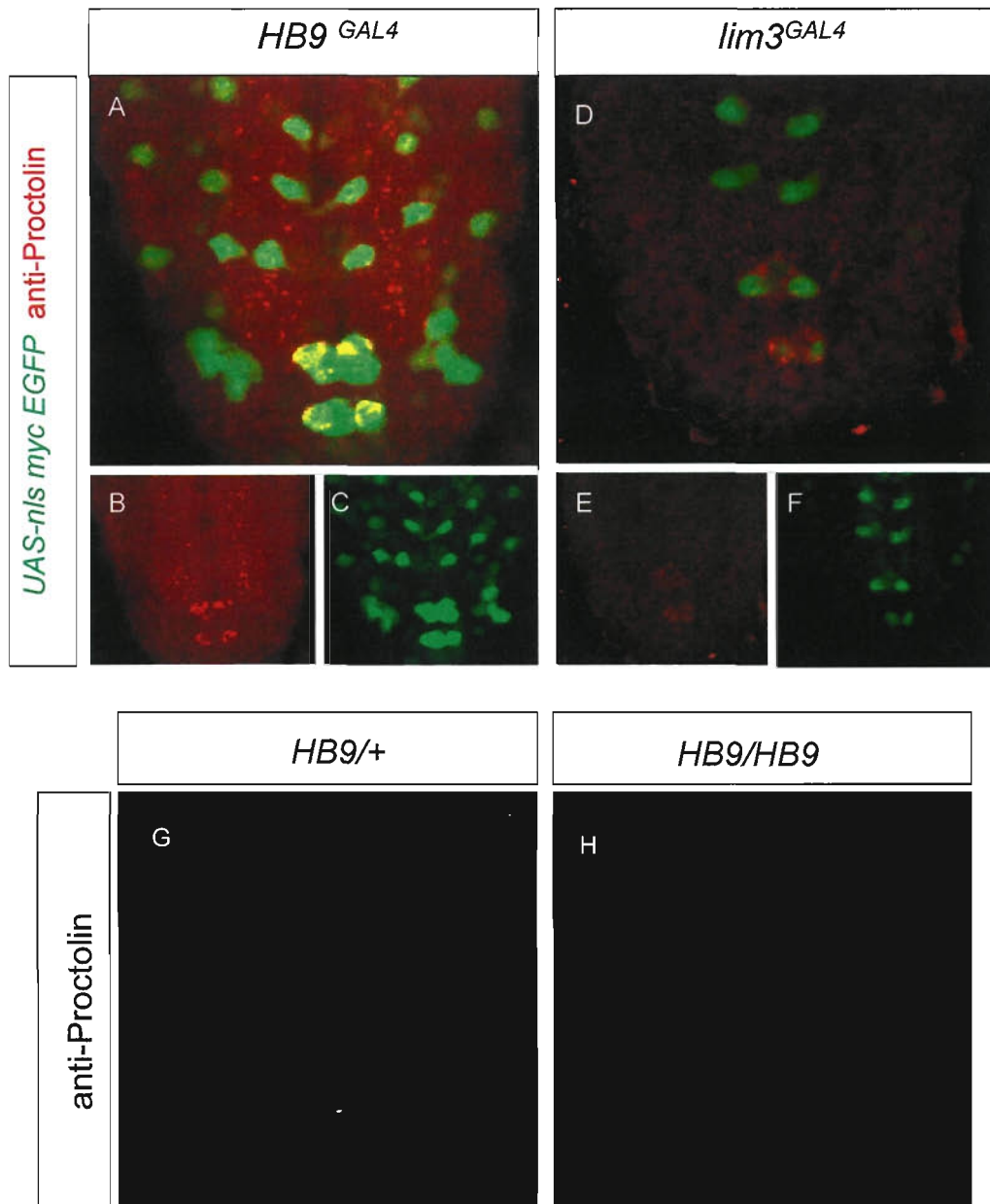
Figure F.6. *HB9* and *lim3* are two homeodomain transcription factors expressed in the *proctolin*-expressing Pc neurons. All VNC were dissected from 1st instar larvae (A) Merged z-stack image of anti-Proctolin immunostaining (B) and *HB9^{GAL4} UAS nls myc EGFP* expression (C). (D). Merged z-stack image of anti-Proctolin immunostaining (E) and *lim3^{GAL4} UAS nls myc EGFP* expression (F). Anti-Proctolin immunostaining for *HB9/+* (G) and *HB9/HB9* (H) showing that HB9 regulates *proctolin*.
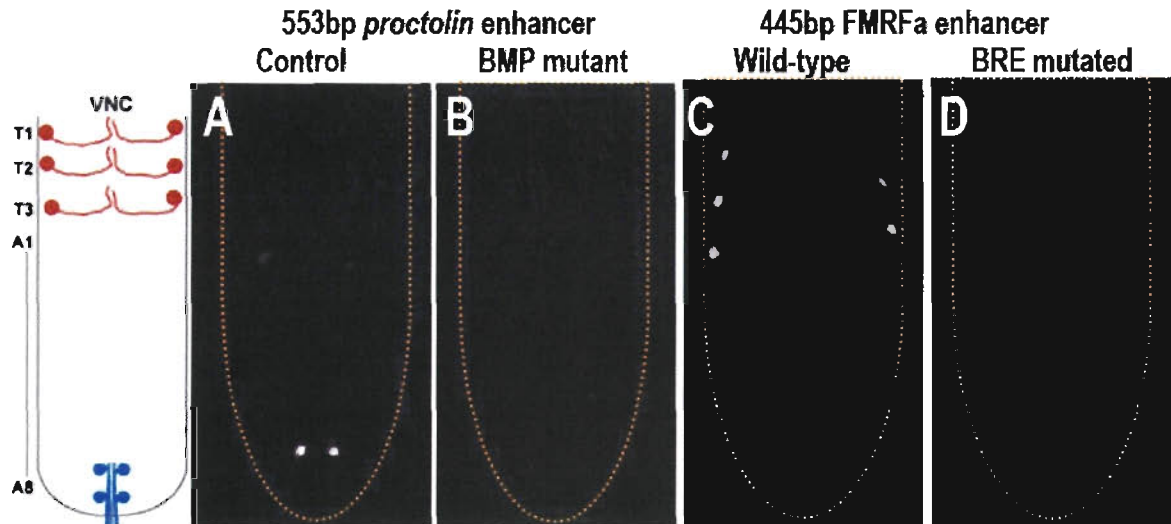
Figure F.7. The 553 bp *proctolin* and 445 bp *FMRFa* enhancer are both BMP-responsive. Left panel depicts a diagram of Tv (red) and Pc (neurons) in the *Drosophila* central nervous system. Visualization of *enhancer GAL4 > reporter* nerve cords from first instar larvae (A-D). Compare A and B for 553 bp *proctolin* reporter expression in wildtype versus BMP mutant. Mutation of Mad binding sequences in the 445bp *FMRAa* reporter result in loss of reporter activity (compare C and D. BRE stands for BMP response elements and are Mad binding sites).

Figure F.8. Cluster of homeodomain (purple) and Mad (yellow) and Medea (green) sequences in the 553 bp *proctolin* and 445 bp *FMRFa* enhancers. Note that shown sequences are not necessarily identical to the consensus ones.
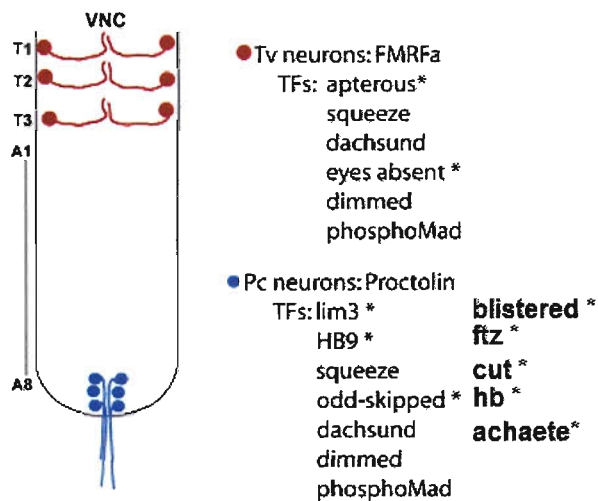
Figure F.9. Expression pattern of *FMRFa* and *Proctolin* in BMP-responsive cells in the ventral nerve cord (VNC) of the *Drosophila melanogaster*. Tv (red), and Pc (blue) neurons. TFs (TFs) in each cell type is listed. Asterisks denote TFs that are uniquely expressed in the Thoracic (Tv) or Posterior cluster (Pc) neurons.

