Characterising the function and evolution of enhancers in *Drosophila*

Alexandra D. Buffry

Degree awarded by

Oxford Brookes University

The thesis is submitted in partial fulfillment of the requirements of the award of Doctor of Philosophy

October 2019

Acknowledgments

First and foremost, I would like to thank everyone in the McGregor lab of whom I got the pleasure to work with during my PhD. Alistair thank you for always encouraging me, supporting me and listening to my stresses/ideas/excitement, also thank you for the countless beers which undoubtedly helped me to get through the last 4 years! Thank you to Dani who has not only provided life advice but many laughs in the process. To Jo, the 4 years spent at Brookes would definitely not be the same without you, thank you for being my best-friend, confidant and fellow seagull. To Brittany, girl we had a wild ride, thank you for always bringing a bit of fun into my life. Thank you to Anna without whom this final thesis would not exist, I have come to value your friendship enormously and think you are one of the most genuine, caring and downright awesome people I have ever met. There are too many other names to mention here but Sebastian and Franzi deserve a shout out for their endless support and advice.

Next thank you to all the friends I have met along the way, the Chepstow bears who have always provided tequila and the Oxford people who made my 4 years sunny and bright. Special mention to my lobster Kaytee, thank you for being the most supportive, wonderful and crazy friend. Thank you to Amy and Liisa – you guys really are 'the best friends anyone could have' I have loved every moment of our friendship.

I would like to say a huge thank you to Dom, I know the PhD and writing this thesis sometimes made me crazy but you were always there to make me laugh or provide a shoulder for me cry on. I really couldn't have done this without you. Thank you for all the meals, cuddles, chocolate bars and love.

Finally, I would like to thank my family. To my Dad, you are and always will be my role model and best-friend, thank you for being my biggest supporter. Even in the worst times of this PhD, hearing your voice was all I needed to realise everything would be fine. To Julie, thank you for everything you have done for me, I don't think I would have ever survived living by myself in Oxford if it wasn't for you. To my Mum, you have so much kindness in your heart and I hope I have inherited even just a little bit of that, thank you. To my Gramp, you are my hero, thank you for instilling in me a curious mind, I am sure living with you and Nanny and all the walks we went on fueled my love of the natural world. To my Nanny, I miss you every single day, you are a light in my life and I owe you so much. To all the other family members and friends that have supported me and encouraged me throughout this process, thank you.

Abstract

Enhancers are important modular cis-regulatory elements that precisely control the spatial and temporal expression of most genes. Therefore understanding how enhancers work and evolve is crucial for understanding the regulation of most developmental processes in animals. In my PhD project, I investigated three important aspects of enhancers to further our understanding of these cis-regulatory elements: enhancer identification, enhancer functionality and enhancer evolution. I identified a novel post-embryonic enhancer of the Hox gene *Ubx*, which provides new insights into how Hox genes are integrated into post-embryonic GRNs that determine fine-scale adult morphology. To help understand the cis-regulatory logic of enhancers, I focussed on a functionally related set of genes in a well characterised gene regulatory network. I identified a common pattern of motifs in candidate enhancers, which evidences that this approach may be a powerful for identifying key transcription factor binding sites in enhancers and de novo enhancer prediction. Finally, I described the evolutionary turnover of binding sites and intervening sequences among natural variants of the *Drosophila hb* P2 enhancer. This suggests that by studying the mutations gained through evolution we can learn more about flexibility and constraints in enhancers and improve on our current knowledge. I then generated tools for the detailed functional comparison of these natural variants, and observed qualitative differences in their function between species. These tools can now be applied to quantify these differences and shows this approach has great potential to better understand the function and evolution other developmental enhancers. Taken together my investigation has broadly contributed to our knowledge of enhancer organisation and functionality and provides a very useful platform for future analyses.

Table of contents

Chapter 1 ..... **17**

General introduction **17**

1.1 What are enhancers? **17**

1.2 The identification of enhancers **19**

1.2.1 Phylogenetic Footprinting 20

1.2.2 ChIP-Seq 20

1.2.3 ATAC-Seq 22

1.3 Enhancer functionality **23**

1.3.1 Enhancer syntax 23

1.3.1.1 Transcription Factor Binding Site Affinity 24

1.3.1.2 Shadow Enhancers 26

1.4 Enhancer Evolution **28**

1.4.1 Enhancer conservation 28

1.4.2 Divergent Enhancers 29

1.4.3 The Molecular Evolution of Enhancers 33

1.4.4 Cis-Regulatory Evolution and Phenotypic Diversification 35

1.4.4.1 The Genetic Basis of Phenotypic Evolution 35

1.4.4.2 Trichome Pattern Evolution 36

1.4.4.3 Enhancers Can Evolve in Different Ways to Facilitate Phenotypic Change 37

1.5 Aims of this thesis **39**

Chapter 2 ..... **42**

Identification and charaterisation of post-embryonic enhancers in Drosophila. **42**

2.1 Introduction 42

2.1.1 Post-embryonic roles of Hox genes 42

2.1.2 Trichome development as a model to study post-embryonic development and enhancers 43

2.1.3 Previous leg trichome enhancer predictions 46

2.2 Methods 48

2.2.1 Fly stocks and genetics 48

2.2.2 Cloning 49

2.2.3 GFP, protein trap and naked valley analysis 49

2.2.4 Analysis of potential TFs that may bind to Ubx VT42733 50

2.3 Results 51

2.3.1 Several regions of open chromatin at the Ubx locus drive expression in *Drosophila* pupal legs 51

2.3.2 Functional testing of putative enhancers 54

2.3.3 Further delimiting the Ubx naked valley enhancer 57

2.3.4 Ubx-1 is capable of generating trichomes in the NV 60

2.3.5 Analysis of larger fragments encompassing the *Ubx* enhancer suggest the presence of potential insulator/repressor sites 62

2.3.6 Functional testing 64

2.3.7 Summary of results from each driver line tested 66

2.3.8 Ubx protein trap 68

2.3.9 Transcription factor binding sites in the Ubx enhancer 69

2.3.10 Alignment and TFBS prediction of the NV enhancer 72

2.4 Discussion 75

2.4.1 Dissection of the NV enhancer region reveals a complex regulatory landscape 75

2.4.2 Larger reporter constructs revealed further complexity 76

2.4.3 The ATAC-seq profile provides insights into enhancer organisation at this locus 77

2.4.4 Ectopic expression? 78

2.4.5 Identification of TFBS in the NV enhancer 80

Chapter 3 81

Analysis of the function and regulation of leg trichome genes 81

3.1 Introduction 81

3.2 Methods 83

3.2.1 Gene identification 83

3.2.2 Accessible chromatin analysis 84

3.2.3 Motif enrichment using i-*cis*Target 84

3.2.4 RNAi screen and trichome analysis 85

3.3 Results 86

3.3.1 Selection of twelve genes expressed in T2 legs, which are putative targets of miR-92a and predicted to be dependent on Svb 86

3.3.2 RNAi against potential trichome genes 88

3.3.2.1 CG14395 RNAi 90

3.3.2.2 CG4678 RNAi 91

3.3.2.3 CrebA RNAi 92

3.3.2.4 forked RNAi 93

3.3.2.5 mwh RNAi 94

3.3.2.6 neo RNAi 95

3.3.2.7 Rab23 RNAi 97

3.3.3 Motif enrichment 98

3.3.4 Enrichment of Svb/Ovo and/or Grh motifs in ATAC-seq peaks 102

3.3.4.1 CG4678 102

3.3.4.2 CG5742 103

3.3.4.3 CG8303 103

3.3.4.4 CG14395 105

3.3.4.5 CrebA 106

3.3.4.6 forked 107

3.3.4.7 mwh *108*

3.3.4.8 neo *111*

3.3.4.9 Rab23 ..... *112*

3.3.4.10 sha *113*

3.3.4.11 svb *114*

3.3.4.12 tal *117*

3.4 Discussion 118

3.4.1 A set of 12 genes that may form a finely-tuned developmental switch for trichomes..... 118

3.4.2 RNAi analysis reveals several of the candidate genes affect the growth of trichomes 119

3.4.3 Overall motif enrichment reveals different motif combinations in open vs closed chromatin 122

3.4.4 Prediction of epidermal trichome enhancers..... 123

Chapter 4..... 125

Exploring the evolution of hunchback enhancers 125

4.1 Introduction 125

4.2 Methods..... 129

4.2.1 Sequence alignment 129

4.2.2 Fly stocks 130

4.2.3 Molecular cloning..... 130

4.2.4 MS2 system..... 132

4.2.5 Sample preparation and live-imaging..... 132

4.3 Results..... 133

4.3.1 Analysis of Zelda sites in the P2 enhancer 137

4.3.2 Testing the functional equivalence of P2 enhancers..... 139

4.3.3 Development and optimisation of tools to study enhancer turnover in P2 140

4.3.4 *D.melanogaster hb* P2 enhancer and promoter 141

4.3.5 *D.yakuba hb* P2 enhancer and promoter 144

4.3.6 *D. virilis hb* P2 enhancer and promoter 146

4.4 Discussion 149

5 General Discussion 152

5.1 Chapter two 152

5.1.1 Approaches to identify post-embryonic enhancers 152

5.1.2 Deciphering post-embryonic GRNs 153

5.2 Chapter three 155

5.2.1 Exploring a potential regulatory switch for trichome development 155

5.2.2 Investigating cis-regulatory logic 156

5.3 Chapter four 157

5.3.1 Turnover in enhancer sequences 157

5.3.2 Combining natural variation and live imaging to understand enhancer function 158

5.4 Conclusions 159

References 161

Supplementary materials 183

Table of figures

Figure 1. A basic overview of the location and function of enhancers in eukaryotic genomes p 18

Figure 2. The arrangement of enhancers at the *hunchback* locus in *Drosophila melanogaster* p 28

Figure 3. Conservation and divergence of the hunchback P2 enhancer in higher Dipterans p 30

Figure 4. The GRN for larval and leg trichomes p 46

Figure 5. The *Ubx* locus showing regions tested for a NV specific enhancer p 47

Figure 6. ATAC-Seq profile of the *Ubx* locus for T2 pupal legs at 21-28 hAPF p 53

Figure 7. VT42732 has no effect on NV size p 55

Figure 8. The driver line VT42733 generates trichomes in the NV p 56

Figure 9. VT42734 may drive weak expression in the NV p 57

Figure 10. Expression driven by sub-regions of VT42733 and VT42734 p 59

Figure 11. Ubx-1 is capable of generating trichomes in the NV p 60

Figure 12. Ubx2 has no effect on NV size p 61

Figure 13. Ubx-3 is also not capable of producing trichomes in the NV p 61

Figure 14. Expression driven by combinations of VT42732, VT42733 and VT42734 p 63

Figure 15. Driver line VT42732+VT42733 has no effect on NV size p 64

Figure 16. Combination line VT42733+Vt42734 significantly reduces the size of the NV p 65

Figure 17. The combination line VT42732+VT42733+VT42734 has no effect on NV size p 66

Figure 18. Summary of the position and chromatin profile of the driver lines tested in this study p 68

Figure 19. Expression of endogenously tagged Ubx in pupae at 24 hAPF p 69

Figure 20. Motifs in Ubx VT42733 multiple species alignment as found by MEME p 74

Figure 21. Process of identifying candidate genes for motif analysis p 86

Figure 22. Analysis of trichome length after RNAi knockdown of *CG14395* in T2 legs using the driver line VT057077/Dicer p 90

Figure 23. Analysis of trichome length after RNAi knockdown of *CG4678* in T2 legs using the driver line VT057077/Dicer p 91

Figure 24. Analysis of trichome length after RNAi knockdown of *CrebA* in T2 legs using heat-shock GAL4 driver line p92

Figure 25. Analysis of trichome length after RNAi knockdown of forked in T2 legs using the driver line VT057077/Dicer p 93

Figure 26. Analysis of trichome length after RNAi knockdown of *mwh* in T2 legs using the driver line VT057077/Dicer p 94

Figure 27. Effect on the morphology of trichomes both on the T2 leg and the wing after knockdown of *mwh* p 95

Figure 28. Analysis of trichome length after RNAi knockdown of *neo* in T2 legs using the driver line VT057077/Dicer p 96

Figure 29. Analysis of trichome length after RNAi knockdown of *Rab23* in T2 legs using the driver line VT057077/Dicer p 97

Figure 30. Top 5 motifs found enriched in Svb-dependent, miR-92a candidate genes p 99

Figure 31. ATAC-Seq profile for *CG4678* showing a region enriched for Svb and Grh motifs p 102

Figure 32. ATAC-Seq profile for *CG5742* in T2 pupal legs p 103

Figure 33. ATAC-Seq profile for *CG8303* in T2 pupal legs p 104

Figure 34. ATAC-Seq profile for *CG14395* in T2 pupal legs p 105

Figure 35. ATAC-Seq profile for *CrebA* in T2 pupal legs p 106

Figure 36. ATAC-Seq profile for *forked* in T2 pupal legs p 108

Figure 37. ATAC-Seq profile for the region containing *mwh* for T2 pupal legs p 110

Figure 38. ATAC-Seq profile for *neo* in T2 pupal legs p 111

Figure 39. ATAC-Seq profile for *Rab23* in T2 pupal legs p 112

Figure 40. ATAC-Seq profile for *sha* in T2 pupal legs p 114

Figure 41. ATAC-Seq profile for *svb* in T2 pupal legs p 116

Figure 42. ATAC-Seq profile for *tal* analysis region in T2 pupal legs p 118

Figure 43. Experimental design for both the rescue and the MS2 experiments p 131

Figure 44. Schematic showing multiple species alignment of the P2 enhancer of *hb* p 134

Figure 45. Analysis of the usage of canonical Zld sites in the P2 enhancer of *hb* p 138

Figure 46. Schematic of the rescue experiment strategy p 140

Figure 47. The anterior of an embryo expression GFP detected by MS2 loops driven by *D. melanogaster* P2E+P2P p 142

Figure 48. Whole embryo expressing GFP detected by MS2 loops driven by the *D. melanogaster* P2E+P2P p 143

Figure 49. Whole embryo expressing GFP detected by MS2 loops driven by the *D. yakuba* P2E+P2P p 145

Figure 50. Whole embryos expressing GFP detected by MS2 loops driven by the *D. virilis* P2E+P2P p 147

Figure 51. The anterior of an embryo expressing GFP detected by MS2 loops driven by the *D. virilis* P2E+P2P p 148

Figure 52. The anterior of an embryo expressing GFP detected by MS2 loops driven by the *D. virilis* P2E+P2P p 148

Supplementary Figure 1. Expression driven by the line AB01 p 183

Supplementary Figure 2. Multiple species alignment of *Ubx* VT42733 for *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. mauritiana* and *D. yakuba* p 187

Supplementary Figure 3. Results from i-cisTarget for motif enrichment in epidermal genes that are not dependent on Svb p 206

Supplementary Figure 4. Multiple species alignment of the hb P2 enhancer for several Drosophila species p 207

List of tables

Table 1. Summary of results from functional testing for all constructs p67

Table 2. List of candidate TFs with binding sites in the VT42733 enhancer p71

Table 3. List of the 12 genes identified in Fig. 21 p87

Supplementary Table 1. The list of genes produced by TargetScan that are predicted to be targeted by miR-92a p188

Abbreviations

BLAST	basic local alignment search tool
bp	base pairs
°C	degrees Celsius
Cas9	caspase 9 endonuclease
CRISPR	clustered regularly interspaced palindromic repeats
DamID	DNA adenine methyltransferase identification
DNA	deoxyribonucleic acid
DNase-seq	DNase I hypersensitive sites sequencing
gDNA	genomic DNA
GFP	green fluorescent protein
kb	kilobases
mRNA	messenger ribonucleic acid
MYA	million years ago
PCR	polymerase chain reaction
Pol II	RNA-polymerase II
RFP	red fluorescent protein
RNA	ribonucleic acid
RNAi	RNA interference
STARR-seq	self-transcribing active regulatory region sequencing

Chapter 1
General introduction

1.1 What are enhancers?

The function of all organisms relies on the regulation of gene expression to produce precise amounts of proteins and other crucial factors at the right time and in the correct cells. While there are many mechanisms through which this is achieved (reviewed in Alonso and Wilkins, 2005), at the transcriptional level, the primary method of control is by enhancers - cis-regulatory DNA sequences that work with the core promoter to facilitate the transcription of genes (Banerji, Rusconi and Schaffner, 1981). Enhancers integrate information from the binding of transcription factors (TFs) and proteins that regulate chromatin accessibility to modulate the expression of target genes. A given gene can have multiple enhancer modules that are each responsible for the regulation of different aspects of its spatial and temporal expression (Fig. 1) (Dynan, 1989; Uchikawa et al., 2003; Carroll, 2008). Enhancer modules are often defined as elements of several hundred bases representing the minimum sequence required to drive expression (often of a reporter gene) that recapitulates an aspect of the native expression of the gene (Arnone and Davidson, 1997; Simpson and Ayyar, 2008). Although this is a useful definition and has greatly aided in the identification of many spatially and/or temporally specific enhancers, it belies the complexity of these cis-regulatory elements since they can overlap in position and expression and even share transcription factor binding sites (TFBS). A focus on the minimal sequence requirements to drive patterns resembling the native

gene may also miss important aspects of enhancer function and the regulation of the gene in question, including, for example, how different modules and primary and shadow enhancers may interact and combine to regulate gene expression in their native context (Fig.1) (reviewed in Halfon, 2019). Indeed, enhancers are notoriously difficult to identify in DNA because currently TF binding sites are degenerate and there is no consensus as to what genomic features mark enhancer regions (reviewed in Halfon, 2019). Therefore, although it is becoming more feasible to study the regulatory genome with new tools such as ATAC-seq, C technologies and CRISPR/Cas9, we still do not fully understand the regulatory logic underlying enhancer function. Given their importance in evolution, development, and disease it is imperative that we continue to study enhancer structure and function in more detail.

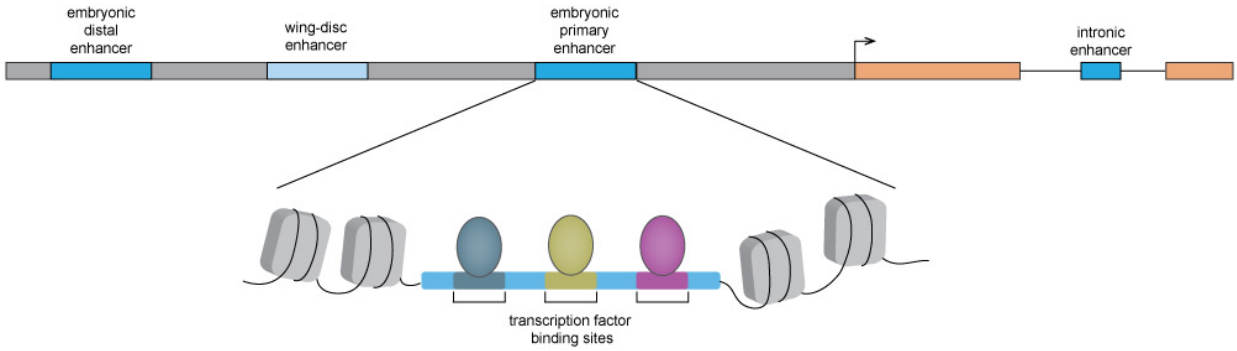


Fig. 1. A basic overview of the location and function of enhancers in eukaryotic genomes. The hypothetical gene in this case (shown in orange) is under control from several different enhancers. There is a proximal or primary embryonic enhancer, a distal or shadow embryonic enhancer, an enhancer which regulates expression of the gene in a different tissue, in this example, the wing disc, and finally an enhancer in the intron of the gene. The embryonic primary enhancer shows an example of the chromatin structure found at enhancers, the chromatin is accessible which allows the binding of transcription factors (TFs) to their specific binding sites (TFBS) to facilitate regulation of gene expression.