Appendix G- dilp7

-1040   +660   GAL4   Dilp7 GAL4

-1040 AGCAACATAA ATGCGTATAT ATAAATGTAT TCTACTGCAT TCCTAGCACA GTTGTGGAAT TGAACAATAT TTGAATTTAG AAAGAAAACA ACCGCGAACA
TCGTTGTATT TACGCATATA TATTTACATA AGATGACGTA AGGATCGTGT CAACACCTTA ACTTGTTATA AACTTAAATC TTTCTTTTGT TGGCGCTTGT
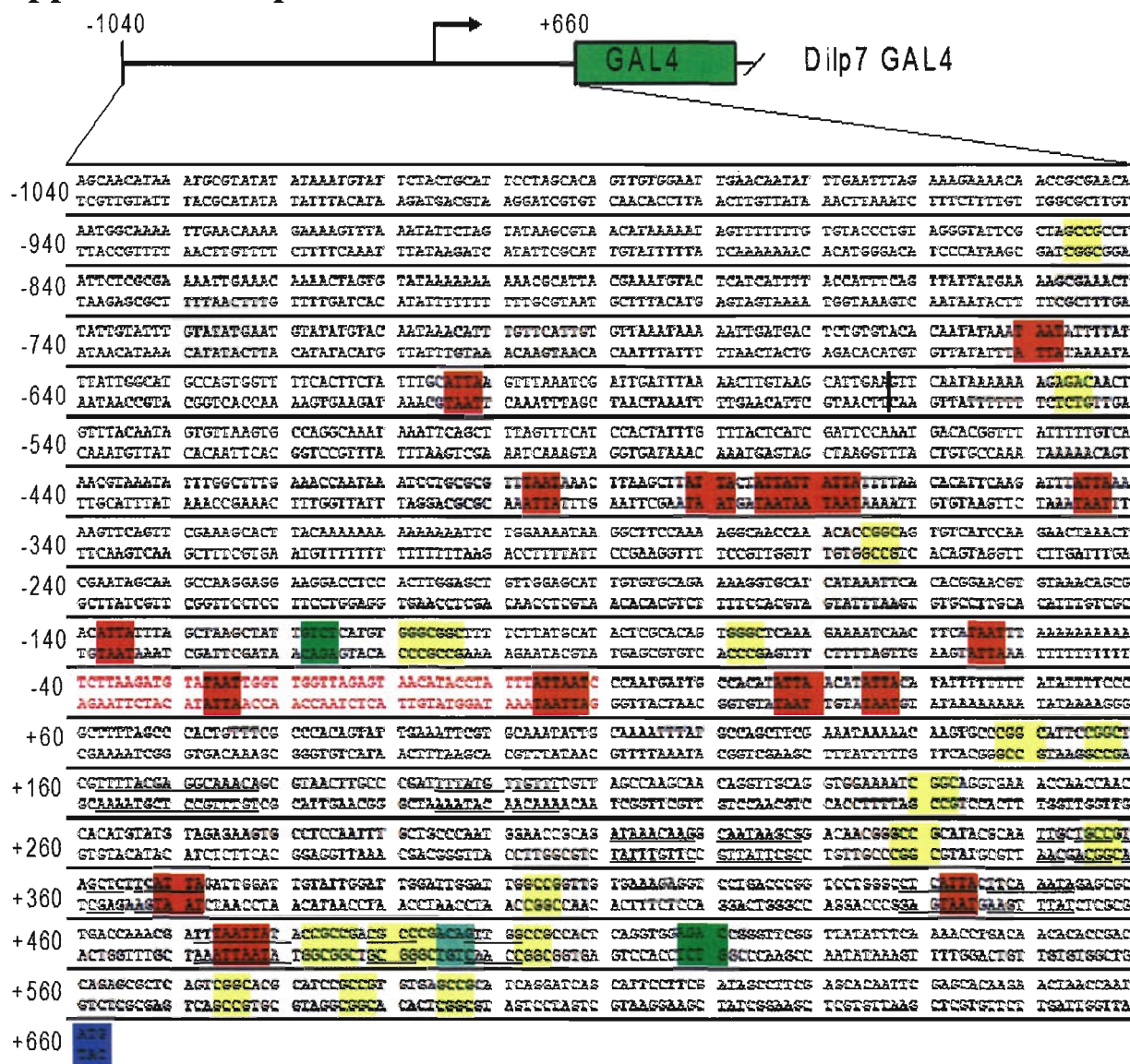
-940 AATGGCAAAA TTGAACAAAA GAAAAGTTTA AATATTCTAG TATAAGCGTA ACATAAAAAT AGTTTTTTTG TGTACCCTGT AGGGTATTCG CTAGCCGCCT
TTACCGTTTT AACTTGTTTT CTTTTCAAAT TTATAAGATC ATATTCGCAT TGTATTTTTA TCAAAAAAAC ACATGGGACA TCCCATAAGC GATCGGCGGA

-840 ATTCTCGCGA AAATTGAAAC AAAACTAGTG TATAAAAAAA AAACGCATTA CGAAATGTAC TCATCATTTT ACCATTTCAG TTATTATGAA AAGCGAAACT
TAAGAGCGCT TTTAACTTTG TTTTGATCAC ATATTTTTTT TTTGCGTAAT GCTTTACATG AGTAGTAAAA TGGTAAAGTC AATAATACTT TTCGCTTTGA

-740 TATTGTATTT GTATATGAAT GTATATGTAC AATAAACATT TGTTCATTGT GTTAAATAAA AATTGATGAC TCTGTGTACA CAATATAAAT AATATTTTAT
ATAACATAAA CATATACTTA CATATACATG TTATTTGTAA ACAAGTAACA CAATTTATTT TTAACTACTG AGACACATGT GTTATATTTA TTATAAAATA

-640 TTATTGGCAT GCCAGTGGTT TTCACTTCTA TTTGCATTAA GTTAAATCG ATTGATTTAA AACTTGTAAG CATTGAAGTT CAATAAAAAA AGAGACAACT
AATAACCGTA CGGTCACCAA AAGTGAAGAT AAACGTAATT CAAATTTAGC TAACTAAATT TTGAACATTC GTAACTATTAA GTTATTTTTT TCTCTGTTGA

-540 GTTTACAATA GTGTTAAGTG CCAGGCAAAT AAATTCAGCT TTAGTTTCAT CCACTATTTG TTTACTCATC GATTCCAAAT GACACGGTTT AITTTTGTCA
CAAATGTTAT CACAATTCAC GGTCCGTTTA TTTAAGTCGA AATCAAAGTA GGTGATAAAC AAATGAGTAG CTAAGGTTTA CTGTGCCAAA TAAAAACAGT

-440 AACGTAAATA TTTGGCTTTG AAACCAATAA ATCCTGCGCG TTAAAAAC ITAAGCTTAT TACTATTATT ATTATTTAA CACATTCAAG ATTATTTAAA
TTGCATTTAT AAACCGAAAC TTTGGTTATT TAGGACGCGC AATTATTTTG AATTCGAATA ATGATAATAA TAATAAAATT GTGTAAGTTC TAATAATTT

-340 AAGTTCAGTT CGAAAGCACT TACAAAAAAA AAAAAAATTC TGGAAAATAA GGCTTCCAAA AGGCAACCAA ACACCGGCAG TGTCATCCAA GAACTAAACT
TTCAAGTCAA GCTTTCGTGA ATGTTTTTTT TTTTTTTAAG ACCTTTTATT CCGAAGGTTT TCCGTTGGTT TGTGGCCGTC ACAGTAGGTT CTTGATTTGA

-240 CGAATAGCAA GCCAAGGAGG AAGGACCTCC ACTTGGAGCT GTTGGAGCAT TGTGTGCAGA AAAGGTGCAT CATAAATTCA CACGGAACGT GTAAACAGCG
GCTTATCGTT CGGTTCCTCC TTCCTGGAGG TGAACCTCGA CAACCTCGTA ACACACGTCT TTTCCACGTA GTATTTAAGT GTGCCTTGCA CATTTGTCGC