1.2 The identification of enhancers

Many enhancers have been identified using functional assays, such as “enhancer bashing”, where sequential DNA fragments are placed up-stream of a reporter gene to test if they can recapitulate aspects of the expression of a nearby gene of interest (Kvon, 2015). This approach has largely been successful in identifying the enhancers of individual genes and the principle has recently been applied to generate panels of reporter construct lines that cover the putative cis-regulatory regions of many genes in some organisms (Kvon, 2015). For example, the FlyLight and VT collections have produced a vast library of transgenic GAL4 reporter lines for use in *Drosophila* (Pfeiffer et al., 2008; Jenett et al., 2012). Similarly, for mammals, the VISTA Enhancer Browser catalogs human enhancers, many of which have been validated in transgenic mice (Pennacchio et al., 2006; Visel et al., 2007).

It is clear that such resources greatly advance the identification and functional validation of the enhancers of many genes, however, a reporter gene-based approach can suffer from the fact that enhancers can often be distant from the genes that they regulate with many intervening genes (e.g. Lettice et al., 2003). Indeed, enhancer bashing large regions of DNA can be time-consuming and low-throughput, and so ideally it is desirable to first identify putative enhancers using bioinformatic tools followed by functional validation using reporter genes and other approaches.

The development of bioinformatic tools to identify enhancer regions has been a challenging task since enhancer regions have no obvious sequence-level features like open reading frames to aid in their discovery (reviewed in Slattery and White, 2013). However, very powerful approaches are now available for the identification of

enhancer sequences (Suryamohan and Halfon, 2015; Kleftogiannis, Kalnis and Bajic, 2016; Hasma and Halfon, 2019).

1.2.1 Phylogenetic Footprinting

Phylogenetic footprinting is a method of identifying putative cis-regulatory elements in the genome based on their conservation among different species (Zhang and Gerstein, 2003; Alonso et al., 2009). It relies on the principal that functional noncoding elements are thought to be under more constraint and hence undergo a slower rate of change than nonfunctional elements (Ganley and Kobayashi, 2007). This general approach has been used in a vast number of studies and can be very powerful, particularly when combined with other tools to study enhancers. Despite the power that phylogenetic footprinting can provide it is limited by the fact that the similarity of sequences generally decreases with phylogenetic distance and is unlikely to be able to detect newly evolved and lineage-specific enhancers (Prabhakar et al., 2006; Hardison and Taylor, 2012; Maeso et al., 2013). In addition, it is thought that despite being evolutionary conserved, in some cases enhancer sequences evolve faster than coding regions (Clarke et al., 2012) which can hinder phylogentic footprinting efforts, and in some cases binding sites within enhancers can evolve more rapidly than flanking nucleotides (Wray, 2006).

1.2.2 ChIP-Seq

Chromatin immunoprecipitation coupled to sequencing (ChIP-Seq) enables genome-wide mapping of protein binding and epigenetic marks (Mundade et al., 2014). ChIP-

seq involves physical cross-linking of TFs to their binding sites *in vivo*, and the resulting complexes can then be isolated via immunoprecipitation with an antibody against the TF or histone mark. Subsequent sequencing then allows identification of the potential enhancers that have been bound by the TF or carry particular histone marks such as monomethylation of histone H3 at lysine 4 (H3k4me1) or acetylation of histone H3 at lysine 27 (H3K427ac) (Heintzman et al., 2007; Visel et al., 2009; Creyghton et al., 2010; Ernst et al., 2011; Rada-Iglesias et al., 2011; Nord et al., 2013; Zhu et al., 2013). ChIP-Seq can also be used to investigate how similar or different TF binding profiles are among species and how rapidly these patterns change. Thus, ChIP-Seq studies have contributed to our understanding of regulatory evolution (Villar, Flicek and Odom, 2014). For example, Prasad et al. (2016) investigated *Ultrabithorax* (*Ubx*) binding among insect species using ChIP-Seq and identified both species-specific and common targets of this transcription factor. Subsequently, using reporter assays, they showed that putative *Apis* enhancer of *vestigial* (*vg*) was able to drive expression in the third thoracic segment of *Drosophila*. However, the previously identified *Drosophila vg* enhancer mediated Ubx repression of this gene in this segment, implicating changes in the Ubx responsiveness of this enhancer in the development of halteres versus wings during the course of insect evolution (Prasad et al., 2016).

Studies that have used ChIP-Seq, like those above, demonstrate the utility of this approach to identify genome-wide histone marks and the regions bound by specific transcription factors with the caveat that comparisons among species will be limited to those for which a cross-reacting antibody is available.

1.2.3 ATAC-Seq

Although there are other methods to identify putative enhancers including ChIP-Seq as discussed above, as well as DamID, DNase-Seq and STARR-Seq (reviewed in Risca and Greenleaf, 2015; Shlyueva, Stampfel and Stark, 2014). Assay for Transposable Accessible Chromatin coupled to next-generation sequencing (ATAC-Seq) is a relatively new method of open chromatin analysis that has great potential for the study of genome-wide transcriptional regulation (Buenrostro et al., 2013).

The basic principle of ATAC-Seq relies on taking advantage of the specificity of Tn5 transposase to insert sequencing adapter oligo-nucleotides into regions of accessible chromatin in the genomic DNA of permeabilized, unfixed cells (Reviewed in Risca and Greenleaf, 2015). Thus, ATAC-Seq captures genome-wide sites of open chromatin and can provide information on TF occupancy, individual nucleosomes, and chromatin compaction at nucleotide resolution (Buenrostro et al., 2013).

In the last few years there has been a large number of studies that have utilized ATAC-Seq data for enhancer and chromatin accessibility analysis, for example in *Drosophila*, ATAC-Seq has been used in various tissues including, but not limited to, the embryo (Blythe and Wieschaus, 2016) the developing eye disc (Davie et al., 2015) and pupal legs (Kittelmann et al., 2018). Further recent developments have also been made in ATAC-Seq, for example, it is now possible to do ATAC-Seq on single cells (sci-ATAC-Seq) giving a comprehensive map of chromatin accessibility for each individual cell (Cusanovich et al., 2018).

ATAC-Seq provides advantages over other methods. For instance, formaldehyde-assisted isolation of regulatory elements (FAIRE)-Seq allows nucleosome-bound DNA to be distinguished from the open chromatin that can then

be sequenced to identify putative enhancers (reviewed in Tsompana and Buck, 2014). However, FAIRE-Seq can suffer from high background signals and therefore may not provide as sensitive detection of regulatory regions as ATAC-Seq (Baek and Sung, 2016). ATAC-Seq also provides advantages over ChIP-Seq, for example, the former does not require an antibody (Baek and Sung, 2016). Despite the advantages of ATAC-Seq, however, this approach is unable to determine the 3-D architecture of interactions, which other methods, such as the chromatin capture-based technologies (3C, 4C, 5C, Hi-C etc) can provide (reviewed in Risca and Greenleaf, 2015).

1.3 Enhancer functionality

1.3.1 Enhancer syntax

Through the composition of their binding sites and other factors such as insulators, enhancers display a 'regulatory logic'. In most cases enhancers have a combination of different affinity TFBS in a specific arrangement - enhancer "syntax" - and it is thought that the majority of enhancers have an optimal arrangement to allow them to precisely regulate gene expression (reviewed in Rickels and Shilatifard, 2018). However, as explained below, there are some cases where this syntax appears to change during evolution without an obvious effect on enhancer function and gene expression. Therefore it is not only crucial that we fully understand the functional motifs and their syntax for individual enhancers, but to explore how flexible enhancer syntax can be. This can help to more fully reveal the functionality of cis-regulatory regions beyond studying a single sequence from one species.

1.3.1.1 Transcription Factor Binding Site Affinity

It has long been recognized that TFs can bind to a range of sequence motifs and that these have different binding affinities. For example, pioneering work on the phage l operator and the yeast promoter Gal1 (Hochschild, Douhan and Ptashne, 1986; Giniger and Ptashne, 1988), as well as the SV40 enhancer, demonstrated the significance of low-affinity binding sites (Lee et al., 1987). In fact, it appears that the highest affinity binding sites might be used infrequently in nature, and instead, TFs may tend to bind to a range of binding sites of lower affinity, the importance of which is supported by many studies (reviewed in Crocker, Noon and Stern, 2016). For example, the *Otx-a* enhancer, which drives neural plate expression in the ascidian, *Ciona intestinalis*, contains binding sites for the TFs ETS and GATA that vary from the consensus binding motifs and display suboptimal spacing (Farley et al., 2015). However, it appears that the "suboptimization" of this enhancer is required to provide weak, but specific expression, because replacement with higher affinity sites results in ectopic expression (Farley et al., 2015). In addition, research on the *Pax6* locus in *Drosophila* has shown that it contains two evolutionary conserved low-affinity binding sites for the TF Prep1. Rowan et al. (2010) replaced the lower affinity sites with high affinity sites and they observed an alteration in the timing of transcriptional activation in addition to ectopic expression.

In *Drosophila melanogaster* embryos, a sharp expression boundary of *hunchback* (*hb*) in response to the concentration gradient of the morphogen Bicoid (Bcd) is, at least in part, coordinated by a cluster of Bcd binding sites in the *hb* enhancer (Schroder et al., 1988; Driever and Nusslein-Volhard, 1989; Struhl, Struhl and Macdonald, 1989; Ma et al., 1996). These Bcd binding sites are a mixture of both higher affinity

canonical motifs and non-canonical motifs of lower binding affinity (Driever and Nusslein-Volhard, 1989; Ma et al., 1996). Functional analysis suggests that the lower affinity sites are used to regulate gene regulation through cooperative binding (Giniger and Ptashne, 1988; Burz et al., 1998; Lebrecht et al., 2005). Furthermore, it is thought that Bcd binding to non-consensus sites, which requires the arginine at position 54 of the Bcd homeodomain, provides more specificity, since other lysine-50 class homeodomain-containing transcription factors cannot bind such motifs (Dave et al., 2000; Zhao et al., 2000a; Liu et al., 2018).

The findings of this work on the specificity of Bcd binding are consistent with research on the gene *shavenbaby* (*svb*). One of the enhancers of *svb* contains low-affinity binding sites for Ubx. Replacement of these low affinity sites with higher affinity sites resulted in ectopic expression (Crocker et al., 2015). It has been proposed that these low-affinity sites confer specificity for Ubx binding and that the ectopic expression observed when they are replaced with high-affinity sites is due to the recruitment of other homeodomain-containing TFs (Crocker et al., 2015).

In addition to studies that reveal the importance of low-affinity binding sites at specific loci, it has also been shown that they can be found genome wide. For example, Tanay (2006) revealed that there are extensive low-affinity transcriptional interactions throughout the yeast genome. Therefore, low-affinity binding sites appear to be a general mechanism for the regulation of gene expression in the genomes of eukaryotes (Crocker, Noon and Stern, 2016). However we do not fully understand the coordination and contribution by low-affinity and high-affinity sites in enhancers to ultimately lead to controlled gene expression. Further testing of different affinity sites requires not only the identification and validation of new enhancers but also the

binding sites that they contain.

1.3.1.2 Shadow Enhancers

The so-called shadow enhancers are cis-regulatory sequences that show similar or overlapping expression to other enhancers of the same gene (Fig. 2) (Barolo, 2012). They were first described as shadow enhancers by Hong, Hendrix and Levine (2008) upon their observation that some *Drosophila* genes had several enhancer elements that could drive similar expression (Hong, Hendrix and Levine, 2008). They termed the more proximal enhancer as the primary enhancer and the second, more distal, copy, the shadow enhancer (Fig. 1) (Hong, Hendrix and Levine, 2008). Since their discovery, it has been found that shadow enhancers are pervasive in the genes of *Drosophila* and other organisms (Cannavo et al., 2016). For example, study of the developing *Drosophila* mesoderm has identified a more complex regulatory landscape than previously thought, where many loci contain primary enhancers and three, four, or even five shadow enhancers (Cannavo et al., 2016).

The *snail* (*sna*) locus in *Drosophila* has provided many of the key insights into shadow enhancers (Dunipace, Ozdemir and Stathopoulos, 2011). The expression patterns driven by both the proximal and distal enhancer are similar but not identical. It has been demonstrated that the distal enhancer limits the expression domain of the proximal enhancer, and the proximal enhancer provides a “damper” for the expression levels driven by the distal enhancer (Dunipace, Ozdemir and Stathopoulos, 2011). Deletion of either the primary or shadow enhancer of *sna* results in less robust gene expression patterns, for example, under more extreme environmental conditions, such as increased temperature (Perry et al., 2010). These findings for the *sna* locus suggest

that the shadow enhancers function in concert with the so-called primary enhancers to ensure robust gene expression, which is consistent with the analysis of other genes.

In *Drosophila*, *svb* is important for the specification of larval trichomes (Wieschaus, Nusslein-Volhard and Jurgens, 1984; Mevel-Ninio et al., 1995; Payre, Vincent and Carreno, 1999; Sucena and Stern, 2000; Delon, Chanut-Delalande and Payre, 2003; Chanut-Delalande et al., 2006) and this gene is regulated by several enhancers with overlapping activities (McGregor et al., 2007; Frankel et al., 2011; Arif, Kittelmann and McGregor, 2015). A study by Frankel et al. (2010) revealed that the development of normal trichome patterns is still achieved at optimal temperatures after the removal of several distal enhancers leaving only a subset of the enhancers. However, when the embryos were allowed to develop at low or elevated temperatures, loss of trichomes was observed (Frankel et al., 2010). Likewise, trichome number was also reduced in the absence of the distal enhancers in a genetically sensitized background, generated by the removal of one copy of the *wingless* (*wg*) gene, which is known to be important for normal trichome patterning (Bokor and DiNardo, 1996; Frankel et al., 2010). These results highlight that shadow enhancers help to regulate robust gene expression by buffering against genetic as well as environmental perturbations.

Further studies have shown that several *Drosophila* gap genes also contain shadow enhancers, as well as primary enhancers (Perry, Boettiger and Levine, 2011). El-Sherif and Levine (2016) dissected the expression of the shadow and primary enhancers of *Krüppel* (*Kr*) and *knirps* (*kni*). They found that they drive largely overlapping expression patterns and the shadow enhancers are crucial for the dynamic anterior shifts in the expression of these genes (El-Sherif and Levine, 2016).

Another important example of shadow enhancers in *Drosophila* has been described at the *hb* locus. The distal enhancer is this case has been shown to aid in the production of a sharp boundary of activation by the Bcd gradient (Perry et al., 2010; Perry, Boettiger and Levine, 2011). When either the primary or shadow enhancer is deleted there is no obvious effect on the overall expression of *hb* and they do not act additively in the anterior of the embryo. However, they are functionally additive in the central domain of expression where the Bcd gradient is low and tight control is needed to switch the expression on and off.

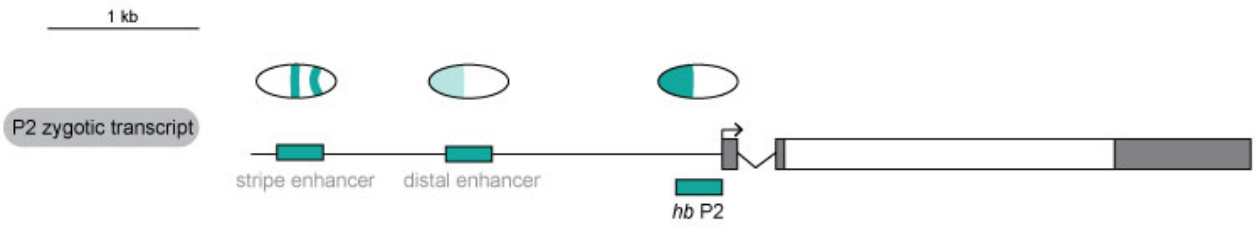


Fig 2: The arrangement of enhancers at the *hunchback* locus in *Drosophila melanogaster*. The transcript shown is the P2 zygotic form of *hb*. There is a proximal or primary enhancer located just upstream of the transcription start site, named *hb* P2, which drives expression in the anterior of the developing embryo. Further downstream is the shadow or distal enhancer, which also drives expression in the anterior of the embryo and aids in the production of a sharp expression boundary. In the *hb* locus there is also a stripe enhancer located further upstream than the shadow enhancer, which drives in a distinct stripe pattern

1.4 Enhancer Evolution

1.4.1 Enhancer conservation

While the protein coding regions of genes and even gene regulatory networks (GRNs) can be conserved across highly divergent taxa, as recently discussed in the case of metazoans, the enhancers themselves appear to be relatively less conserved among distantly related organisms (Maeso et al., 2013). This reflects the finding that while

phylogenetic footprinting shows that enhancers can be identified based on sequence conservation, the extent of conservation varies among clades and fades with evolutionary distance with the rate of turnover of enhancer sequences dependent on parameters including the mutation rate, effective population size, and genome size (Prabhakar et al., 2006; Villar, Flicek and Odom, 2014).

Nevertheless, the sequence of some enhancers, which are categorized as conserved noncoding sequences (CNSs) or highly conserved noncoding regions (HCNRs), has been maintained over several hundred million years during animal (Maeso et al., 2013) and plant (Burgess and Freeling, 2014) evolution. Furthermore, it has been observed by Burgess and Freeling (2014) that CNSs regulate similar categories of genes in both plants and animals including developmental genes and contain many TF binding sites.

Interestingly, enhancers can be conserved in function even when the primary DNA sequence is not (e.g. see Maeso et al., 2013). Therefore, with the application of recently developed tools to assay for epigenetic marks that signify open chromatin, and modeling of the 3-D structure of enhancers and their interactions with their target genes, as well as utilizing positional information, it will likely be possible to identify larger numbers of homologous enhancers across distantly related organisms in the near future (Maeso et al., 2013; Maeso and Tena, 2016).

1.4.2 Divergent Enhancers

Since the first comparative analyses of characterized enhancers between species, it is clear that highly divergent enhancer sequences can drive conserved expression patterns. This means that enhancer function can be maintained despite considerable

evolutionary loss and gain of binding sites, or turnover, as well as changes in the sequence, spacing, and orientation of individual binding sites. For example, there has been extensive turnover of the Bcd binding sites in the *hunchback* (*hb*) and *tailless* enhancers among higher dipterans to the extent that they are unalignable between some species, however they are still able to drive similar expression patterns (Lukowitz et al., 1994; Bonneton et al., 1997; McGregor et al., 2001; Shaw et al., 2002; Wratten et al., 2006) (Fig. 3).