-140 ACATTATTTA GCTAAGCTAT TTTCTCATGT GGGCGGCTTT TCTTATGCAT ACTCGCACAG TGGGCTCAAA GAAAATCAAC TTCATAATTT AAAAAAAAAA
TGTTAATAAT CGATTCGATA ACAGAGTACA CCCGCCGAAA AGAATACGTA TGAGCGTGTC ACCCGAGTTT CTTTTAGTTG AAGTATTAAA TTTTTTTTTT

-40 TCTTAAGATG TATAATTGGT TGGTTAGAGT AACATACCTA TTTATTAATC CCAATGATTG CCACATATTA ACATATTACA TATTTTTTTT ATATTTCCC
AGAATTCTAC ATATTAACCA ACCAATCTCA TTGTATGGAT AAATAATTGA GGTTACTAAC GGTGTATAAT TGTATAATGT ATAAAAAAA TATAAAAGGG

+60 GCTTTTAGCC CACTGTTTCG CCCACAGTAT TGAAATTCGT GCAAATATTG CAAAATTTAT GCCAGCTTCG AAATAAAAAC AAGTGCCCGG CATTCCGGCT
CGAAAATCGG GTGACAAAGC GGGTGTCATA ACTTTAAGCA CGTTTATAAC GTTTTAAATA CGGTCGAAGC TTTATTTTTG TTCACGGCGT GTAAGGCCGA

+160 CGTTTTACGA GGCAAACAGC GTAACTTGCC CCATTTTATG TTGTTTGTT AGCCAAGCAA CAGGTTGCAG GTGGAAAATC GGCAGGTGAA ACCAACCAAC
GCAAAATGCT CCGTTGTCG CATTGAACGG GGTAAAAATAC AACAAAACAA TCGGTTCGTT GTCCAACGTC CACCTTTTAG CCGTCCACTT TGGTTGGTTG

+260 CACATGTATG TAGAGAAGTG CCTCCAATTT GCTGCCCAAT GGAACCGCAG ATAAACAAGG CAATAAGCGG ACAACGGGCC GCATACGCAA TTGCTGCCGT
GTGTACATAC ATCTCTTCAC GGAGGTTAAA CGACGGGTTA CCTTGGCGTC TATTTGTTCC GTTATTCGCC TGTTGCCCGG CGTATGCGTT AACGACGGCA

+360 AGCTCTTCAT TAGATTGGAT TGTATTGGAT TGGATTGGAT TGGCCGGTTG TGAAAGAGGT CCTGACCCGG TCCTGGGCGT CATTATTCA AATAGAGCGC
TCGAGAAGTA ATCTAACCTA ACATAACCTA ACCTAACCTA ACCGGCCAAC ACTTTCTCCA GGACTGGGCC AGGACCCGCA GTAATAAGT TTATCTCGCG

+460 TGACCAAACG ATTAATTAT ACCGCCGACG CCCGACGATT GGCCGCCACT CAGGTGGTGA CGGGTTCGG TTATATTTCA AAACCTGACA ACACACCGAC
ACTGGTTTGC TAATTAATA TGGCGGCTGC GGGCTGCTAA CCGGCGGTGA GTCCACCACT GCCCAAGCC AATATAAAGT TTTGGACTGT TGTGTGGCTG

+560 CAGAGCGCTC AGTCGGCACG CATCCGCCGT GTGAGCCGCA TCAGGATCAG CATTCCTTCG ATAGCCTTCG AGCACAATTC GAGCACAAGA ACTAACCAAT
GTCTCGCGAG TCAGCCGTGC GTAGGCGGCA CACTCGGCGT AGTCCTAGTC GTAAGGAAGC TATCGGAAGC TCGTGTTAAG CTCGTGTTCT TGATTGGTTA

+660 ATG
TAC

Figure G.1. Detailed analysis of the *dilp7* cis-regulatory region included in the *dilp7*^{GAL4} reporter. Evolutionarily conserved cluster of sequences across *Drosophila* species ((*D.melanogaster, D. erecta, D. persimilis, D. pseudoobscura, D.virilis, and D.grimshawi*) are underlined. The putative *dilp7* core promoter is in red font. Clusters of homeodomain consensus (highlight in red), Mad consensus or near consensus (highlight in yellow or purple)/Medea consensus or near consensus (highlight in green or light green) sequences was found. Start methionine of *GAL4(*and also *dilp7)* is highlighted in blue.
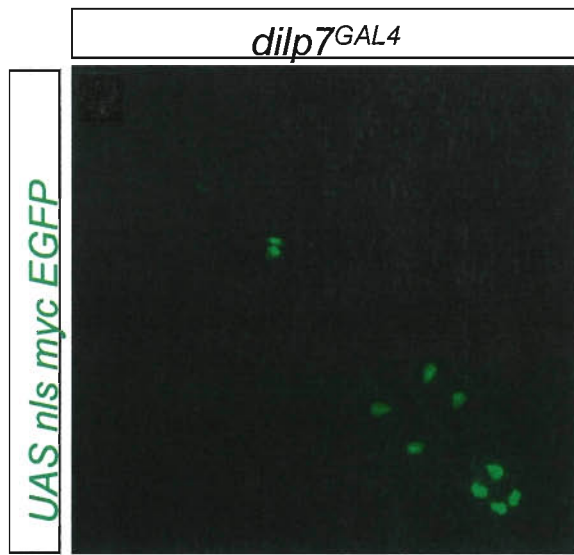
‘

Figure G.2. Expression of *dilp7-^{GAL4}* in 1st Instar larvae of *Drosophila*. Projected whole VNC confocal z-series of the 1st Instar larval VNC . B) *dilp7^{GAL4}; UAS nls.myc. EGFP*. Both are expressed exactly in a pattern consistent with previous reports.

Appendix H – Bioinformatics

Black capital letters represent bases in the *D.melanogaster* reference sequence that are conserved in the *D.simulans, D.sechellia, D.erecta, D.yakuba, D.virilis and D.grimshawi* orthologous DNAs

```
ccatctgcagacgtggttttcgaacgtatttatattgattatgggtgatcgtcaacaagagcagtggaca
cccaataaacctgtccaaaaacccgacacatttctgcccagtcatgcgtggtggacaatagccaaatgcc
attgatgagactcgtctccaaaaactTTgGCCtTTtgcccGGCCGTAATTACAGACTTccgtcttttgaac
agttttttcagccccacccaagagtcgagtcttgaaaagctggctgggatggggtggtttcgggtgctgG
AcGagaTGCcagAGGCGCCACAAtGTATCCtgttacagGTTACAGGgCCATAAAgcgcCATAAacgccGC
GACGgCAAtGgCAAATTATAaCGCATACggACAcGTAGtcgatccactggctagaaGGCTAATTGGACGt
gcccggccaggatgtccctgctcat
```

Black capital letters represent bases conserved in all species and colored bases represent sequences present in all species except *D.simulans, D.sechellia, D.erecta, D.yakuba, D.virilis or D.grimshawi*

```
ccatctgcagacgtggttttcgaacgtatttatattgattatgggtgatcgtcaacaagagcagtggaca
cccaataaacctgtccaaaaacccGACaCATTTCTGCccAGTCATGcgtggtggacaatagccaaatgcc
attgatgagactcgtctccaAAACTTTTGGCCTTTTGccgGGCCGTAATTACAGACTTCCGtCTTttgaac
agttttttcagccccacccaagagtcgagtcttgaaaagctggctggGATGGgGTGGtttcGGgtgctgG
AcGAGATGCCAgAGGCGCCACAAtGTATCCtgttacagGTTACAGGGCCATAAAgCgCCATAAAcgccGC
GACGGCAAtGgCAAATTATAaCGCATACGgACACGTAGtcgatccactggctagaAGGCTAATTGGACGT
GCcCGgCCAGGATGtccctgctcat
```

Black capital letters represent bases in the *D.melanogaster* reference sequence that are conserved in the *D.simulans, D.sechellia, D.erecta, D.yakuba, D.pseudoobscura, D.virilis and D.grimshawi* orthologous DNAs

```
ccatctgcagacgtggttttcgaacgtatttatattgattatgggtgatcgtcaacaagagcagtggaca
cccaataaacctgtccaaaaacccgacacatttctgcccagtcatgcgtggtggacaatagccaaatgcc
attgatgagactcgtctccaaaaactTTgGCCtTTtgccgGGCCgTAATTAcAGACTTccgtcttttgaac
agttttttcagccccacccaagagtcgagtcttgaaaagctggctgggatggggtggtttcgggtgctgg
acgagaTGCcagAGGCGCCACAAtGTATCCtgttacagGTTACAgGgCCATAAAgcgccATAAacgccGC
GACggcaatGgCAAATTATAaCGCATACggACAcGTAGtcgatccactggctagaaGGCTAATTGGACGt
gcccggccaggatgtccctgctca
```