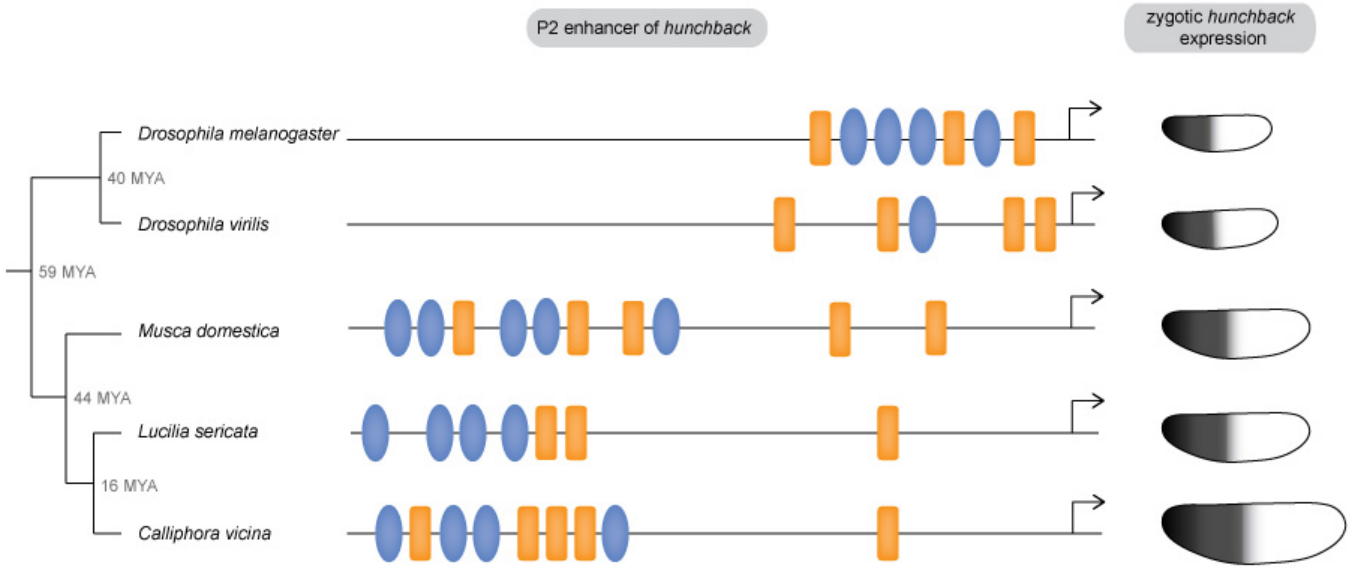


Fig. 3: Conservation and divergence of the *hunchback* P2 enhancer in higher Dipterans. The phylogenetic relationship between 5 different fly species is illustrated on the left. The structure of the each P2 enhancer is illustrated in the centre. The activation of *hb* expression in the anterior of the embryo is dependent on binding of Bcd to the P2 enhancer. In each species the P2 enhancer is composed of both high affinity Bcd sites (shown as orange boxes) and low affinity Bcd stes (shown as blue ovals), however, as illustrated, the number, spacing and composition of these sites differs dramatically between Dipteran species. For each of the species the expression of *hb* that is driven by the P2 is shown on the right. Evidently, despite extensive turnover in the Bcd sites and surrounding nucleotides, the anterior expression of *hb* is conserved in these species, over 50 MYA of evolution. Figure adapted from McGregor et al., (2001) and Buffry et al., (2016)

Similarly, the *eve* stripe 2 enhancer has diverged in sequence both among *Drosophila* species and between drosophilids and sepsids, and yet is still functionally conserved because these divergent sequences can drive very similar expression when introduced into *D. melanogaster* embryos (Ludwig and Kreitman, 1995; Ludwig, Patel and Kreitman, 1998; Hare et al., 2008). It has also been shown that the sparkling enhancer of *dPax2* has maintained the functional syntax required for eye expression, despite being subject to rapid turnover among *Drosophila* species (Swanson, Schwimmer and Barolo, 2011). The above examples illustrate the maintenance of enhancer function despite binding site turnover for a few well-characterized individual enhancers, but clearly this cis-cis coevolution is a genome-wide phenomenon (Moses et al., 2006; Schmidt et al., 2010; Paris et al., 2013; Stefflova et al., 2013; Arnold et al., 2014; Gordon, Arthur and Ruvinsky, 2015; Barkoulas et al., 2016). For example, comparison of the binding of several TFs between humans and mice has shown that even when these proteins regulate the same gene in each species, two-thirds of their binding sites have diverged (Odom et al., 2007). Carl and Russell (2015) also recently showed that there has also been turnover in Dichate and SoxNeuro binding sites among the genomes of *Drosophila* species, although interestingly the most conserved sites were bound by both of these TFs.

How are the mutations that lead to such turnover in binding sites tolerated? It was shown previously by Ludwig et al. (2000) using interspecific chimeric *eve* stripe 2 enhancers that such patterns of evolution with respect to this enhancer have been driven by stabilizing selection on the functional output of this enhancer despite turnover of binding sites (Ludwig et al., 2000). In support of this, it was found that compensatory changes between repressor (Gt) and activator (Bcd) binding sites in the

eve stripe 2 enhancer have evolved during the course of evolution of drosopholids (Martinez et al., 2014). It has also been shown recently that compensatory changes between *Krüppel* shadow enhancers underlie the conserved expression of this gene among *Drosophila* species even though the expression driven by the individual enhancers has diverged between these species (Wunderlich et al., 2015). Some studies have also suggested that TFs that bind to a number of degenerate motifs and clusters of binding can confer mutational robustness, this theory may help explain how enhancers composed of binding sites that differ in number and sequence can evolve but still drive the same spatio-temporal expression (True and Haag, 2001; Payne and Wagner, 2014; Payne and Wagner, 2015).

As well as having the ability to bind to a wide range of motifs (Badis et al., 2009), many transcription factors bind to similar sequences to other TFs (Payne and Wagner, 2014), and therefore the binding sites of the genotype network for one TF may be only one or two mutations from a motif that can be bound by a different transcription factor. Payne and Wagner (2014) suggested that this means TF binding degeneracy, in addition to conferring robustness, also increases evolvability even in the case of small genotype networks. Although such changes to binding sites in enhancers are often likely to be deleterious, this may help explain how the binding sites for different transcription factors evolve in enhancers thus facilitating coevolution, developmental systems drift (DSD), co-option, and the evolution of new gene expression patterns (True and Haag, 2001).To fully understand the functionality and logic of enhancers we must continue to investigate occurrences of mutational tolerance and DSD within cis-regulatory DNA, this will not only improve understanding of enhancers themselves but the whole regulatory landscape governing

gene expression.

1.4.3 The Molecular Evolution of Enhancers

Clearly compensatory changes within enhancers can maintain function and facilitate coevolution. This turnover in the binding site composition of enhancers can be generated by point mutations, slippage, unequal crossing over, and transposition (Tautz, Trick and Dover, 1986; Hancock et al., 1999; Dover, 2000; Kunarso et al., 2010; MacArthur and Brookfield, 2004). There is evidence that these changes can evolve quite rapidly. For example, it has been calculated that the gain/loss of binding sites among *Drosophila* species occurs at a rate of nearly 1% per million years (Costas et al., 2003).

What constraints and other factors are involved in such binding site evolution? A study by Tugrul et al. (2015), which modeled TFBS evolution using population genetics and biophysical parameters, suggests that the evolution of a single binding site in isolation is extremely slow in the absence of strong selection and that long binding sites with high affinity were unlikely to evolve in the time frame of eukaryotic speciation events (Tugrul et al., 2015). This may in part explain the presence of multiple homotypic binding sites of weaker affinity that have been observed in eukaryotic enhancers and in some cases functionally validated (reviewed in Crocker, Noon and Stern, 2016). This is also consistent with the results of simulations by He, Duque and Sinha (2012), which showed that there are many more possible ways to generate a functional enhancer using multiple weak binding sites than fewer binding sites with high affinity. Indeed, shorter binding site sequences generally confer weaker affinity but provide greater mutational robustness and this may help explain why

binding sites tend to be around 10 bp or less (Stewart, Hannenhalli and Plotkin, 2012).

Several studies have highlighted that binding sites could evolve where "pre-sites" or even previous degenerated sites were already present thus requiring fewer mutations to generate a binding site (MacArthur and Brookfield, 2004; Tugrul et al., 2015). Since there will also obviously be more pre-sites available for shorter TFBS, and the evolution of binding sites will also be faster in longer stretches of DNA, these findings are congruous with observed binding site lengths and the identification of enhancers that have evolved considerable distances from the genes that they regulate (Koshikawa et al., 2015) (Tugrul et al., 2015; Stewart, Hannenhalli and Plotkin, 2012).

It is also apparent that binding sites can evolve faster in circumstances where there is cooperative binding between TFs (Stefflova et al., 2013; Tugrul et al., 2015). In fact, Duque and Sinha (2015) considered cooperative binding and other factors in simulations to show that a complex enhancer, like those employed during *Drosophila* development, could evolve on a time scale of 0.5-1.5 million years, which is consistent with estimates by Arnold et al. (2014).

How long does it then take for compensatory changes, i.e., the loss and gain of binding sites to take place in enhancers? Durrett and Schmidt (2008) addressed this question by modeling how long it takes for two such mutations to evolve. They found that two mutations that facilitated switching between two binding sites could evolve in less than 1 million years given the presence of pre-sites (see above) even when these changes are nearly neutral (Durrett and Schmidt, 2008). These results fit with the turnover of binding sites among *Drosophila* species and are broadly consistent with

the more recent model by Tugrul et al. (2015) .

It is clear from the above studies that the evolution of new functional TFBS in enhancers relies on the presence of nearby pre-sites (MacArthur and Brookfield, 2004; Durrett and Schmidt, 2008; Tugrul et al., 2015). Are pre-sites always already randomly distributed in the genome or are there mutational mechanisms that may generate them more frequently? This question was addressed in a study by (Nourmohammad and Lassig, 2011) that provided some interesting insights into the origin of pre-sites. These authors found evidence that many neighboring TFBS in the *Drosophila* genome likely arose from the local duplication of an ancestral sequence (Nourmohammad and Lassig, 2011). Therefore, such duplications not only give rise to clusters of binding sites for a given TF (the loss and gain of which will lead to different compositions of binding site arrangements observed between species as described earlier) but also can potentially generate pre-sites for subsequent point mutations to provide functional binding sites for new TFs, which can lead to DSD or even neofunctionalization and phenotypic evolution (Nourmohammad and Lassig, 2011).

1.4.4 Cis-Regulatory Evolution and Phenotypic Diversification

1.4.4.1 The Genetic Basis of Phenotypic Evolution

It is clear that phenotypic evolution is predominantly caused by changes in gene regulation (Carroll, 2008). Furthermore, although still debated, there is an emerging consensus that changes in enhancer sequences is the major underlying mechanism (versus, for example, gene duplication or changes in coding regions alone in the absence of enhancer changes) (Carroll, 2008; Lynch and Wagner, 2008). In support

of this argument, there is a large and growing number of examples of phenotypic evolution caused by changes in enhancers (e.g. Belting, Shashikant and Ruddle, 1998; Marcellini and Simpson, 2006; Wang and Chamberlin, 2004; Chan et al., 2010; Loehlin and Werren, 2012; e.g. Arif et al., 2013; Koshikawa et al., 2015; Indjeian et al., 2016) although only in a relatively few cases have the actual mutations in these enhancers been identified and functionally validated (Gompel et al., 2005; Jeong, Rokas and Carroll, 2006; Prud'homme et al., 2006; Frankel et al., 2011; e.g. Arnoult et al., 2013; O'Brown et al., 2015). For a comprehensive overview, see Martin and Orgogozo (2013)

1.4.4.2 Trichome Pattern Evolution

The *Drosophila* trichome network (see also Results Chapters 2 and 3 for further details) has provided a classic example of mutations in cis-regulation leading to phenotypic changes. It has been demonstrated that changes in several different enhancers of the gene *svb* are required for the loss of dorsal 4○ trichomes in the first instar larvae of *Drosophila sechellia* compared to other species in the *D. melanogaster* species subgroup (Sucena and Stern, 2000; McGregor et al., 2007; Frankel et al., 2011; Stern and Frankel, 2013). Detailed analysis of the E6 enhancer has shown that at least five mutations in *D. sechellia* are required in this enhancer for the loss of trichomes (Frankel et al., 2011). All of the *D. sechellia*-specific changes act in the same direction (i.e., cause a reduction on trichome number), which is suggestive of directional selection (Frankel et al., 2011). Interestingly, the effects of the *D. sechellia*-specific changes in E6 individually do not add up to their cumulative effect on trichome numbers, which indicates that these changes combine epistatically to

contribute to the loss of trichomes in *D. sechellia* (Frankel et al., 2011).

Understanding the interactions and evolution of the GRN that produces trichomes has great potential to further understand enhancers, GRN architecture and evolution and phenotypic evolution.

1.4.4.3 Enhancers Can Evolve in Different Ways to Facilitate Phenotypic Change

Studies that have identified causative changes in enhancers have found that a range of different types of mutations are involved: from single point mutations in individual enhancers to the accumulation of multiple mutations in several enhancers of an individual gene, as well as transposition events. These studies have shown that mutations in enhancer sequences not only act additively but can also interact epistatically, which is likely to impact the relative order that mutations can accumulate to alter enhancer function. Furthermore, these changes can result in the modification of enhancers to increase or reduce gene expression or the co-option of existing enhancers or even the generation of new enhancers to create novel domains of gene expression. As has been pointed out previously (Stern and Frankel, 2013), the fact that large phenotypic effects can be caused by the accumulation of many mutations of small effect can reconcile phenotypic evolution at the microevolutionary and macroevolutionary scales. Therefore, microevolutionary studies with the power to identify mutations in enhancers underlying phenotypic change, as well as helping to understand the evolutionary forces that have shaped them, can provide important and broadly applicable insights into the genetic basis of phenotypic evolution over longer timescales (Nunes et al., 2013).

As the number of examples where the genetic basis of phenotypic evolution has been identified has grown, it appears that the enhancers of some genes appear to be "hot spots" for the convergent evolution of similar traits in different lineages (Martin and Orgogozo, 2013). For example, changes in *yellow* (*y)* enhancers underlie differences in abdominal pigmentation as well as wing spots (Arnoult et al., 2013; Gompel et al., 2005; Prud'homme et al., 2006); changes in the enhancers of *svb* have independently caused the loss of larval trichomes in the *Drosophila virilis* group as well as the *D. melanogaster* group (Sucena et al., 2003; Stern and Frankel, 2013; Frankel, Wang and Stern, 2012). It has been concluded that enhancer evolution is facilitated by the modular structure of enhancer sequences that limits the pleiotropic effects of mutations in these sequences (Stern, 2011; Carroll, 2008). Furthermore, it has been suggested that some genes occupy key nodes in GRNs and act as "input-output" devices, and therefore the accumulation of mutations in the modular enhancers of these genes is favored and may explain convergent evolution (Stern and Orgogozo, 2008; Stern and Orgogozo, 2009). Examples of such input-output devices that have been characterized to date include *svb* (larval trichomes) (Sucena and Stern, 2000; Sucena et al., 2003; McGregor et al., 2007; Frankel, Wang and Stern, 2012), as well as *Optix* (butterfly wing patterns) (Reed et al., 2011; Martin et al., 2014) and *Dll* (*Drosophila* wing spots) (Arnoult et al., 2013).

This has led to the argument that if we understand GRNs in detail then it may be possible to predict the loci of evolution (Stern and Orgogozo, 2008; Stern and Orgogozo, 2009; Stern, 2011). However, the finding that a cis-regulatory change leading to differences in the expression of *microRNA-92a* underlies the evolution of *Drosophila* leg trichome patterns suggests that this predictability may not always be

straightforward because the tissue-specific context as well as whether the trait represents a gain or loss is likely to be important (Arif et al., 2013; Arif, Kittelmann and McGregor, 2015; Kittelmann et al., 2018). Indeed, further examples of trait differences mapped to the resolution of the causative genes and individual mutations in the context of well-characterized GRNs are needed to further explore the predictability of the genetic basis of phenotypic evolution.

1.5 Aims of this thesis

Building on both what is known about enhancers as well as the open questions about these cis-regulatory elements as I have reviewed here in the Introduction to my thesis, the overall experimental aims of my thesis were to identify new enhancers, understand the binding site composition of known enhancers and to study the evolution of binding site syntax. These aims were addressed in the following three Results chapters.

Chapter 2

To combine bioinformatics and functional reporter assays to identify and characterize new enhancers by:

- Analyzing recently generated ATAC-seq data for the T2 legs in *Drosophila melanogaster* to identify regions of accessible chromatin in the Hox gene *Ubx*.
- Analyzing candidate regions identified above with reporter assays to determine expression patterns.
- Assessing the activity of candidate enhancers with functional studies to see if

there is an effect on trichome development.

- Investigating the potential TFs that might bind to candidate enhancers using bioinformatics approaches.

Chapter 3

To combine bioinformatics and genetic methods to investigate motif composition at a sub-set of enhancers in *Drosophila* by:

- Identifying a group of genes with related functions and regulators, in this case genes that are predicted targets of *svb* and potentially also *miR-92a* during trichomes patterning
- Functionally testing predicted genes for phenotypic effects on *Drosophila* trichomes
- Performing a thorough motif analysis using existing bioinformatic tools to identify common motifs in the selected gene set.
- Analyzing the composition of motifs in open chromatin in these genes and combine this data with other bioinformatic methods with the goal of predicting novel enhancers.

Chapter 4

To investigate the binding site turnover and its potential functional consequences in *Drosophila hb* enhancers by:

- Identifying nucleotide changes in a large number of *Drosophila* species for the Bicoid dependent P2 enhancer of *hb* to assess the extent of enhancer turnover.

- Based on the above findings to generate tools for live imaging to fully characterize the effects of mutations in the P2 on transcription.

Chapter 2

Identification and charaterisation of post-embryonic enhancers in *Drosophila*

2.1 Introduction

2.1.1 Post-embryonic roles of Hox genes

Hox genes are a highly important and conserved set of genes that function during embryogenesis to determine positional identity along the anterior-posterior axis of all bilaterian animals (Maeda and Karch, 2009; Mallo, Wellik and Deschamps, 2010; Pearson, Lemons and McGinnis, 2005). However, Hox genes also function post-embryonically, where they can have more specific roles such as the specification of organs and cell-types (e.g Rux and Wellik, 2017). Several post-embryonic roles of Hox genes have been identified in *Drosophila*, for example, the specification of certain subtypes of cells in the *Drosophila* CNS (Estacio-Gomez et al., 2013; Kannan et al., 2010) and the development of a single cell type, the larval oenocytes, by *abdominal-A* (*abd-A*) (Brodu, Elstob and Gould, 2002). Furthermore, *Ultrabithorax* (*Ubx*) has been shown to determine segment specific size and bristle patterns of the developing limbs of *Drosophila* (Davis et al., 2007; Lohmann and McGinnis, 2002; Rozowski and Akam, 2002; Stern, 1998; Stern, 2003). For example, *Ubx* regulates differences in the morphology of the legs on the second (T2) and third (T3) thoracic segments (Stern, 1998; Stern, 2003) and variation in the expression levels of *Ubx* have been associated with differences in the patterning of leg trichomes (Davis et al., 2007; Stern, 1998). Moreover, the patterning of mechanosensory bristles, michrochaete, on T1 and T3

legs is regulated by *Sex combs reduced* (*Scr*) and *Ubx* respectively (Barmina and Kopp, 2007; Shroff, Joshi and Orenic, 2007). However, despite a growing number of studies, compared to their roles during embryonic development, relatively little is known about how Hox genes are integrated in to downstream post-embryonic gene regulatory networks (GRNs) to control more fine scale aspects of development. It has been suggested that to achieve these downstream fine scale patterning functions, Hox levels need to be tightly regulated in space and time (Mallo and Alonso, 2013), but the mechanisms that underlie the finely tuned post-embryonic regulation of Hox genes are not well understood.

One explanation is that the control of Hox genes by specific post-embryonic enhancers may allow the genes to be integrated into downstream GRN tiers through tight temporal and spatial control, which would allow Hox genes to specify certain cell-types and fine-scale morphological structures. To date, however, only a few of post-embryonic enhancers of Hox genes have been identified, for example an enhancer of *Scr* which is responsible for patterning the first leg in *Drosophila* (Eksi et al., 2018). Therefore, it is necessary to identify and fully investigate more such enhancers of Hox genes to get a better understanding of how they fit into downstream developmental pathways and their roles in the regulation of fine-scale morphology.

2.1.2 Trichome development as a model to study post-embryonic development and enhancers

The development and patterning of trichomes in *Drosophila* species has proven to be an excellent model to study GRNs (Reviewed in Arif, Kittelmann and McGregor, 2015). Furthermore, the involvement of *Ubx* in leg trichome patterning makes it an

excellent system to uncover more about the post-embryonic function and regulation of Hox genes (Davis et al., 2007; Stern, 2003). Trichomes are short, non-sensory, actin protrusions that are found on insect bodies throughout all stages of life (Reviewed in Arif, Kittelmann and McGregor, 2015). Despite being an excellent biological model to study genetic regulation, the function of a number of these structures is enigmatic, but they have been hypothesised to be involved in processes such as aerodynamics, thermal regulation and larval locomotion (Balmert et al., 2011; Ditsche-Kuru et al., 2011; Goodwyn, Voigt and Fujisaki, 2008; Goodwyn et al., 2008; Inestrosa et al., 1996). The larval cuticle of *Drosophila* displays a distinct pattern of trichomes which differs between species, the network of genes that produce this pattern has been studied in great detail and we now have a well characterised GRN for the formation of larval trichomes (Reviewed in Arif, Kittelmann and McGregor, 2015) (Fig. 4A). In brief, the gene *shavenbaby* (*svb*) acts as a so called 'input-output' device in the network, integrating information from a number of upstream factors, including Ubx, and directing the expression of a downstream battery of effector genes which determine the formation of the trichomes and their morphology (Fig. 4A) (Chanut-Delalande et al., 2006; Delon, Chanut-Delalande and Payre, 2003; Menoret et al., 2013; Stern, 1998; Sucena et al., 2003; Sucena and Stern, 2000). Further research on the evolution of this network showed that each case of divergence in the pattern of larval trichomes among *Drosophila* species, is associated with changes in the enhancers of *svb* (Frankel et al., 2011; Frankel, Wang and Stern, 2012; McGregor et al., 2007; Preger-Ben Noon, Davis and Stern, 2016; Sucena et al., 2003; Sucena and Stern, 2000). This supports the hypothesis that it is commonly changes in enhancers that ultimately lead to phenotypic change and highlights the importance that

enhancers can play in GRNs and morphological evolution (Carroll, 2008).

The second leg (T2) of *D. melanogaster* and related species also display a distinct trichome pattern: a patch of cuticle on the posterior of the femur remains free from trichomes - this is known as the naked valley (NV) (Fig. 4E-F). This phenotypic feature is also used as a model for GRN studies and evolutionary phenotypic change (Reviewed in Arif, Kittelmann and McGregor, 2015). Recently, we started exploring the differences between the GRN employed to make leg trichomes versus the GRN underlying the formation of larval trichomes by combining RNA-seq data from both larvae and legs with ATAC-seq data for the legs, in the window of developmental time when trichomes are specified (Kittelmann, Buffry et al., 2018). This allowed us to not only compare which genes are expressed between these two developmental stages, but also to look at the chromatin structure in the nuclei of T2 leg cells as an indicator of which regions of chromatin are accessible and therefore may indicate enhancers (Kittelmann et al., 2018). This comparison led to the finding that some of the key genes involved in the larval network are not expressed in the leg, meaning the trichome GRN architecture is different between these two developmental contexts (Fig. 4B) (Kittelmann et al., 2018)

In particular, we were also interested in investigating the role of *Ubx* in these two GRNs. In the larvae network *Ubx* has been shown to activate *svb*, which leads to the activation of trichome formation (Crocker et al., 2015) (Fig.4). In the leg network however, *Ubx* represses trichomes (Davis et al., 2007; Inestrosa et al., 1996; Stern, 1998) and it has been proposed that *Ubx* activates *miR-92a* to achieve this leading to the development of naked cuticle (Fig.4) (Arif et al., 2013). Interestingly, in a previous study variation in the expression of *miR-92a* has shown to be responsible for the

differences in NV size seen between strains of *D. melanogaster* (Fig. 4) (Arif et al., 2013).

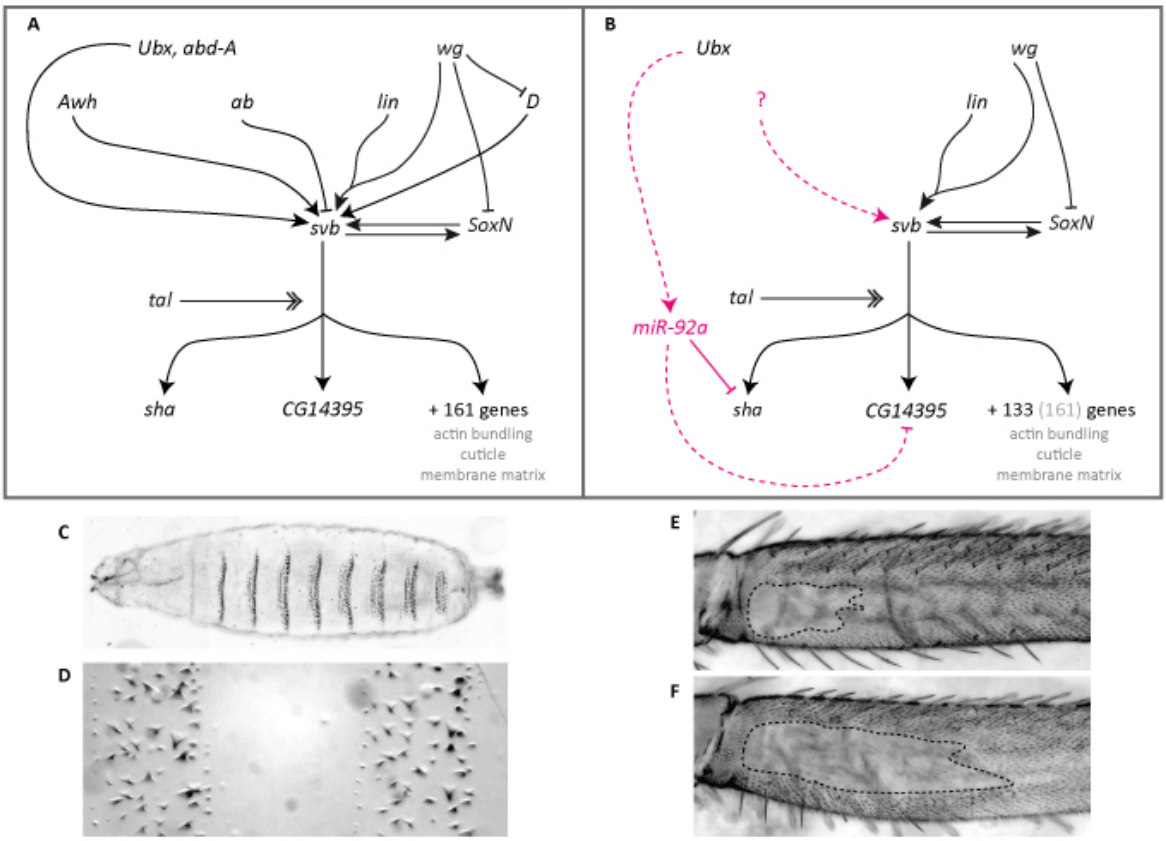


Fig.4. The GRN for larval and leg trichomes. (A) a simplified GRN for the development of larval trichomes, for a review see Arif et al., 2015. (B) a simplified GRN for the development of leg trichomes, pink lines indicate some of the interactions that differ between these two contexts as outlined in Kittelmann, Buffry et al., 2018. Note that out of the 161 genes identified to be downstream of *svb* in the larval network, only 133 of these genes can be identified in the leg network. Also note the difference in the role and interactions of *Ubx* (C-D) trichome patterns on the larval epidermis of *Drosophila*. (E-F) Trichomes on the posterior femur of the second leg, dashed lines show the region of trichome-free cuticle known as the naked valley. (E) OregonR strain of *Drosophila* (F) *e, wo, ro* strain of *Drosophila*. Figure adapted from Kittelmann, Buffry et al., 2018.

2.1.3 Previous leg trichome enhancer predictions

Previous studies have also found that the size of the NV varies substantially within and between species, from large NVs in *D. simulans* to small NVs in some strains of *D. melanogaster* to the lack of NV in *D. virilis* (Stern, 1998). These differences were

associated with differences in the expression of *Ubx* (Stern, 1998). Through comparative analysis of the distribution of Ubx protein in different *Drosophila* species, it was shown that Ubx protein is present in *D. melanogaster* T2 legs in the region of the naked valley, but that this expression is missing in the T2 legs of *D. virilis* (Davis et al., 2007; Stern, 1998). This research suggested that evolutionary changes in *Ubx* expression in T2 legs are likely to be attributable to the presence of a leg-specific enhancer of *Ubx* (Davis et al., 2007). After the prediction of this novel post-embryonic *Ubx* enhancer, Davis et al. (2007) assayed all available regulatory mutations of the *Ubx* locus as well as testing new deficiencies (Fig. 5). This allowed them to exclude around 100 kb in and around the *Ubx* locus as containing the T2 posterior specific enhancer. They then assayed a further 30 kb using reporter constructs, but none of these regions was able to drive leg expression. In total they investigated over 95% of the *Ubx* locus, however, they were unable to identify any cis-regulatory sequences that could drive post-embryonic expression of *Ubx* in T2 (Fig. 5) (Davis et al., 2007).

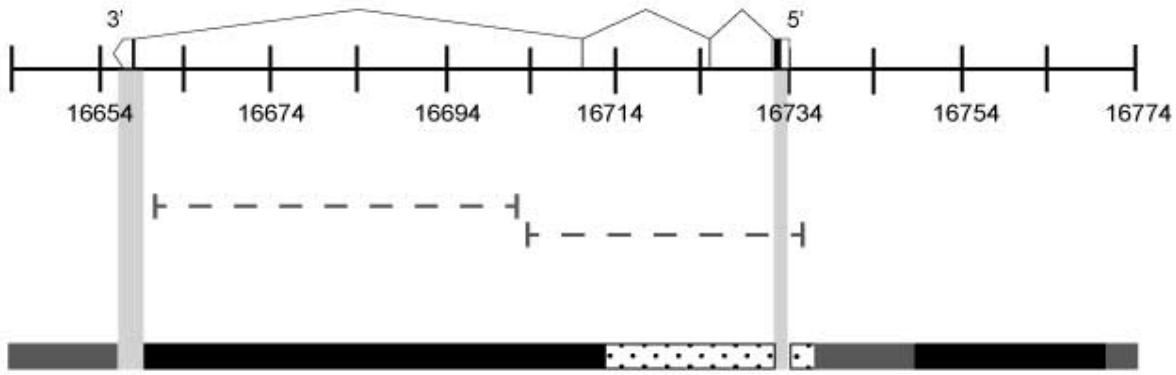


Fig.5. The *Ubx* locus showing regions tested for a NV specific enhancer. The 4 exons of *Ubx* are depicted above and the bars underneath the genomic positions indicate regions tested by Davis et al., 2007 for the presence of a NV enhancer. Grey horizontal bars indicate regions tested by inversions, black bars show regions tested using deficiencies, the spotted box shows the region scanned with reporter constructs, the dotted lines indicate two large deficiency lines used which cover most of the *Ubx* locus and finally grey vertical bars indicate regions that were not tested. Figure adapted from Davis et al., 2007.

Systematically testing enhancer expression *in vivo* remains a critical step for the identification and functional characterization of enhancers (Barriere and Ruvinsky, 2014; Wittkopp and Kalay, 2011). To better understand the mechanisms responsible for the spatial modulation of post-embryonic expression of Hox genes in downstream patterning networks, I sought to identify enhancers regulating the post-embryonic expression of *Ubx* and specifically in T2 legs. By using the ATAC-seq data for the T2 legs I have discovered a putative enhancer of *Ubx* which is functional in the NV region of the femur. Furthermore, by studying smaller and larger fragments surrounding this enhancer element I can conclude that the regulation at the *Ubx* locus in this post-embryonic context, much like that in embryonic development, is highly complex. However, I hypothesise that by evolving specific post-embryonic enhancers Hox genes can control the patterning and development of fine-scale morphological structures. Moreover, I have started investigating the factors that bind to this novel enhancer to determine how this enhancer is integrated into the GRN underlying the development of leg trichomes.

2.2 Methods

2.2.1 Fly stocks and genetics

All stocks used were kept on standard yeast extract-sucrose medium at 25˚C. Reporter lines VT42732, VT42733 and VT42734 were obtained from the Vienna *Drosophila* Resource Centre (VDRC, www.vdrc.at). Reporter lines 31C06, 32B04, 31F12, 32B03 and 31E11 were obtained from the FlyLight enhancer collection (Jenett et al., 2012). To test the expression driven by all lines, they were crossed to a UAS-stinger-GFP line (Bloomington stock #65402), and/or UAS-Sha-ΔUTR (Bloomington stock #32096).

2.2.2 Cloning

Fragments VT2732+VT42733, VT42733+VT42734, VT42732+VT42733+VT4273, *Ubx-1*, *Ubx-2*, *Ubx-3*, *Ubx-4* and *Ubx-5* were cloned from gDNA (*D. melanogaster* Oregon R strain) and initially inserted into a TOPO/D vector. LR gateway cloning was then used to subclone the fragments into the pBPGUw plasmid upstream of GAL4 (gift from Gerald Rubin, Addgene plasmid # 17575). These constructs were used for phiC-31 mediated germline transformation into the stock y, w; M(eGFP, vas-int, dmRFP)ZH-2A;; M(attP)ZH-86Fb (Bloomington stock 24749) by either BestGene Inc, Cambridge fly injection facility or in-house based on the protocol from Gompel & Schröder (2015).

2.2.3 GFP, protein trap and naked valley analysis

Prepupae from reporter lines crossed to UAS-Stinger GFP were collected and allowed to age to between 20-28 hours after pupal formation (hAPF), the window when T2 trichome patterning is regulated by Ubx (Stern, 1998). GFP expressing pupae were imaged on a Zeiss Axiozoom. The endogenously tagged *Ubx* line (GFP-Ubx$^{3.005}$) was kindly donated by I. Lohmann from the publication Domsch et al. (2019). Pupae were prepared the same as above but imaged on Zeiss 800 confocal with the 10x objective and 2% laser power. For the analysis of the size of the naked valley, T2 legs were dissected from adults and mounted in Hoyer's medium/lactic acid 1:1 and imaged under a Zeiss Axioplan camera using ProgRes MF cool camera (Jenaoptik, Germany). The size of the naked valley was measured using Fiji software (Schindelin et al., 2012) and statistical analysis was performed in R-Studio version 1.2.1335 (RStudio Team,

2015). In all cases where the size of the NV was measured both parental lines were used as controls as well as a control cross between the trichome activating line, UAS-Sha-ΔUTR and the attP containing line which was used to create the driver line being tested.

2.2.4 Analysis of potential TFs that may bind to Ubx VT42733

Analysis of potential TFs was carried out using the JASPAR database (Khan et al., 2018) with the relative profile score threshold set to 90%. All TFs returned for the search with VT42733 were cross checked with the RNA-seq data for T2 pupal legs (Kittelmann et al., 2018). The function of remaining TFs was taken from FlyBase. For the analysis using MEME (Bailey et al., 2009), sequences for VT4733 were extracted from the genomes of *D. melanogaster* (r6.18), *D. simulans* (r2.02), *D. sechellia* (r1.3), *D. mauritiana* (r1.0) and *D. yakuba* (r1.05) using BedTools (Quinlan and Hall, 2010). Sequences were then aligned using T-Coffee (default parameters) (Madeira et al., 2019) and viewed using BoxShade. Sequences were then submitted to the MEME web-server (Bailey et al., 2009) with the following parameters: ANR (any number of repetitions), sites may be located on either strand, maximum number of motifs 10, minimum motif width 6, maximum motif width 10, maximum number of sites per motif 25 and minimum number of sites per motif 2. The resulting motifs were processed using TomTom (Gupta et al., 2007), and only the hits with the lowest E-values from TomTom were reported.

2.3 Results

2.3.1 Several regions of open chromatin at the Ubx locus drive expression in *Drosophila* pupal legs

To identify putative post-embryonic leg enhancers of *Ubx*, I used the ATAC-seq profile for T2 legs as a guide (Kittelmann et al., 2018). ATAC-seq peaks indicate the presence of accessible chromatin regions and therefore putative cis-regulatory elements at several locations in this locus. Note that in this study some peaks were not tested because they either correspond to a transcription start site (TSS), known *Ubx* enhancers or they were already covered with reporter fragments by (Davis et al., 2007) (Fig. 5).

To identify enhancer regions, I then took advantage of existing reporter constructs (VDRC and FlyLight), to assay for expression in T2 legs driven by several of the open chromatin peaks in the 70 kb of the *Ubx* locus (Fig 3). Three overlapping lines were tested from the Vienna collection - VT42732, VT42733 and VT42734 – each of these lines contains approximately 2.2 kb of DNA from the *Ubx* locus and overlap by ~400 bp.

VT42732 did not drive any detectable expression in any pupal tissues at 24 hAPF (data not shown), the window when trichomes are specified (Stern, 1998). In contrast, VT42733 was able to drive strong expression in several tissues including, but not limited to, the legs, antenna, genitalia and the ventral region of thoracic segments (Fig. 6). VT42734 also drove expression at 24 hAPF but in a pattern much more specific to the developing legs (Fig. 6).

I then tested lines from the FlyLight collection for their ability to drive

expression in the developing legs in pupae. GMR31C06 and GMR32B04 did not drive expression in any tissues in the pupae (data not shown). GMR31F12 produced a stripe like expression pattern which appeared to be localised to internal tissues in the developing abdomen (data not shown). GMR32B03 showed strong expression in the wings, legs and other head cuticle tissues (Fig.6) and GMR31E11 was also able to drive expression in the developing legs although it appears this expression is restricted to the joint separating the femur from the tibia (Fig. 6) GMR31E11 also displayed localised spots of expression along the periphery of the abdomen (Fig. 6). Finally, line AB01 displayed very specific expression in the developing histoblasts but no other tissues (supplementary fig. 1).