Black capital letters represent bases conserved in all species and colored bases represent sequences present in all species except *D.simulans, D.sechellia, D.erecta, D.yakuba, D.pseudoobscura, D.virilis or D.grimshawi*

```
ccatctgcagacgtggttttcgaacgtatttatattgattatgggtgatcgtcaacaagagcagtggaca
cccaataaacctgtccaaaaacccgacacatttctgcccagtcatgcgtggtggacaatagccaaatgcc
attgatgagactcgtctccaAAACTTTGGCCTTTtgccgGGCCGTAATTACAGACTTCCGtCtTttgaac
agttttttcagccccacccaagagtcgagtcttgaaaagctggctgggatggggtggtttcgggtgctgG
AcGagaTGCCAgAGGCGCCACAAtGTATCCtgttacagGTTACAGGGCCATAAAgCgCCATAAAcgccGC
GACGgCAAtGgCAAATTATAaCGCATACGgACACGTAGtcgatccactggctagaaGGCTAATTGGACGT
GCccGgCcAGGatgtccctgctcat
```

Figure H.1. Evoprint of Tv enhancer showing results after filtering various *Drosophila* species.

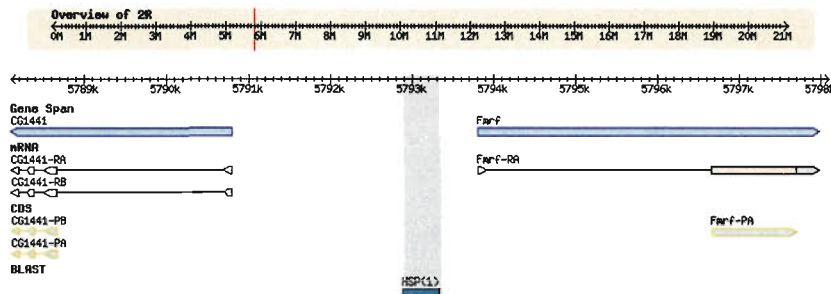**Explanation of genomic BLAT of Tv sequence to *Drosophila species***

Comparative Genomics: For each eBLAT (1-3) that aligned to the *D.melanogaster* TV enhancer piece, we performed both a BLAST search (in Flybase) and a BLAT search (in UCSC Browser) of that sequence on the pertinent species' genome. We then examined whether that sequence was near the FMRFa gene for that species. For every species, onlt the 1st BLAT was in the vicinity of the FMRFa for each genome, and in all cases was found 5' of the FMRFa gene. To account for the potential failure of annotation of additional FMRFa genes within the region of BLAT2 or 3 sequences, we performed TBLASTN 2.2.21 (NCBI: Search translated nucleotide databases using *D.melanogaster* FMRFa full length amino acid sequence). In every case, only one FMRFa gene was found in each genome and BLAT2/3 sequences were not found to be in the proximity of that FMRFa gene. These data, utilizing the sequenced genomes currently available for each *Drosophila* species, suggest that the Tv enhancer has not been subjected to rearrangement or duplication throughout the evolution of *Drosophila*, spanning *D.melanogaster* to *D.grimshawi*. These data further establish that the Tv enhancer is upstream of the FMRFa gene in all species.

[Stephen F. Altschul, Thomas L. Madden, Alejandro A. Schäffer, Jinghui Zhang, Zheng Zhang, Webb Miller, and David J. Lipman (1997), "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs", Nucleic Acids Res. 25:3389-3402].

In the following pages, we show the genomic location of the sequence that best matches the Tv enhancer.

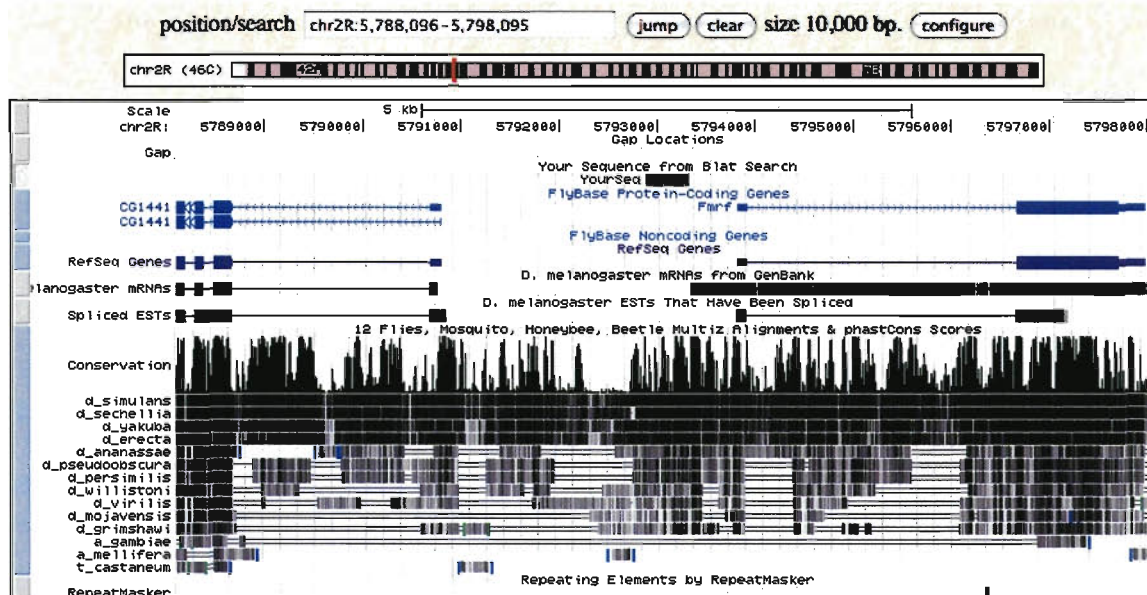## Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

Location of Tv piece relative to *D.melanogaster* FMRFa gene
(showing 2R:5788096..5798095)

Tv piece on UCSC Browser, showing Comparative Genomics, 2R: 5792874..5793318
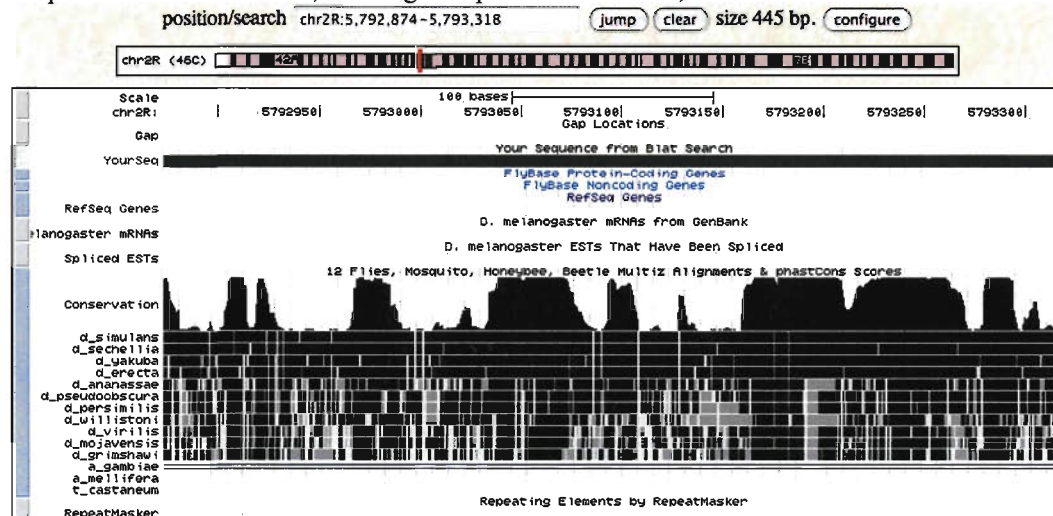
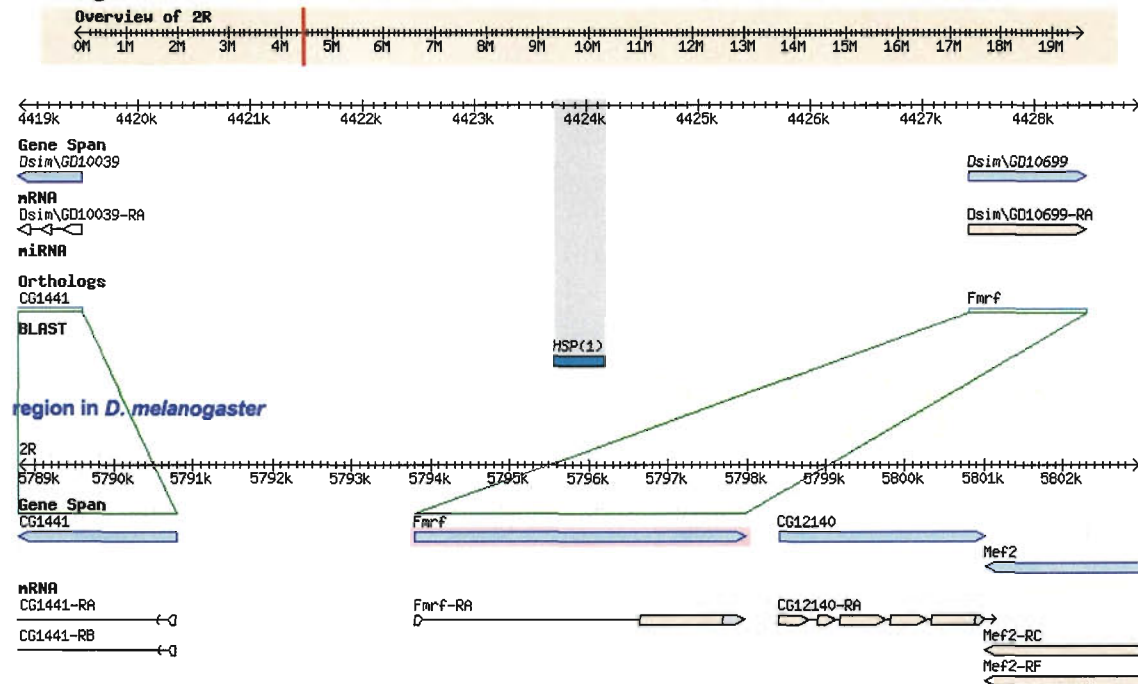Location of Tv piece relative to FMRFa gene
(showing 2R:5788096..5798095)

Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

D.simulans – GENE ID: 6733792 Dsim\GD10699 NCBI locus:XM_002080800.1
Blat 1 only one that maps to FMRFa
Release=r1.3 4423713.. 4424156
Showing 2R:4418935..4428934

Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

D.sechelia <u>GENE ID: 6608431 Dsec\GM21167; NCBI accession XM_002033129.1|</u>
– Blat 1 only one that maps to FMRFa
release=r1.3 Scaffold 1: 3403612.. 3404057
Showing scaffold_1:3398835..3408834
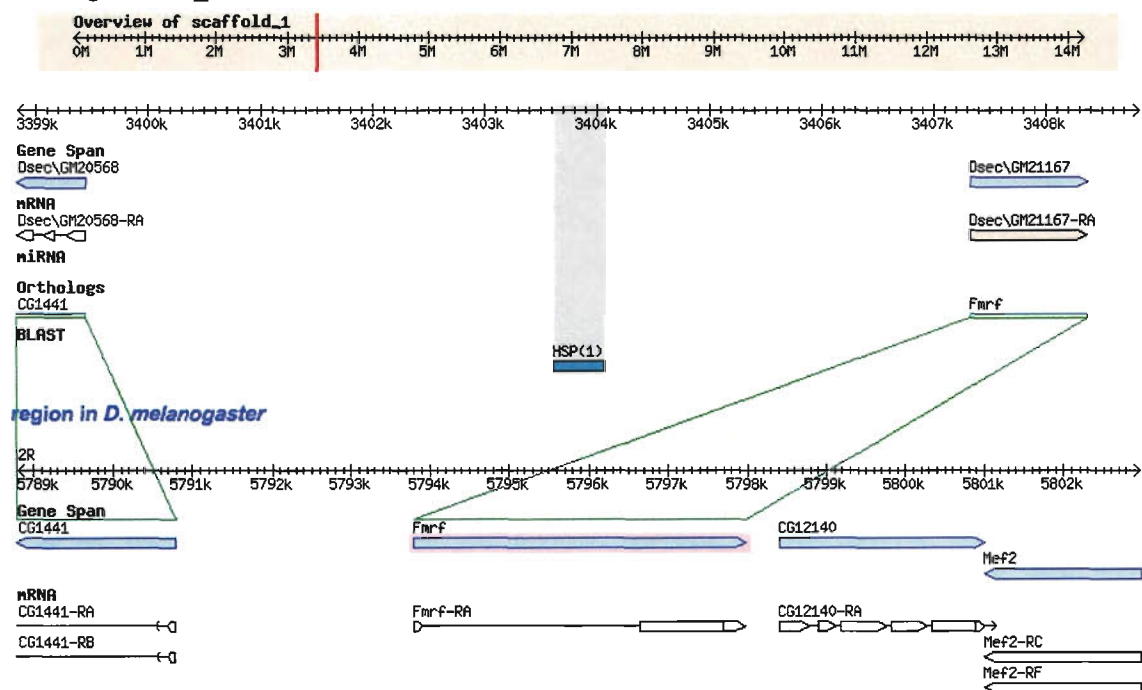
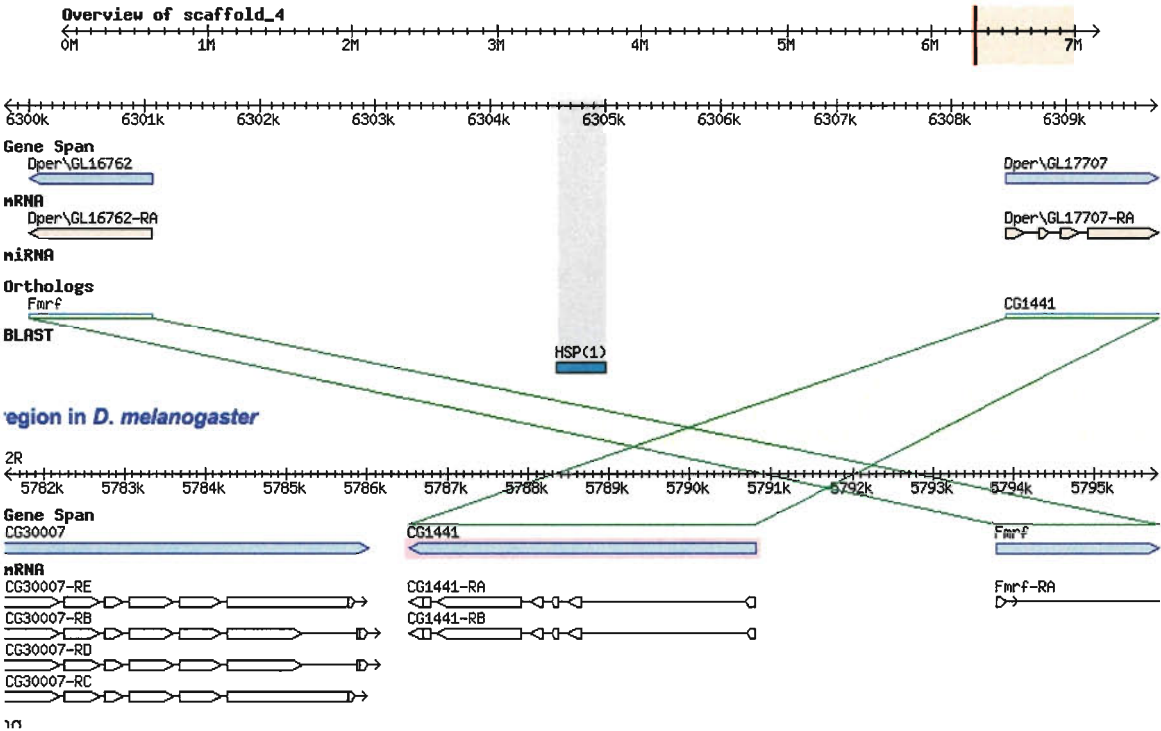## Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