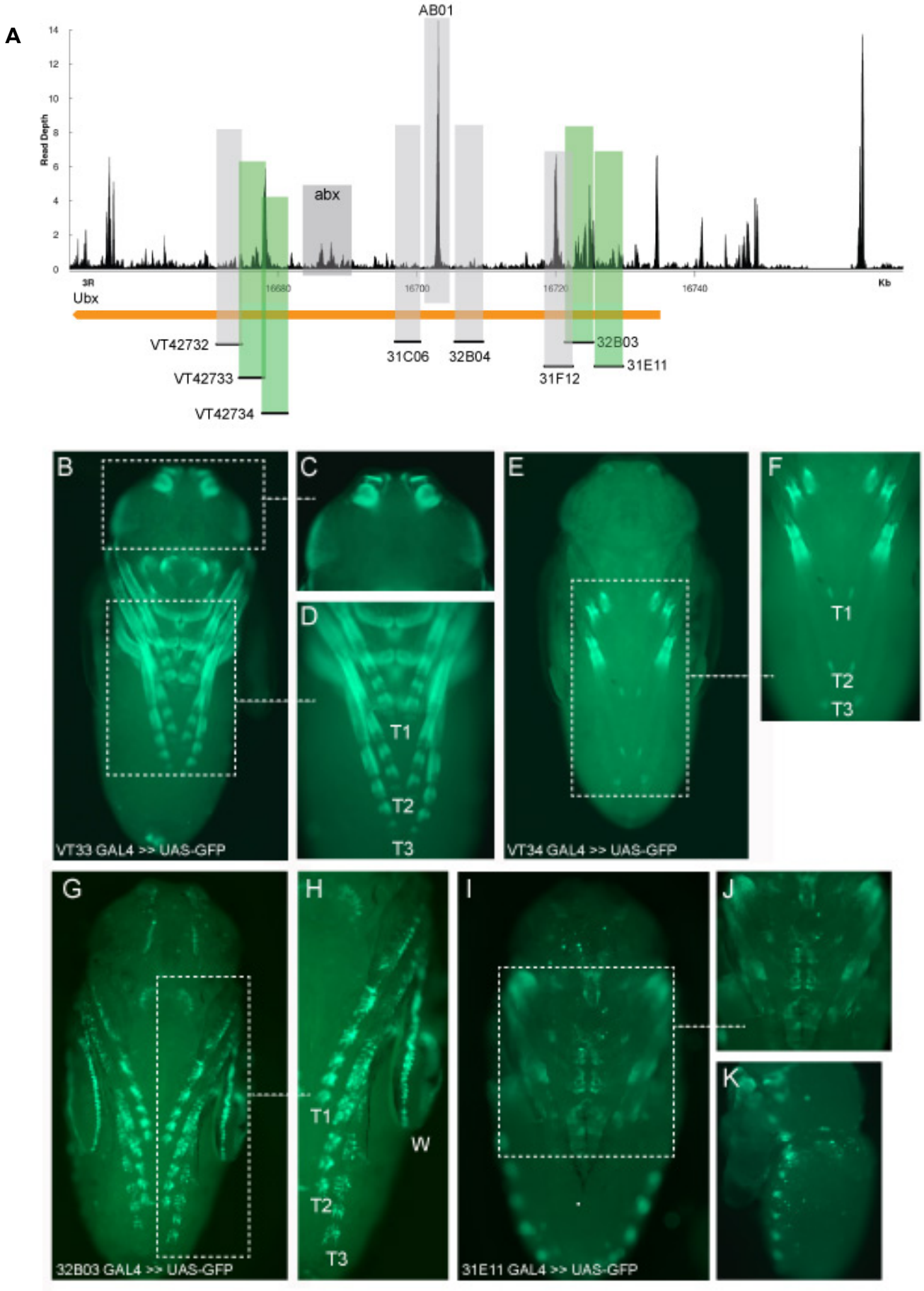

A
AB01
Read Depth
abx
3R
Ubx
16680
16700
16720
16740
Kb
VT42732
VT42733
VT42734
31C06
32B04
31F12
32B03
31E11
B
C
D
E
F
T1
T2
T3
VT33 GAL4 >> UAS-GFP
VT34 GAL4 >> UAS-GFP
G
H
I
J
K
W
32B03 GAL4 >> UAS-GFP
31E11 GAL4 >> UAS-GFP

Fig. 6. ATAC-Seq profile of the *Ubx* locus for T2 pupal legs at 21-28 hAPF (Kittlemann et al., 2018). (A) ATAC-seq profile showing location of driver lines tested. Bars over peaks indicate lines that were tested for expression in 24-hour pupal legs. *abx* is a previously tested and characterised embryonic enhancer of *Ubx* and was therefore not tested. All lines except AB01 are from either the Vienna GAL4 collection or the FlyLight collection. Grey bars indicate lines that did not express in pupal legs 24 hAPF, green bars indicate lines that did express in T2 pupal legs. (B-K) GFP expression in whole pupae at 24 hAPF. Expression from VT42733 is shown in panels B-D, this line expresses strongly throughout the pupal legs and also in the developing antenna. VT42734 driven expression can be seen in panels E-F, this expression pattern is much more localised to the legs of the pupae. GMR32B03 is depicted in panels H-I, this line also drives expression strongly in the developing limbs of the pupae, in a stripy pattern. This line also gives a line of expression at the periphery of the developing wing. GMR31E11 is shown in panels I-J, this expression also seems to be in the developing leg but localised to more proximal limb segments. This line also drives expression on the periphery of the abdomen in localised spots (J). T1 – first leg, T2 – second leg, T3- third leg, w – wing.

2.3.2 Functional testing of putative enhancers

To test the ability of the putative enhancer regions identified from the ATAC-seq and expression data to functionally specify the trichome pattern seen on the T2 legs I crossed the GAL4 driver lines to a leg trichome activating line - *UAS-shaΔUTR*. All lines were tested and compared to the parental lines as controls. Where the region being tested is not a NV enhancer I expected to see an intermediate size of NV when compared to the parental lines. Due to *UAS-shaΔUTR* being a trichome activating driver, if the region being tested is in fact an enhancer I expected to see trichome growth in the NV and a reduction in size of the NV in comparison to the parental lines. Consistent with what was seen in the expression analysis where VT42732 does not express in the NV, when crossing it to *UAS-shaΔUTR* I did not see a reduction in the size of the NV (Fig. 7).

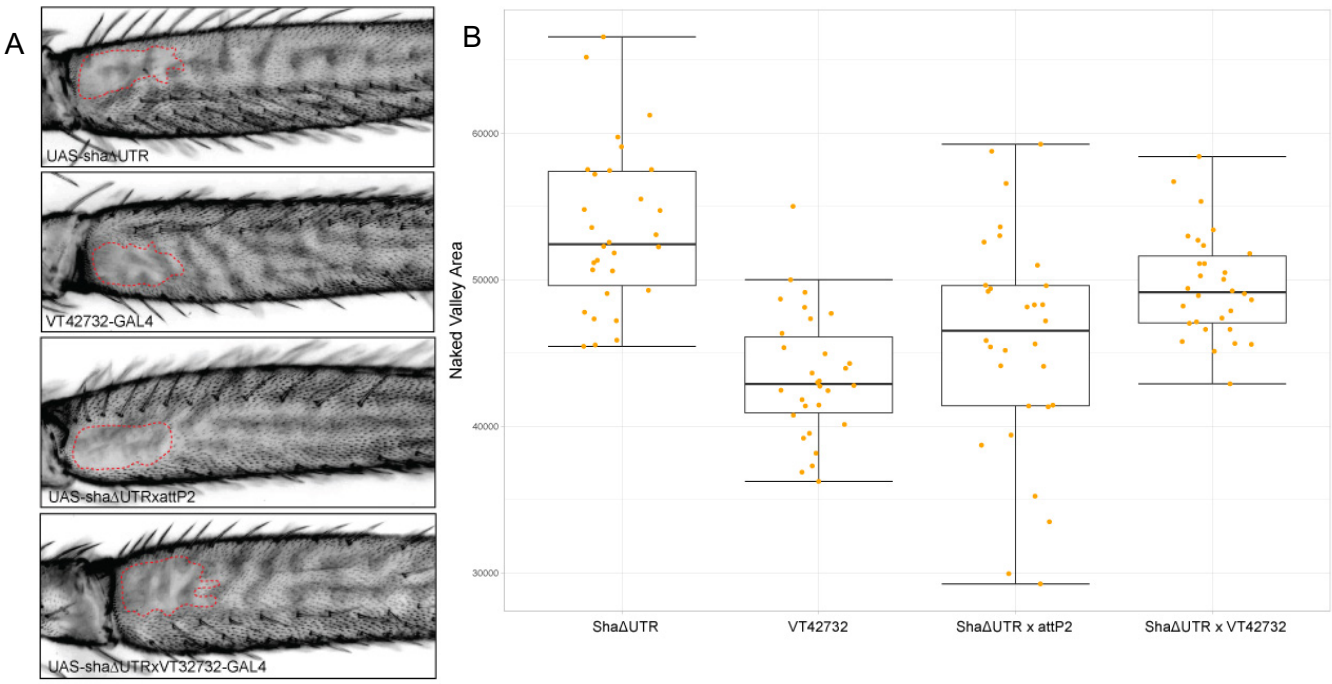


Fig. 7. VT42732 has no effect on NV size. (A) Second legs of control lines and VT42732xUAS-ShaΔUTR showing representative NV sizes. (B) NV area measurements from 30 individuals from each line. The driver line VT42732 has no effect on the size of the NV compared to controls. Data was tested using Shapiro-Wilk followed by an ANOVA.

However, VT42733 did produce a significant reduction in the size of the NV compared to all three controls (Fig. 8).

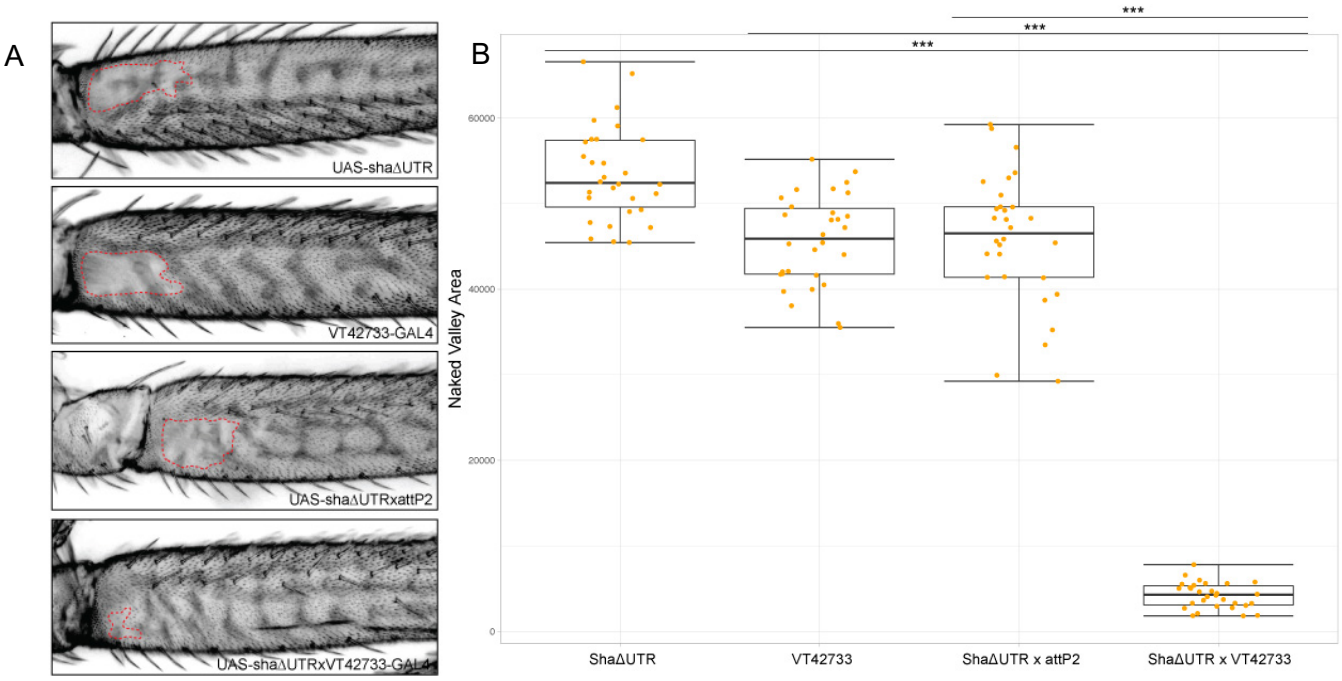


Fig. 8. The driver line VT42733 generates trichomes in the NV (A) T2 legs from lines tested showing representative NV sizes (B) Measurements of NV area from 30 individuals for each genotype. Line VT42733 produces a significant decrease in the size of NV in comparison to all control lines (*** $p>0.001$). Data was tested using Shapiro-Wilk followed by an ANOVA. Significance between groups was tested using Tukey's multiple comparison test.

VT42734 did also show a reduced NV size in comparison to both the parental controls (Fig. 9) but when compared against the control cross, UAS-ShaΔUTRxattP2, this difference was not significant (Fig. 9). From these results it is possible that this line drives some weak expression in naked valley cells.

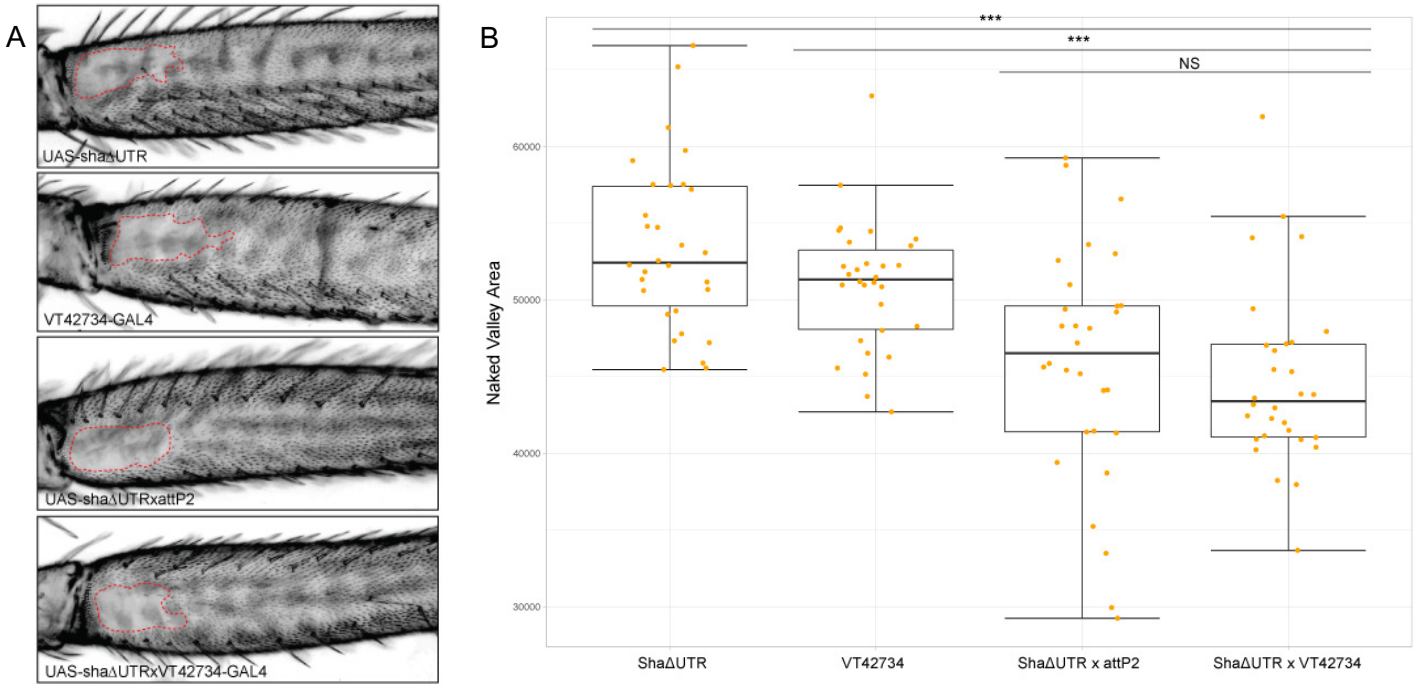


Fig 9. VT42734 may drive weak expression in the NV. (A) Representative T2 legs from all lines showing the area of the NV outlined in red. (B) Measurements of the NV area from 30 individuals for each genotype. The size of the NV is significantly smaller for VT42734xUAS-ShaΔUTR when compared to the parental lines, (*** p-value < 0.001) Tukey's multiple comparison test. However, there is no significant difference between the control cross, UAS-ShaΔUTRxattP2 and VT42734xUAS-ShaΔUTR (NS p-value < 0.887). Data was tested using Shapiro-Wilk followed by an ANOVA. Significance between groups was tested using Tukey's multiple comparison test.

2.3.3 Further delimiting the Ubx naked valley enhancer

VT42733 encompasses 2.2 kb of DNA of the *Ubx* locus and partially overlaps with VT42734 (Fig. 6). To determine if a smaller region within this sequence is responsible for the NV function I chopped *Ubx* VT42733 and *Ubx* VT42734 into five overlapping fragments of around 700 bp each. These fragments were then crossed to UAS-GFP to

assay the expression in pupal legs (Fig. 10).

Ubx-1 drives expression in the developing legs in pupae, and also in the antennae and developing retinal cells (Fig. 10). *Ubx-2* drives much more restricted expression in small patches in pupal legs and in the head cuticle (Fig. 10). *Ubx-3* also drives expression in the legs but mainly in the head and thoracic segments - interestingly this line also shows a stripe-like pattern on the ventral side of the abdomen which was not seen in any of the other driver lines tested (Fig. 10). For *Ubx-4*, despite several attempts both in-house and with commercial injection facilities, no transgenic flies were produced for this line due to a high level of lethality in the G0s, this suggests that this insert may be toxic to the fly. *Ubx-5* does not drive any expression in the pupal legs at this stage of development, however it did drive a small amount of expression on the ventral abdomen which seems to overlap the expression of *Ubx*-3 (Fig. 10).

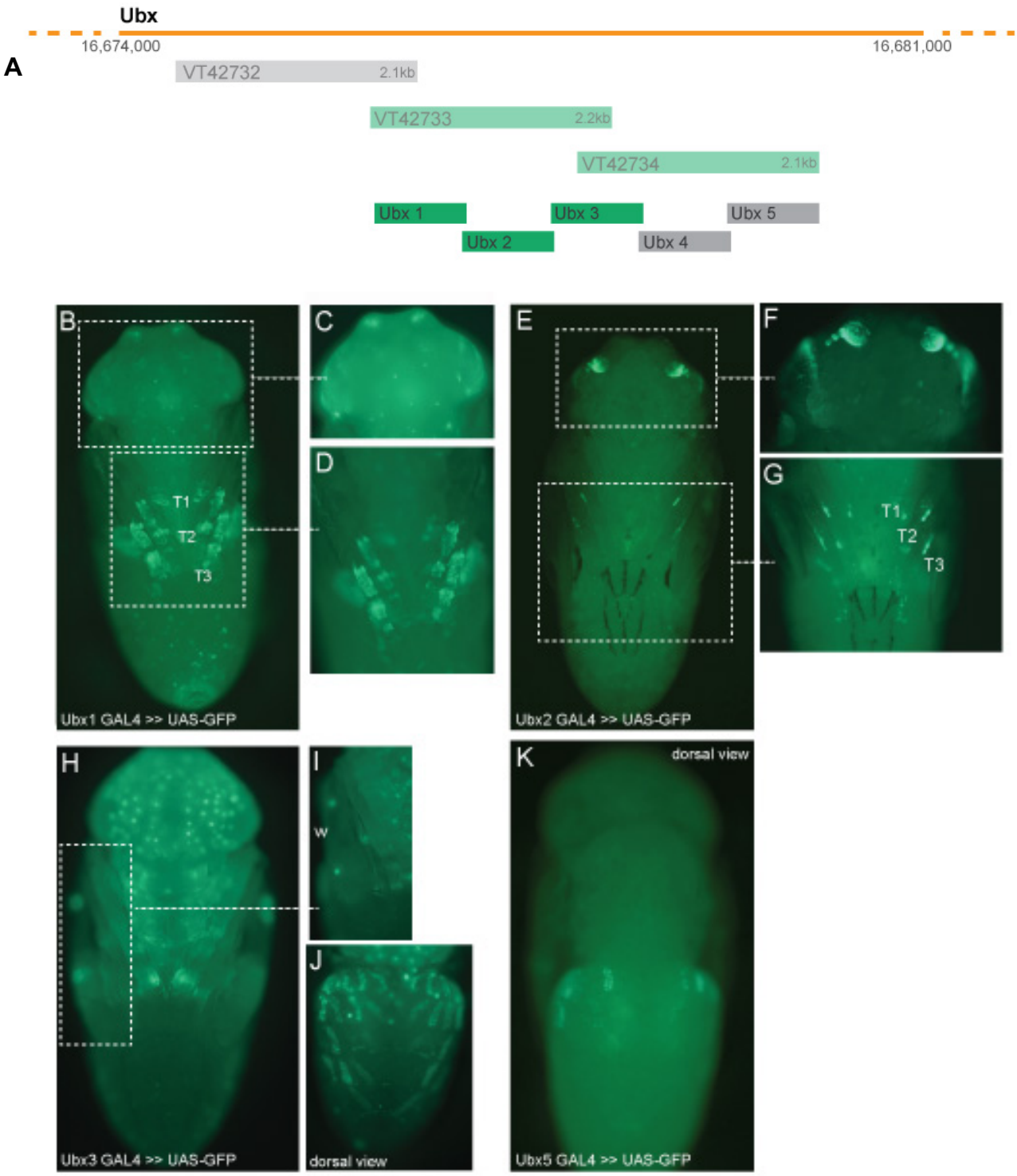


Fig. 10. Expression driven by sub-regions of VT42733 and VT42734. (A) Positions of the five overlapping fragments of around 700 bp each (*Ubx 1-5*) used to drive GFP in pupae that was assayed at 24 hAPF. *Ubx-1* shows expression in the developing legs and also various tissues in the developing head (B-D). *Ubx-2* drives minimal expression in the developing limbs but shows strong expression in the antennae (E-G). *Ubx-3* produces some very localised dots of expression in the wing and some further expression in the head cuticle and developing limbs (H, I). *Ubx-3* also drives an interesting pattern on the dorsal abdomen of the developing pupae (J). *Ubx-5* also drives on the dorsal abdomen, although reduced in comparison to *Ubx-3*, the expression seems to be overlapping (K).

2.3.4 Ubx-1 is capable of generating trichomes in the NV

To test the functionality of these new smaller enhancer regions, I crossed all five new fragments to the trichome activating line *UAS-shaΔUTR*. *Ubx-2, Ubx-3* (Fig 12-13) and *Ubx-5* (data not shown) did not show any effect on the size of the NV in comparison to controls. However, *Ubx-1* was able to produce trichomes in the NV, albeit in a patchy and irregular pattern (Fig. 11)

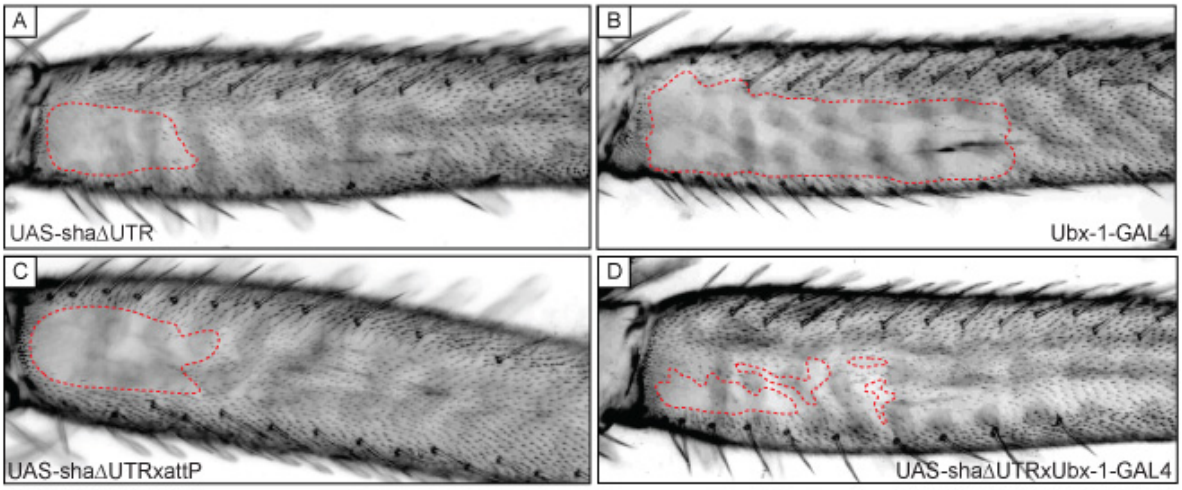


Fig. 11. *Ubx-1* is capable of generating trichomes in the NV. (A-C) Representative legs from control flies. (D) Representative *UAS-shaΔUTRxUbx-1-GAL4* leg showing a patchy trichome pattern in comparison to the effect seen with the full length VT42733 line (N=30) (Fig. 8).

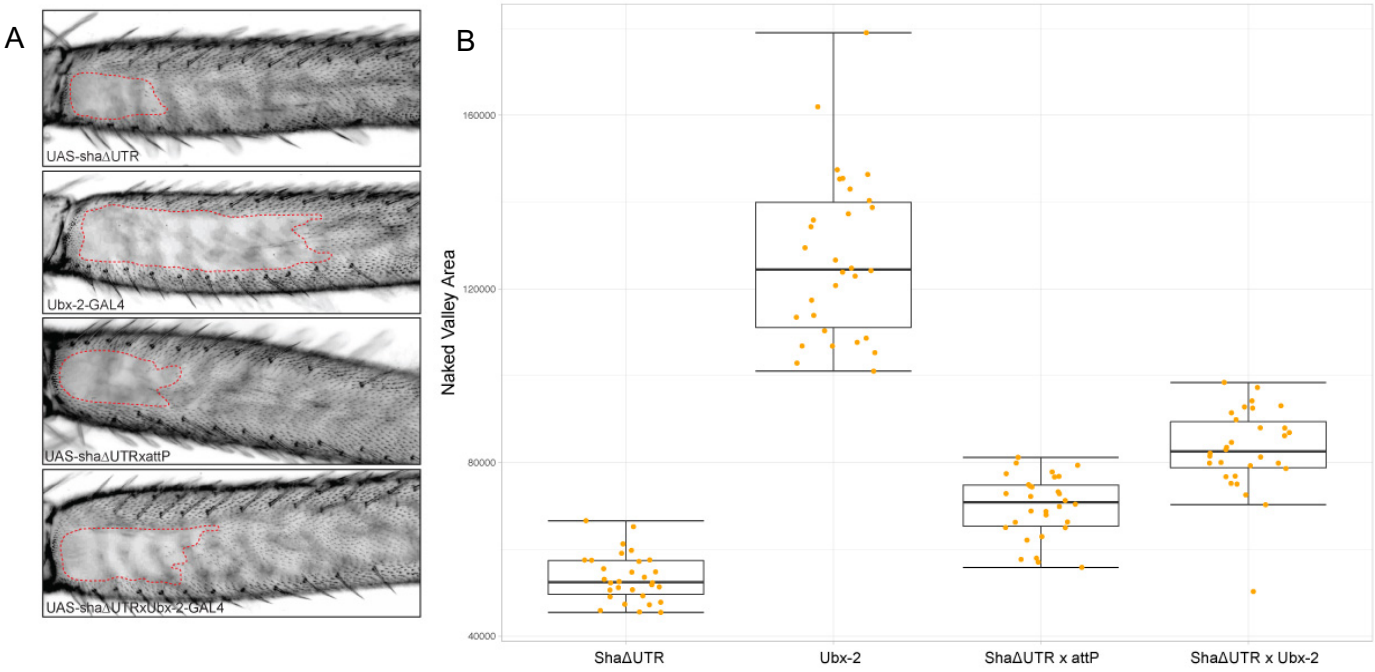


Fig. 12. Ubx-2 has no effect on NV size. (A) Second legs of control lines and *Ubx-2xUAS-ShaΔUTR* showing representative NV sizes. (B) NV area measurements from 30 individuals from each line. The driver line *Ubx-2* has no effect on the size of the NV compared to controls. Data was tested using Shapiro-Wilk followed by an ANOVA.

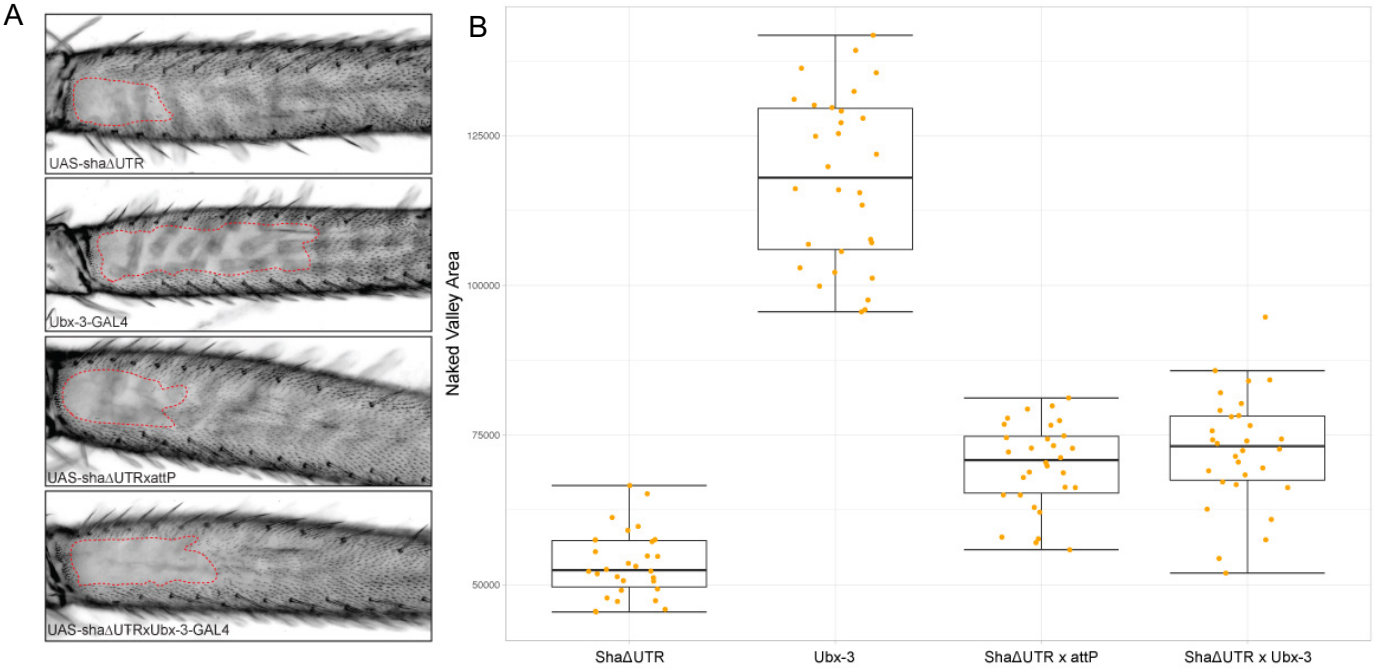


Fig. 13. *Ubx-3* is not capable of producing trichomes in the NV. (A) Second legs of control lines and *Ubx-3x UAS-ShaΔUTR* showing representative NV sizes. (B) NV area measurements from 30 individuals from each line. The driver line *Ubx-3* has no effect on the size of the NV compared to controls. Data was tested using Shapiro-Wilk followed by an ANOVA.

2.3.5 Analysis of larger fragments encompassing the *Ubx* enhancer suggest the presence of potential insulator/repressor sites

There has been some recent speculation that concentrating on the minimal enhancer region may miss important aspects of enhancer logic and function (Halfon, 2019). Therefore, I decided to also make larger constructs across putative *Ubx* enhancer region, which combine lines VT42732, VT42733 and VT42734 (Fig. 14). I hypothesised that this may help to determine if aspects of the expression of the putative enhancer line VT42733 are ectopic perhaps as a result of using small regions missing regulatory sequences. It may also help to explain why VT42734 appears to be able to reduce the size of the NV but not as well as VT42733. Moreover, for further studies it is important to define the boundaries of enhancers.

VT32732+VT32733, which contains ~4.1 kb, drives expression in the legs and the developing antenna of the pupae (Fig. 14). Unlike the other two combination lines this driver also generated expression in localised dots on the periphery of the abdomen (Fig. 14). VT42733+VT42734, which contains ~3.9 kb, also shows expression in the legs and the antenna (Fig. 14), but unlike the VT32732+VT32733 this leg expression extended to more distal leg segments (Fig. 14). The final larger combination line, VT42732+VT42733+VT42734, which contains ~5.6 kb was also able to drive leg and antennal expression (Fig. 14). However, what was striking about the pattern of these combination lines is that they seemed to show very similar expression to one another. Furthermore, by increasing the size of the regions tested the expression appears to be reduced in comparison to the Vienna lines tested (Fig 6).

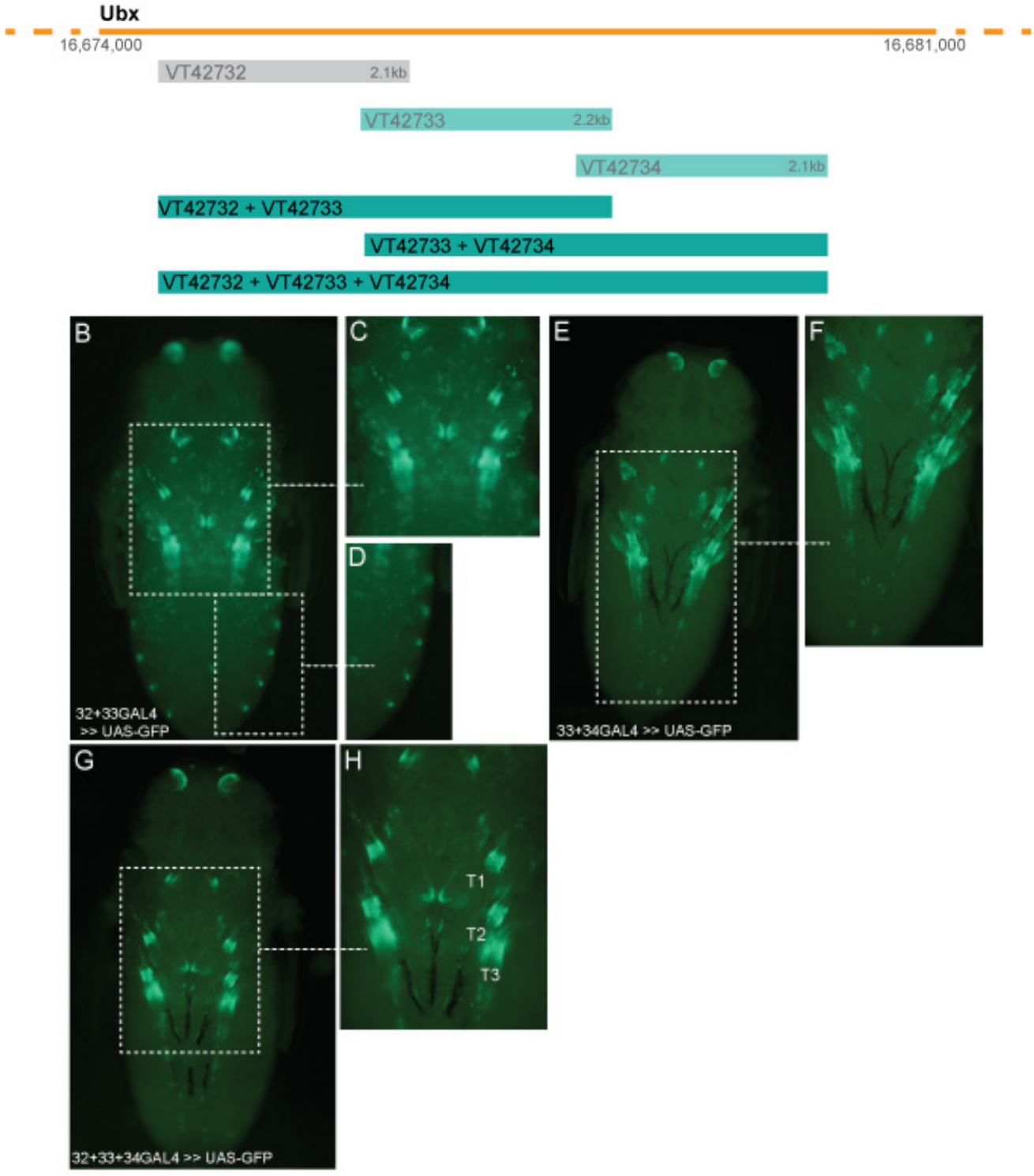


Fig. 14. Expression driven by combinations of VT42732, VT42733 and VT42734. (A) Location of the combination lines. (B-H) Expression patterns driven by the combination lines at 24 hAPF. The expression driven from VT42732+VT42733 can be seen in the developing legs, antennae and also at the periphery of the abdomen in localised spots (B-D). VT42733+VT42734 also drives expression in the developing legs but this seems to extend more distally than the VT42732+VT42733. This line also expresses in the developing antennae (E-F). VT42732+VT42733+VT42734 also drives expression throughout the developing legs of the pupae (G-H). T1 – first leg, T2 – second leg, T3 – third leg.

2.3.6 Functional testing

The combination lines were again crossed to UAS-*shaΔUTR* and the effect on the trichome pattern of adults was measured.

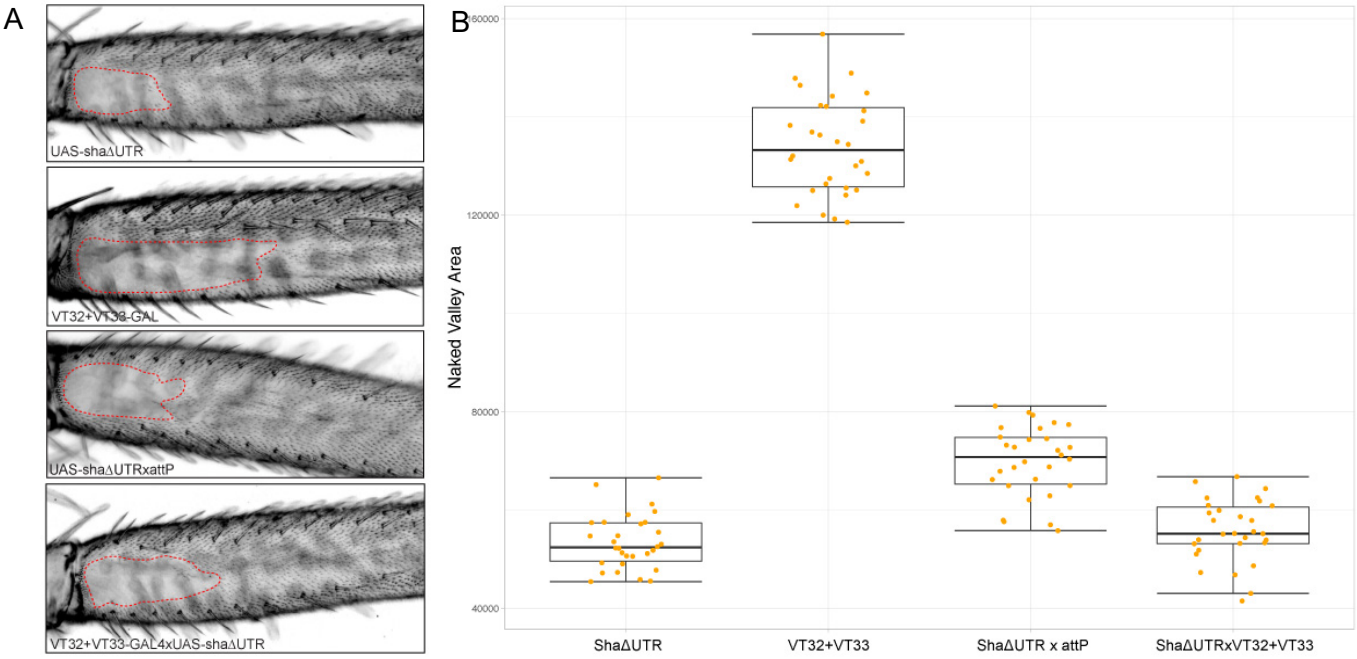


Fig. 15. Driver line VT42732+VT42733 has no effect on NV size. (A) Second legs of control lines and VT42732+VT42733xshaΔUTR (B) measurement of NV area for 30 individuals for each line tested. The driver line has no effect on the size of NV compared to parental controls, however is significantly smaller than the control cross, shaΔUTR x attP (*p-value* < 0.001). Data was tested using Shapiro-Wilk followed by an ANOVA. Significance between groups was tested using Tukey's multiple comparison test.