D.erecta GENE ID: 6541511 Dere\GG24119; NCBI accession XM_001969091.1|
– Blat 1 only one that maps to FMRFa
release=r1.3; scaffold_4929: 8575920.. 8575478
showing scaffold_4929:8570699..8580698
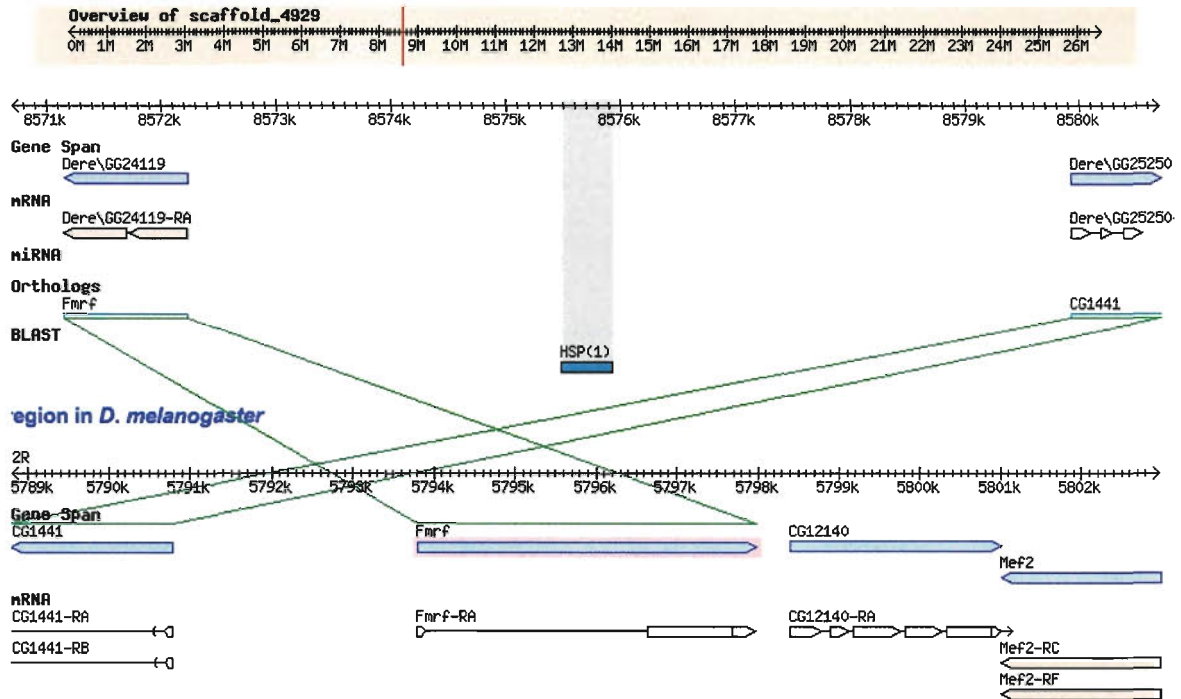
Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

D.yakuba GENE ID: 6528833 Dyak\GE19316, NCBI accession XM_002089829.1|
– Blat 1 only one that maps to FMRFa
release=r1.3: 2L:18418229.. 18418669   and
release=r1.3: 2L:18418701.. 18418870
showing 2L:18413449..18423448

Blast for 2L:18418229.. 18418669
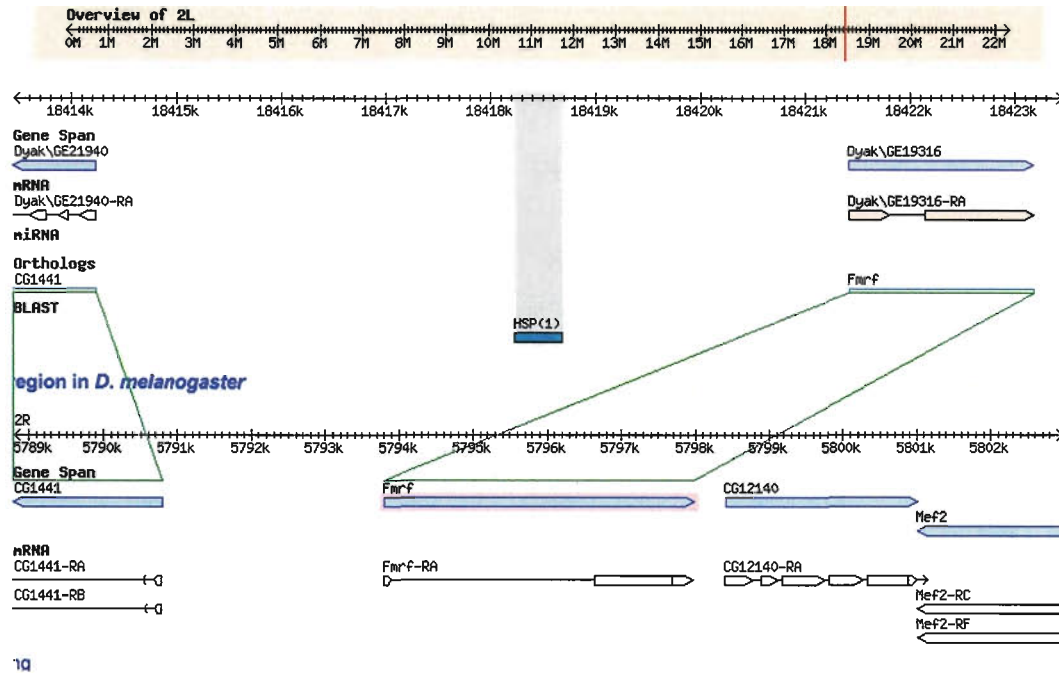
Blast for 2L:18418701.. 18418870
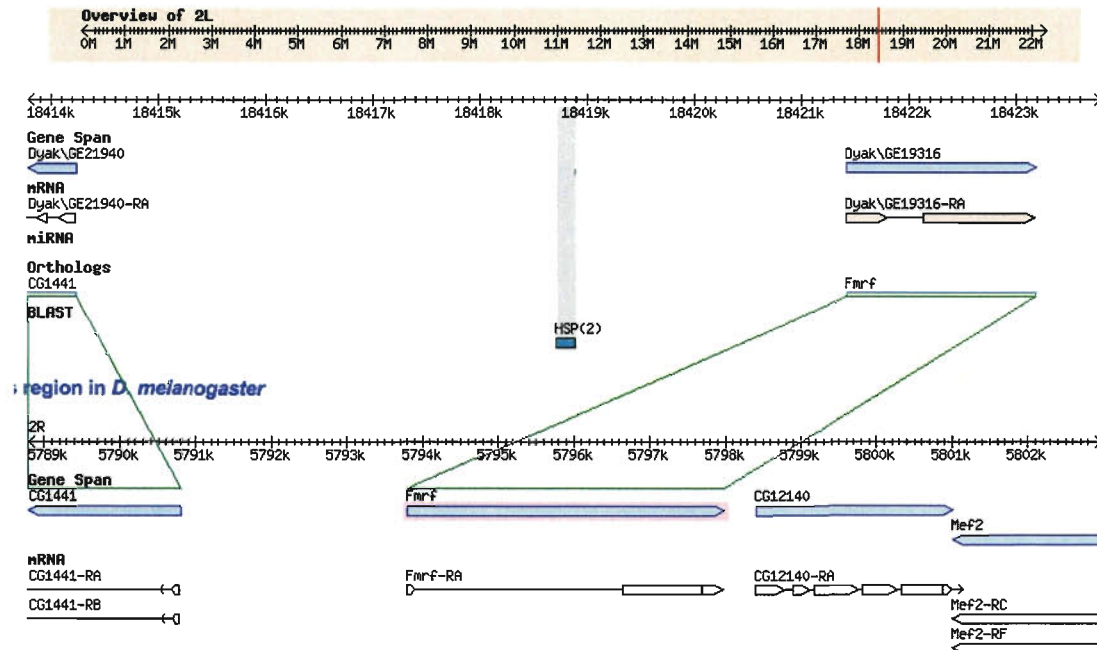
Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