Curiously, despite containing VT42733, the enhancer capable of substantially reducing the size of the NV, VT42732+VT42733 did not have a significant effect on NV size (Fig. 15)

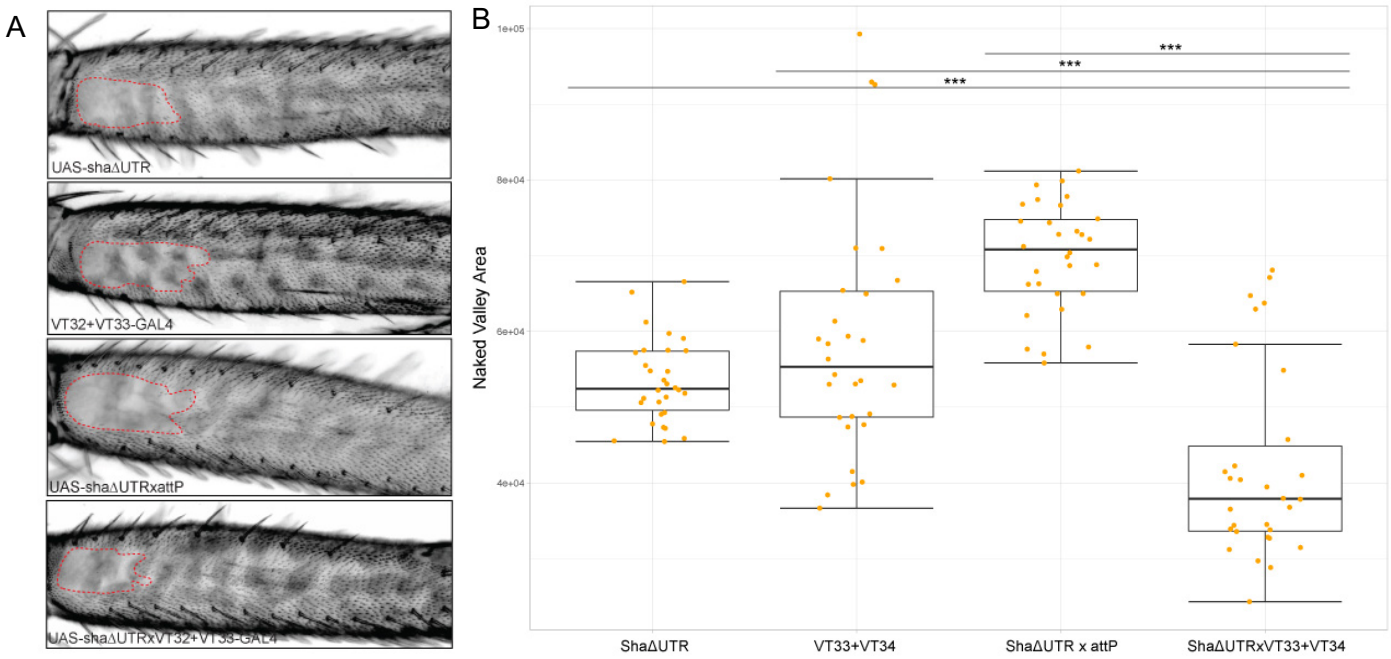


Fig. 16. Combination line VT42733+VT42734 significantly reduces the size of the NV. (A) T2 legs from all lines showing representative NV sizes. (B) NV area measurements from 30 individuals for each line. Line VT42733+VT42734 produces a significant decrease in the size of the NV in comparison to all control lines (*** *p-value* < 0.001). It should also be noted that the control cross displays a significantly larger NV than that of the parents which is not an expected result (*p-value* < 0.005) Data was tested using Shapiro-Wilk followed by an ANOVA. Significance between groups was tested using Tukey's multiple comparison test.

The combination line containing VT42733+VT42734 significantly reduced the size of the NV compared to all controls (Fig. 16). However, when combining all of the VT-GAL4 lines, VT2732+VT42733+VT42734 and crossing to the trichome activating line, there was no significant effect on the size of the NV - the progeny displayed an intermediate size of NV with respect to the parental lines (Fig. 17).

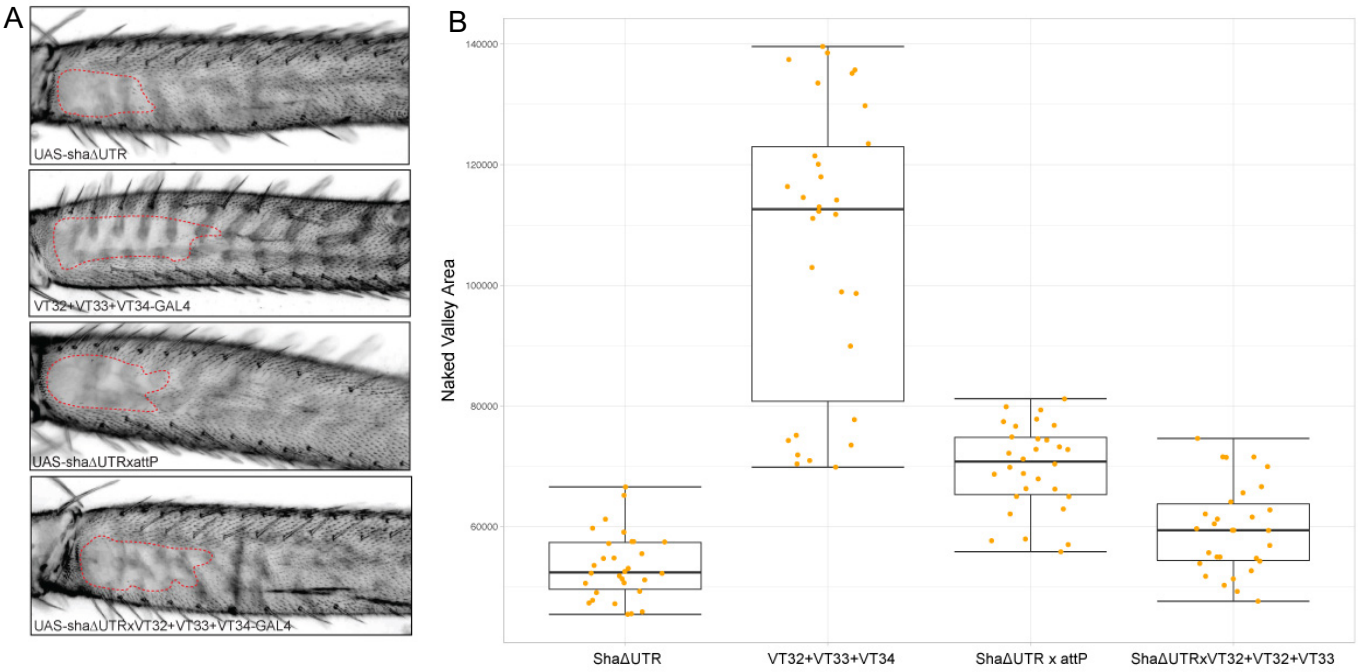


Fig. 17. The combination line VT42732+VT42733+VT42734 has no effect on NV size. (A) T2 legs of control lines and VT42732+VT42733+VT42734x*shaΔUTR* showing representative NV sizes (B) measurement of the NV area from 30 individuals for each line tested. The driver line VT42732+VT42733+VT42734 has no effect on the size of the NV compared to controls. Data was tested using Shapiro-Wilk followed by an ANOVA.

2.3.7 Summary of results from each driver line tested

In total 10 driver were lines tested in this investigation, 8 of these lines drove GFP expression in T2 pupal legs (Table. 1). When functionally tested for an effect on trichomes in the NV, VT42733 was able to induce the development of trichomes throughout the NV region (Table 1). The driver lines VT42734 and VT42733+VT42734 were able to significantly reduce the size of the NV in comparison to controls (Table 1). Furthermore, *Ubx-1* was able to induce the formation of trichomes throughout the full NV region but in a patchy manner (Table. 1)

Table 1. Summary of results from functional testing for all constructs. The constructs used to analyse the NV enhancer are listed. Green boxes show the constructs that express in the T2 pupal leg. The final column shows a representation of the effect that construct has on the trichome pattern in the NV. Grey boxes indicate there was no effect in the NV, Black dashes in the grey boxes represents the trichomes in the NV when the driver line is functionally tested using the trichome activating line.

Construct	Length	Expression in T2	Effect on NV
VT42732	~2.1 kb		
VT42733	~2.2 kb		
VT42734	~2.1 kb		
Ubx-1	780 bp		
Ubx-2	780 bp		
Ubx-3	780 bp		
Ubx-5	826 bp		
VT42732+VT42734	~4.1 kb		
VT42733+VT42734	~3.9 kb		
VT42732+VT42733+VT42734	~5.7 kb		

The chromatin profile for the three driver lines that had an effect on the NV shows that they are several open peaks of chromatin in these regions. The largest peak is located in the overlap between driver lines VT42733 and VT42734 (Fig. 18).

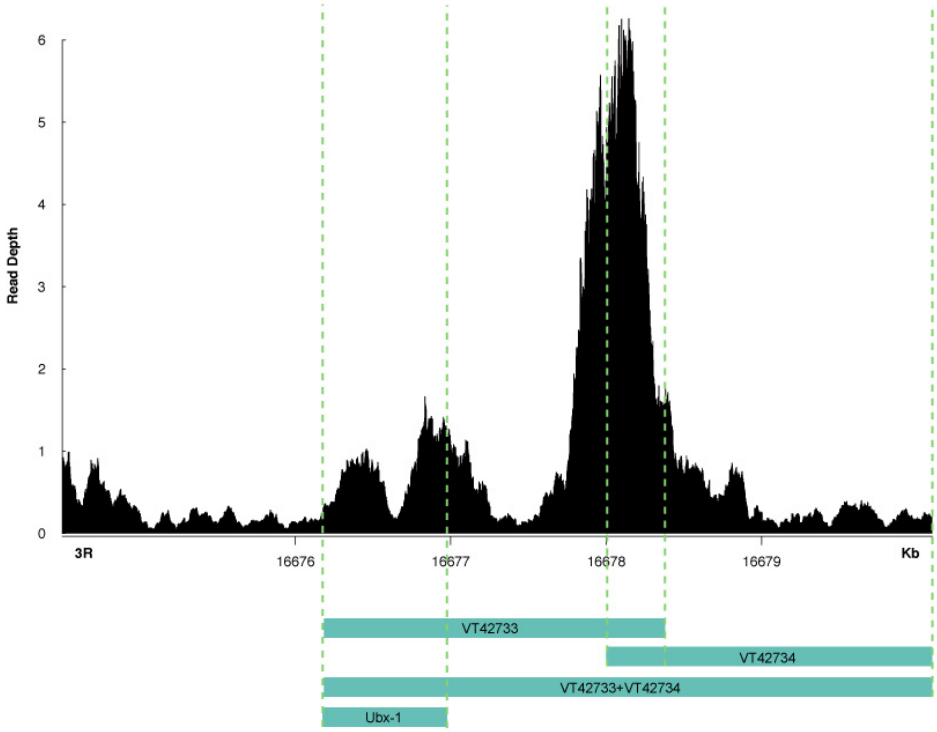


Fig. 18. Summary of the position and chromatin profile of the driver lines tested in this study that had an effect on trichome development in the NV. Above is the ATAC-seq profile in this region. Boundaries of the driver lines and the overlap between VT47233 and VT42734 are shown with green dashed lines.

2.3.8 Ubx protein trap

The expression of Ubx in pupae at 24 hAPF was not previously described, which means I could not compare the expression patterns driven by the enhancer lines tested to that of endogenous Ubx in pupae to determine if the expression from reporter constructs reflected endogenous expression or was ectopic in places. However, recently, a fly line was created by Domsch et al. (2019) in which Ubx is endogenously tagged with GFP.

I therefore looked at the expression of Ubx-GFP in whole pupae at 24 hAPF (Fig. 19). The expression driven by this protein trap is very weak but I did observe fluorescence in the T3 leg at this time point (Fig. 19). However, although there might also have been expression in the other legs and the antennae, like with reporter

constructs, since the signal was relatively weak with respect to the background auto-fluorescence, I was unable to be certain that Ubx is also expressed in these other tissues.

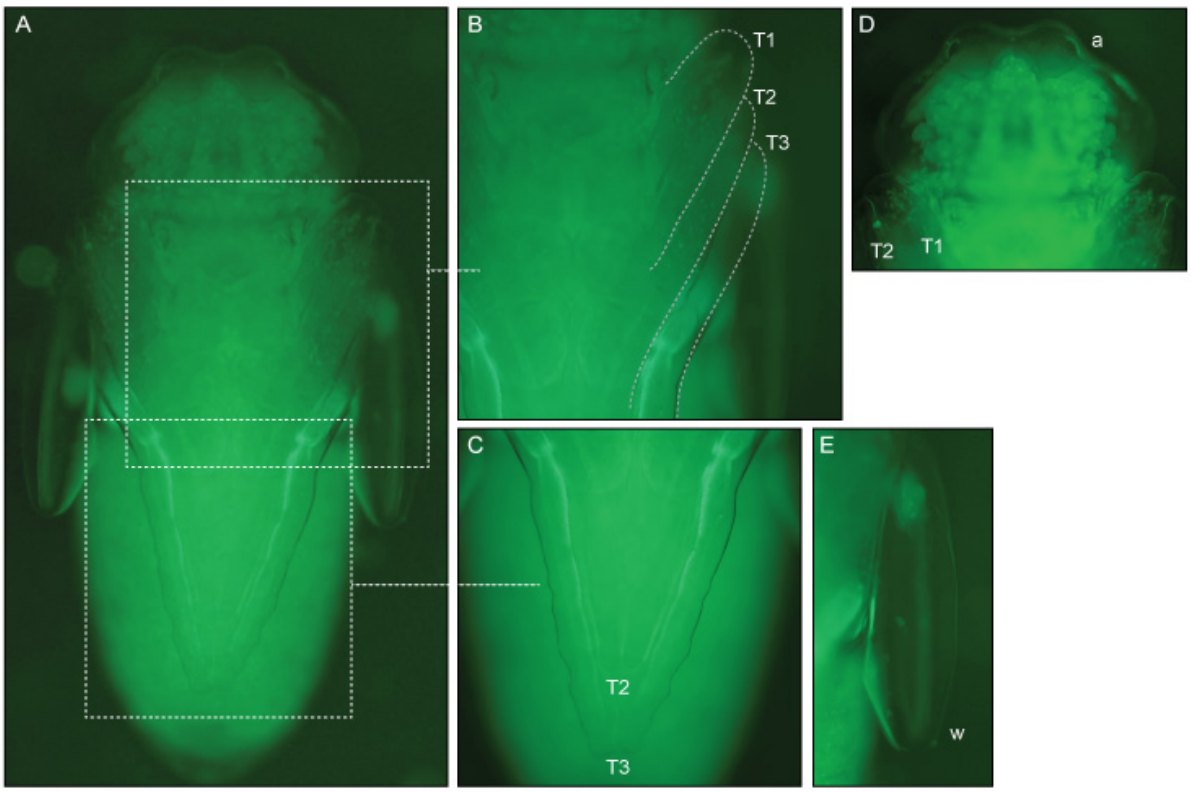


Fig. 19. Expression of endogenously tagged Ubx in pupae at 24 hAPF. (A) Expression can clearly be seen in the third leg. (B, C) Higher magnification views of the three pairs of legs, expression in T3 is the strongest and while there may be weak expression driving in the other legs this is difficult to conclude because the expression of protein trap is weak. (D) View of a pupal head. It appears there might be some weak expression in the antennae, but again this is difficult to conclude. (E) Wing of a pupae, which shows there may be some weak expression in the hinge of the wing but again requires further investigation. T1 – first leg, T2 – second leg, T3 – third leg, a – antennae, w – wing.

2.3.9 Transcription factor binding sites in the Ubx enhancer

To further characterise the putative *Ubx* NV enhancer I studied which transcription factors might bind to the VT42733 driver line. This will help us to understand where Ubx sits in the GRN and potentially identify novel factors that contribute to trichome patterning.

To do this, a bioinformatic approach was taken combining the results from the JASPAR database (Khan et al., 2018) and the RNA-Seq data for the pupal legs (Kittelmann et al., 2018) i.e. TFBS that are present in the *Ubx* NV enhancer and which of those TFs are expressed specifically in the T2 leg. The resulting candidates are shown in Table 2. In total, binding sites were identified for 36 TFs. These candidate TFs then were surveyed for their known biological functions using FlyBase (Thurmond et al., 2019, FlyBase 2.0) to search for those that have a known role in leg development. Out of the 36 TFs only 10 have known roles in leg development and *Dll, C15, ap, hth* and *svb/ovo* have already been shown to physically interact with *Ubx* via a number of methods including fluorescent microscopy and yeast-two-hybrid (Y2H) screens (Baeza et al., 2015; Bischof et al., 2018; Bryantsev et al., 2012). However, this includes predictions of protein-protein interactions, and so it still remains to be tested in the future if any of these TFs also bind to *Ubx* cis-regulatory DNA.

Table 2: List of candidate TFs with binding sites in the VT42733 enhancer. The JASPAR TF database was used to generate a list of putative TFs that have binding sites in the NV enhancer, this list was then cross-referenced against RNA-seq data for the second legs (Kittlemann, Buffry et al., 2018). Only genes that have a predicted binding sites in the enhancer and a FKPM above 1 are listed below. TFs highlighted in grey have a previously described role in leg/leg disc development according to data on FlyBase. Genes highlighted in yellow do not currently have enough information to determine a role

TF	FKPM	FlyBase ID	Number of sites in VT42733
unc-4	16.43	FBgn0024184	1
so	9.05	FBgn0003460	2
B-H2	10.40	FBgn0004854	6
Ubx	12.59	FBgn0003944	35
B-H1	10.40	FBgn0011758	7
mirr	29.82	FBgn0014343	19
dve	10.28	FBgn0020307	2
onecut	2.75	FBgn0028996	4
caup	29.82	FBgn0015919	16
ara	29.82	FBgn0015904	24
NK7.1	16.35	FBgn0024321	8
vvl	30.97	FBgn0086680	18
Dll	119.45	FBgn0000157	7
C15	5.38	FBgn0004863	6
CG32352	18.02	FBgn0052532	3
ct	5.13	FBgn0004198	2
ems	20.25	FBgn0000576	9
en	29.48	FBgn0000577	1
Deaf1	10.21	FBgn0013799	14
E5	25.32	FBgn0008646	7
Schlank	26.42	FBgn0040918	2
Rx	1.15	FBgn0020617	1
Scr	9.74	FBgn0003339	3
Dr	5.06	FBgn0000492	1
ap	8.39	FBgn0267978	2

repo	7.5	FBgn0011701	1
inv	39.76	FBgn0001269	1
nub	21.54	FBgn0085424	7
Trl	179.59	FBgn0013263	1
hth	4.5	FBgn0001235	1
cnc	11.27	FBgn0262975	1
maf	139.45	FBgn0034534	1
btn	8.57	FBgn0014949	2
Eip74ef	17.17	FBgn0000567	1
pan	10.29	FBgn0085432	5
ovo	23.52	FBgn0003028	1

2.3.10 Alignment and TFBS prediction of the NV enhancer

Due to the limited number of TF profiles available in the JASPAR database I also sought to use additional databases to identify further putative TFs that might bind to *Ubx* VT42733. Therefore, I used the MEME suite, which searches for novel motifs that recur in fixed-length patterns in the users sequence (Bailey et al., 2009).

To do this I first created a multiple species alignment of the VT42733 region with the *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. mauritiana* and *D. yakuba* sequences using T-coffee alignment (supplementary fig 2). This alignment was then submitted to the MEME web server and I used TomTom (Gupta et al., 2007) on the resulting motifs to search for similarity to known motifs in several *Drosophila* motif databases (Fig. 20). The 10 motifs found by MEME were present in high numbers (23+) across the five species alignment (Fig. 20). I then cross-referenced the predicted binding TFs found with the leg RNA-seq data (Kittelmann et al., 2018) to identify those that are expressed in T2 pupal legs (Fig. 20). Surprisingly I did not find

any overlap with the TFs identified using the JASPAR database, but this may be explained by both the limited number of profiles on the JASPAR database, and that the MEME search was limited to output 10 motifs. This approach identified some interesting TFs, including Pho which is known to work with Ubx to stabilise lineage choice of cells throughout developmental stages (Domsch et al., 2019) as well as other TFs for future testing.