D.persimilis GENE ID: 6592788 Dper\GL16762; NCBI accession XM_002018403.1|
– only Blat 1 hits FMRFa
release=r1.3; scaffold_4: 6304994..6304577
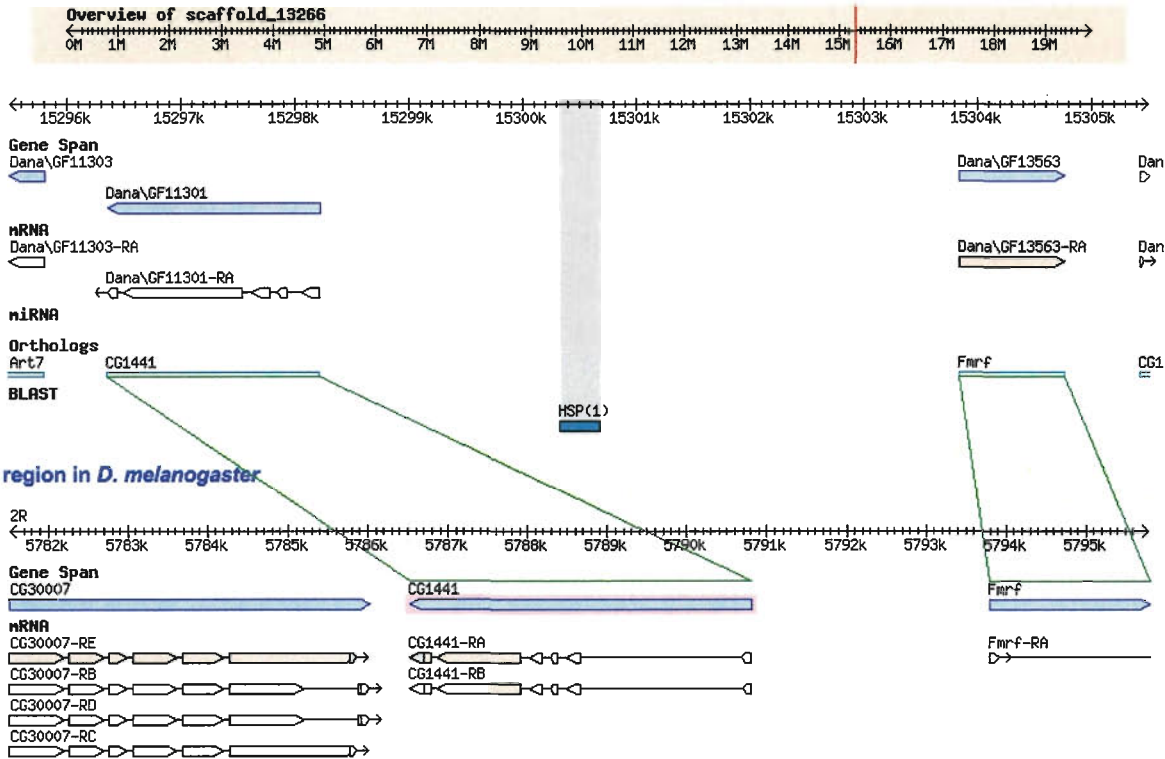showing scaffold_4:6299786..6309785

Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

D.ananassae. GENE ID: 6496402 Dana\GF13563 NCBI Accession XM_001960803.1|
Only Blat1 hits FMRFa
Release r1.3; scaffold_13266: 15300329.. 15300672
Showing scaffold_13266:15298329..15302672

Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

D.pseudoobscura, GENE ID: 4804717 Dpse\GA15356 NCBI accession: XM_001361198.2
Blat1 only one to hit FMRFa
release=r2.4: 3: 10983437.. 10983147
showing 3:10978292..10988291

Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

D.virilis (Dmel FMRFa TBLASTN hits two records in D.virilis – but both are the same gene)
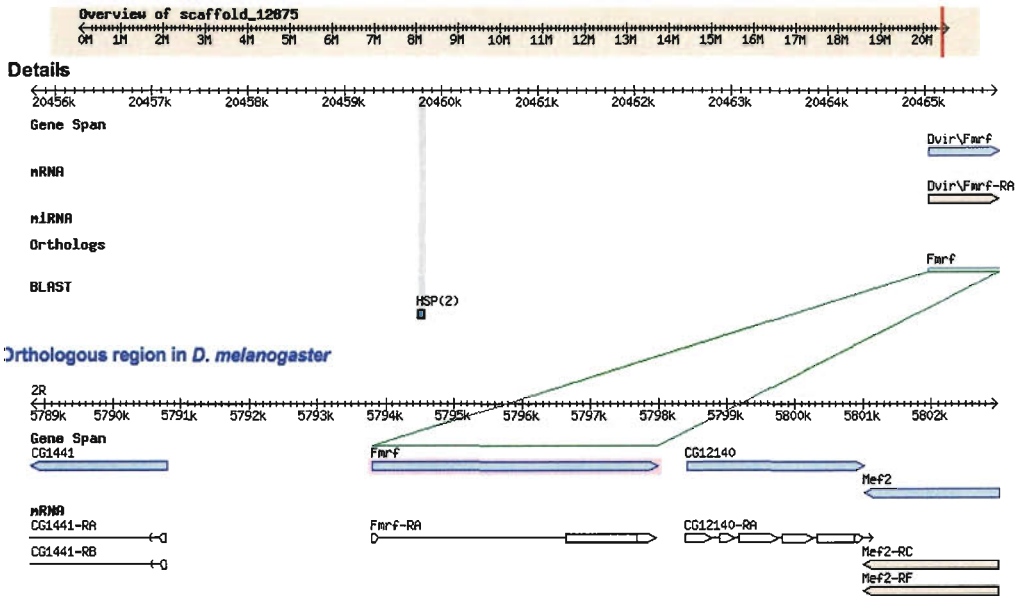GENE ID: 6625712 Dvir\Fmrf; NCBI accession XM_002051008.1|
chrom12875   +   20483908 20485734   1827
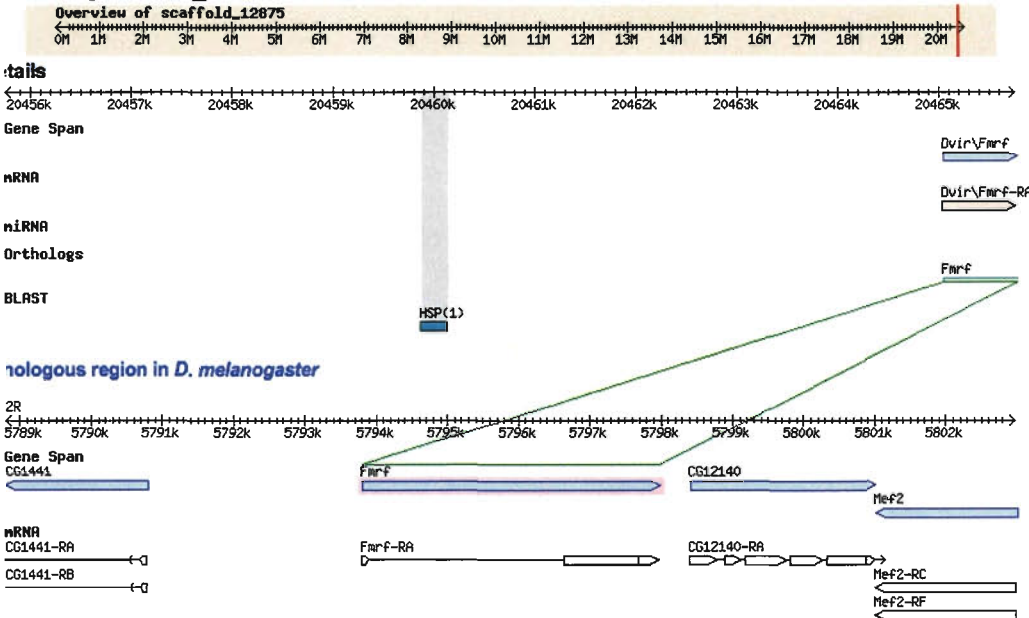chrom12875   +   20484455 20485495   1041
Release=r1.2; scaffold_12875 : 20459755.. 20459825
Release=r1.2; scaffold_12875 : 20459859..20460123

Showing scaffold_12875:20455751..20465750. NCBI accession
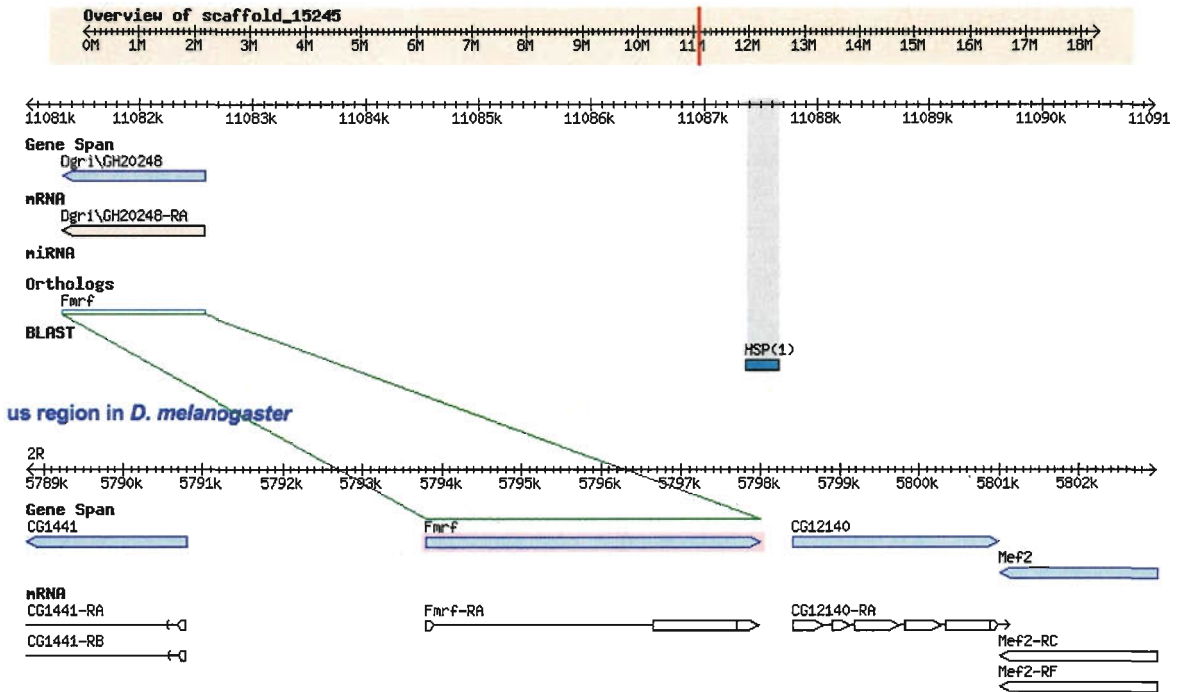

Showing scaffold_12875:20455751..20465750

Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

D.grimshawi GENE ID: 6560226 Dgri\GH20248 NCBI accession XM_001986904.1|
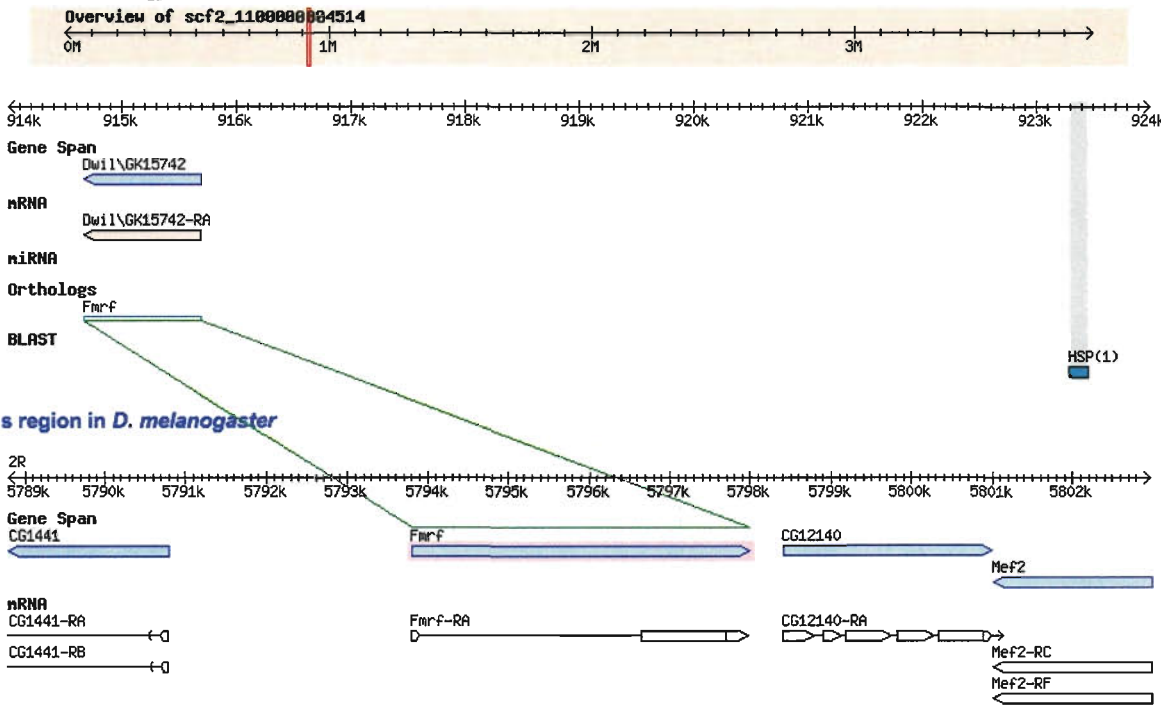**Release=r1.3; scaffold_15245:** 11087651.. 11087366
Showing scaffold_15245:11081000..11091000

Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

D.willistoni <u>ENE ID: 6640994 Dwil\GK15742, NCBI Accession XM_002063706.1|</u>
Release=r1.3; scf2_1100000004514: 923454.. 923290
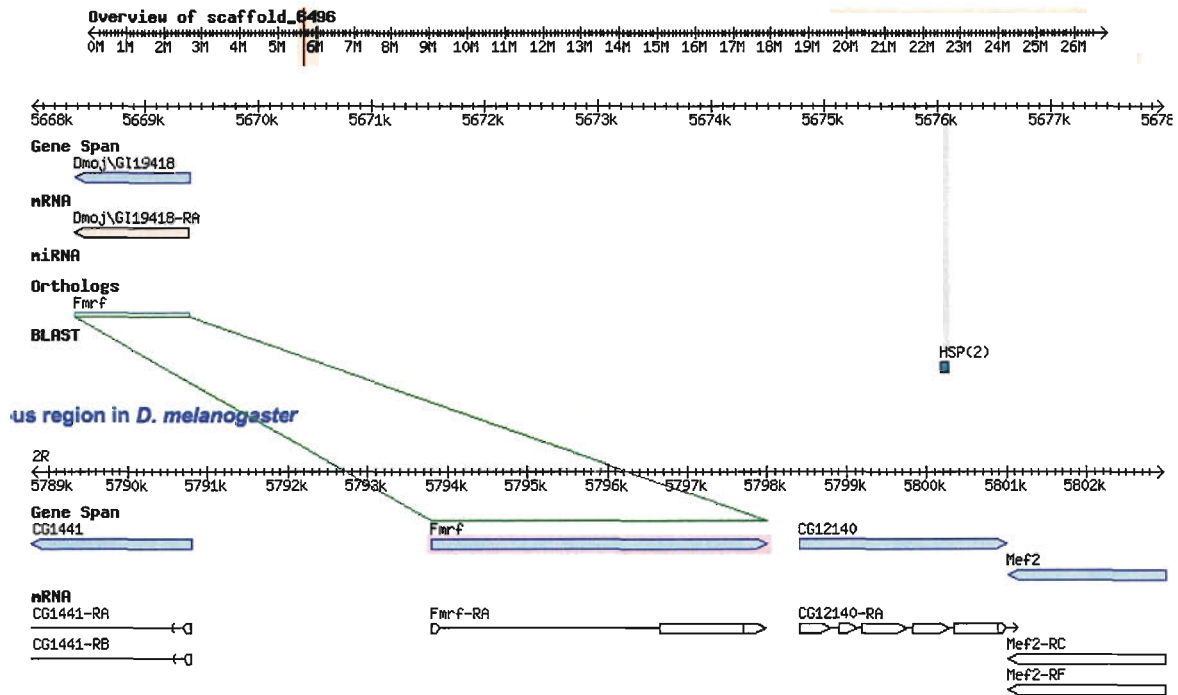
Showing scf2_1100000004514:914000..924000

Explanation of genomic BLAT of Tv sequence to *Drosophila species* (continued)

D.mojavensis GENE ID: 6578865 Dmoj\GI19418 NCBI accession XM_002004728.1|
BLAT 1 is the only one to hit FMRFa. Seq of BLAT1 BLASTs to two close seqs.

Release=r1.3: scaffold_6496: 5676089.. 5676021
Release 1.3: scaffold_6496: 5675955.. 5675789

Showing scaffold_6496:5668000..5678000