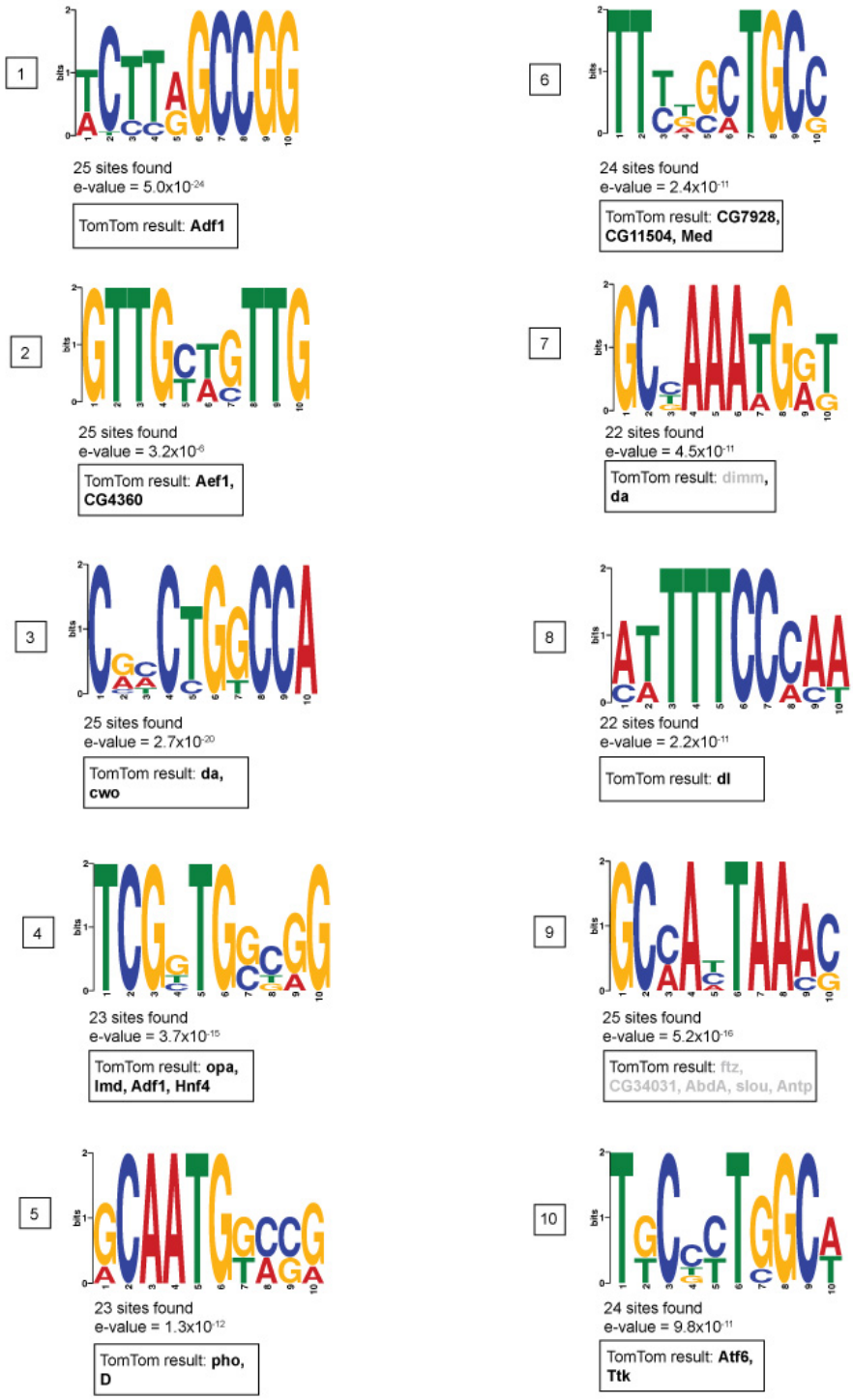


Fig. 20. Motifs in *Ubx* VT42733 multiple species alignment as found by MEME (Bailey et al., 2009). Ten motifs are shown along with the number of sites found and the e-value of each motif, which represents an estimate of the number of motifs (with the same width and number of occurrences) that would occur if random sequences were used. Underneath each motif is a box showing the result when the motif is submitted to TomTom (Gupta et al., 2007) ,which looks for similarity to known TFs from other *Drosophila* TF databases. The top five TFs from the TomTom search are shown: where multiple hits to the same TF occurred, replicates were counted in the top 5. E.g. for motif 1, the top 5 motifs were different configurations of the binding site for the TF Adf1. The TFs shown in grey are those that are not expressed in T2 pupal legs according to the RNA-seq data (Kittelmann et al., 2018)

2.4 Discussion

By using the data obtained from the ATAC-seq analysis I have carried out focussed testing of the *Ubx* locus for the presence of a previously hypothesised NV enhancer (Davis et al., 2007). Only testing regions of open chromatin that are accessible in the window of time when the trichomes are specified greatly reduced the workload needed to screen for an enhancer element, and also the need for complicated genetic methods which may ultimately miss the enhancer. Therefore, this investigation has highlighted the utility of ATAC-seq data when screening for cis-regulatory elements.

2.4.1 Dissection of the NV enhancer region reveals a complex regulatory landscape

While four out of the nine lines tested from existing enhancer trap collections produced expression in the developing pupal legs, none of the lines drove expression specific to just the NV (Fig. 6, 10, 14). However using functional tests I was able to find a region of *Ubx* that is able to drive expression in NV cells at the right developmental time point to induce the formation of trichomes. Interestingly, I also identified another driver, VT42734, which does express in T2 pupal legs and reduces the size of the NV in comparison to the parental controls but this effect was not significant in comparison to the control cross (Fig. 9). VT42733 and VT42734 overlap by 400 bp so it may be that VT42734 is able to carry out part of the function of VT42733 due to this overlap or that the former has a few additional binding sites for TFs that are active in the NV (Fig. 16). By comparing their expression patterns, it does appear that the expression of VT42734 overlaps with that of VT42733 which may provide evidence for sharing the enhancer function in the NV. A significant decrease

in the size of the NV was also seen with the combination line VT42733+VT42734, this suggests that VT42733 is the NV enhancer (hereafter referred to as NV enhancer) and that the small overlap with VT42734 captures some of the NV enhancer functionality. This would also explain why the combination of VT42733 and VT42734 significantly reduces the size of the NV, as this driver contains the full region of the NV enhancer.

2.4.2 Larger reporter constructs revealed further complexity

It appears that when adding the DNA from VT42732 into these combination lines it represses the expression driven by the NV enhancer, suggesting that this region contains cis-regulatory elements that repress enhancer-promoter interactions. Moreover, the finding that the larger combination lines tested showed reduced expression in comparison to the NV enhancer line could also be explained by the presence of binding sites for repressors in region VT42732 which would not only reduce expression, but effect the functionality of the enhancer in the NV.

It is well known that *Ubx* regulation is highly complex with many additional elements that repress or silence transcription (Maeda and Karch, 2006) for example, cp190 is critical for the correct regulation of the bithorax complex (Savitsky et al., 2016). Despite several studies showing the presence of insulators and silencers in the bithorax complex (Gruzdeva et al., 2005; Iampietro et al., 2008; Kyrchanova et al., 2015; Savitsky et al., 2016; Zhou et al., 1996), a thorough analysis of existing data did not reveal the presence of known insulators in the region of interest surrounding and within the NV enhancer, however this does not mean an insulator is not present. A more complete analysis of this region of DNA at the developmental time point of

interest is needed to fully conclude whether insulators are present.

To further characterise the novel NV enhancer, I made smaller constructs of both the VT32733 and VT42734 (Fig. 10). For *Ubx-1* and *Ubx-2* I saw a reduced expression pattern compared to the individual regions, which seems spatially restricted to the legs and antenna. Interestingly, when the *Ubx-1* line was functionally tested for the ability to produce trichomes in the NV, it produced a patchy and irregular trichome pattern. Perhaps this smaller construct is carrying out part of the function of the larger NV enhancer region and/or this effect is a result of removing binding sites in adjacent regions so that not all of the required binding sites are present. Based on the observations with the smaller *Ubx* fragments and the combination lines it appears as the whole NV enhancer is required for enhancer function.

2.4.3 The ATAC-seq profile provides insights into enhancer organisation at this locus

When looking at the ATAC-seq profile for the region analysed in this study (Fig. 18) there is also a large peak of open chromatin that is shared by the NV enhancer and VT42734 which might help to further explain the partial effect seen when crossing VT42734 to the trichome activator (Fig. 9). Moreover, there is also a peak of chromatin at the start of the NV enhancer line which is covered by the *Ubx*-1 fragment (Fig. 18). Therefore it could be speculated that both of these accessible chromatin regions are needed for the activity of the full NV enhancer, which would also explain the patchy trichome pattern effect seen with *Ubx-1* and the reduction in the NV size with VT42733+VT42734 when both are crossed to the trichome activating line. This

would be consistent with several studies that have highlighted that multiple enhancers with varying degrees of functional redundancy can contribute to the expression of a gene (Cannavo et al., 2016; Frankel et al., 2010; Kalay et al., 2019; Letelier et al., 2018; Osterwalder et al., 2018; Perry et al., 2010).

2.4.4 Ectopic expression?

My analysis of post-embryonic *Ubx* enhancers also suggested that the NV enhancer drives expression in several tissues in addition to the T2 legs. To determine if this expression is ectopic I analysed the expression of a protein trap of Ubx (Fig. 19). The protein trap is expressed in the third leg of the pupae, which is where we would expect Ubx to express (Fig. 19). However, the expression from the protein trap was weak and required a high laser power to visualise, so I cannot say for certain that it is absent from other tissues. Therefore, I was unable to absolutely determine whether expression from the reporter constructs is ectopic. It could also be that the enhancer lines do not actually recapitulate endogenous Ubx expression because they exclude epigenetic features that normally repress 'ectopic' Ubx expression in other tissues. A further explanation could be that this enhancer is in fact pleiotropic and acts in several other tissues as well as the legs. Indeed, recently it has become clear that enhancer pleiotropy is more common than previously thought. At the *svb* locus, for example, the majority of the enhancers have been shown to be pleiotropic (Preger-Ben Noon et al., 2018). It is thought that this could be achieved through two different mechanisms - either through the use of common TF binding sites or through using independent TF sites for each tissue type (Preger-Ben Noon et al., 2018). However, the promiscuous expression of the NV enhancer could be the result of using reporter constructs and is

in fact ectopic in some of the tissues. Interestingly, previous studies on the post-embryonic leg enhancers of *Scr* screened using Janelia GAL4 lines did not find any regions (other than those already identified) that drove endogenous *Scr* expression and several of the lines tested drove ectopic expression in the developing legs, in comparison to antibody stains for *Scr* in the same tissues (Eksi et al., 2018). This suggests ectopic expression may be a feature of post-embryonic regulatory regions or are a consequence of using non-endogenous reporter constructs (Casas-Tintó, Arnés and Ferrús, 2017).

The expression patterns driven by the other smaller enhancer fragments were interesting, especially the dorsal abdominal pattern seen for *Ubx3* as this expression was not seen in any of the larger lines tested, however this phenomenon has been reported in other studies. For example, in a recent analysis of pupal enhancers at the *yellow* locus in *Drosophila*, the authors noticed that several smaller enhancer elements tested drove expression patterns not seen in reporter genes containing larger cis-regulatory sequences (Kalay et al., 2019). This is very similar to what we have seen in this investigation and may be a common theme with post-embryonic enhancers or again, analysis of enhancer fragments using reporter constructs. i.e. small regions that exclude functional repressor sites that are present in larger regions.

The complex expression patterns driven by both multiple and overlapping fragments at the *Ubx* locus, in addition to the various effects seen when functionally testing these driver lines, may suggest that this particular region does not contain a classical modular enhancer, and may involve multiple regions working together to output the NV phenotype.

2.4.5 Identification of TFBS in the NV enhancer

For the identification of transcription factors that may bind the NV enhancer, the bioinformatic search has highlighted some interesting candidates. For example, *distal-less* (*Dll*) has a well characterised role in leg development and additionally a known interaction with the enhancer of another Hox gene, *Scr*, in appendage patterning (Eksi et al., 2018). Further interesting candidates are the B-H1 and B-H2 genes which also require *Dll* expression to specify leg segments (Kojima, Sato and Saigo, 2000).

Unfortunately, the JASPAR database does not contain an exhaustive list of the characterised TFs in the *Drosophila* genome. Therefore to obtain a more complete understanding of the TFs that might bind the *Ubx* NV enhancer, I decided to use the MEME suite tools (Bailey et al., 2009) to assess the presence of TFs in *Ubx* VT42733 while taking sequence conservation into account. I did not find any overlap with the TFs identified using the JASPAR database, but this may be due to the limited number of profiles on the JASPAR database, and that the MEME search was limited to output 10 motifs. The MEME search revealed the presence of many additional TFs (Fig. 20), including Pho, which is a Polycomb Group protein that is maintained at Ubx targeted genomic regions to stabilise lineage choice (Domsch et al., 2019). This interaction has been suggested as a reason to why the Hox code is maintained throughout the lifecycle (Domsch et al., 2019). Further functional investigation into the putative binding of TFs is needed, perhaps using a yeast one hybrid (Y1H) assay (Hens, Feuz and Deplancke, 2012). This would allow cross referencing of candidates from the Y1H with both the JASPAR and RNA-seq to filter candidates for functional testing.

Chapter 3

Analysis of the function and regulation of leg trichome genes

3.1 Introduction

Transcription factors (TFs) are recruited to enhancers to regulate transcription through binding to specific DNA motifs. Transcriptional activation by enhancers typically requires the cooperative binding of multiple TFs to a range of motifs (reviewed in Long, Prescott and Wysocka, 2016). TFs generally recognise degenerate motifs around 6-12 bp long, and given this low specificity it is not fully understood which TFs individually, or in combination can bind to a given enhancer, also referred to as TF occupancy, to generate robust and controlled patterns of gene expression (reviewed in Spitz and Furlong, 2012). Indeed, these short, degenerate motifs, mean that genome-wide searches for putative TFBS results in thousands of hits, hindering the prediction of *cis*-regulatory elements (Li et al., 2008; Rister and Desplan, 2010; Slattery, Negre and White, 2012). There has been some effort to understand the logic of TF occupancy at certain enhancers in *Drosophila* embryos from the perspective that enhancers that are functionally related (i.e. acting in the same tissues at the same developmental time-point) can frequently share a combination of TFBS (Pennacchio et al., 2007). By using this combination or code of motifs, some studies have been able to successfully predict spatio-temporal cis-regulatory activity in *Drosophila* embryos (Erives and Levine, 2004; Markstein et al., 2004; Zinzen et al., 2009; Menoret et al., 2013).

This investigation aims to study a post-embryonic gene regulatory network to

explore if there is a pattern of TF occupancy for functionally related genes to help understand their cis-regulatory logic and how this compares to embryos, and eventually even predict enhancers.

Due to the vast amount of information already available as outlined in Chapter 1 and the new knowledge gained in Chapter 2, I chose to use the NV trichome GRN to study the cis-regulatory logic of the genes involved. This complements a previous study by Menoret et al. (2013) who sought to identify a cis-regulatory signature or code for Svb-dependent enhancers for genes expressed in trichome bearing cells in the embryonic epidermis. They showed that DNA containing certain motifs has a high chance of being a Svb-responsive enhancer in the embryonic trichome network (Menoret et al., 2013).

Furthermore, as previously discussed (Chapter 2) miR-92a represses the formation of trichomes on the second leg in *Drosophila*, resulting in the NV (Arif et al., 2013). It has already been shown that miR-92a directly post-transcriptionally represses the Svb-target *sha* to block the formation of trichomes (Arif et al., 2013; Schertel et al., 2012). However, when *sha* is over-expressed in the NV, the resulting trichomes are not completely normal (Kittelmann et al., 2018). This indicates that other genes are also required for the formation of normal leg trichomes and therefore that miR-92a potentially targets multiple Svb-dependent target genes in T2 legs (Kittelmann et al., 2018). It is thought that the gene *CG14395* may be one of these target genes, and indeed, this gene has shown to have effects on the growth of trichomes (Kittelmann et al., 2018 and this study). However, it has been postulated that there are several other genes regulated by Svb and miR-92a in legs (Kittelmann and McGregor 2019).

In this chapter I first identified a unique and informative set of genes that are predicted to be activated by Svb to generate trichomes, on one hand, but post-transcriptionally repressed by miR-92a, on the other hand, to form the naked valley – thus together forming a genetic switch for the fine-scale morphology of cells. I then sought to understand the cis-regulatory logic for the transcriptional activation of the genes forming this switch: specifically to verify if putative leg enhancers of these genes (as assayed by ATAC-Seq) are enriched in binding sites for Svb and which other motifs they contain compared to flanking DNA regions. The results of this analysis provide additional insights into structure and functionality of the trichome GRN and compared to what is known about the regulation of larval trichomes, may provide a predictive framework for identifying potential new enhancers in the leg trichome network, and more information about post-embryonic developmental enhancers.

3.2 Methods

3.2.1 Gene identification

To identify potential genes involved in leg trichome development that are both targeted by the miRNA and dependent on *svb*, I initially looked for predicted targets of miR-92a using TargetScanFly (R7.2), a web-based server which predicts the biological targets of microRNAs based on the presence of conserved sites that match the seed region of each miRNA (Ruby et al., 2007; Agarwal et al., 2018). The resulting list of predicted target genes (Table S1) was cross referenced against RNA-seq data for T2 pupal legs at 24 hAPF (Kittelmann et al., 2018) to find out which of the predicted targets are expressed in the legs during the window of developmental time when trichomes are specified. This set of genes was further cross-checked against known Svb targets, as determined by micro-array, in the larvae from Menoret et al

(2013) to produce the final set of genes for analysis (Fig. 21). For further information on each gene please see Table 3..

3.2.2 Accessible chromatin analysis

Coordinates (*Drosophila melanogaster* genome R6.23, annotation dm6) were extracted for each gene either 10 kb upstream and downstream or to the nearest next gene (for coordinates used please refer to Table 3). Using the ATAC-seq data from T2 pupal legs at 24 hAPF (Kittelmann et al., 2018) each of the 12 target genes was analysed for accessible chromatin. Subsequently, FASTA files for any peaks in the ATAC-seq dataset that were called by MACS2 (Kittelmann et al., 2018) and within the coordinates of interest for each gene were extracted.

3.2.3 Motif enrichment using i-*cis*Target

Fasta files containing sequences corresponding to ATAC-seq peaks for the 12 genes of interest (ATAC-seq data generation and MACS analysis from Kittelmann et al. (2018) were converted to Bed format and submitted to i-*cis*Target, a web-based server that uses motif analysis and sequence conservation to produce a list of enriched TF motifs in user provided sequences (Herrmann et al., 2012; Imrichová et al., 2015). Parameters used were as follows: PWM database (version 5.0), minimum fraction overlap 0.4, NES threshold 3.0. Control datasets used in this investigation were DNA regions not called as peaks in ATAC-seq data for the 12 genes of interest and epidermal genes that are not known to be dependent on Svb (Menoret et al., 2013). The results of the i-*cis*Target were then further analysed to locate regions of DNA for each of the 12 genes that may contain a potential enhancer. The i-*cis*Target database

contains candidate regulatory regions (CRRs) spanning the entire *D. melanogaster* non-coding genome (dm3, FlyBase r5.37, 136353 CRRs). The user supplied data is mapped to the CRR database and any CRRs that overlap with the user input set of sequences becomes the foreground dataset (Heintzman et al., 2007; Imrichová et al., 2015). I selected motifs matching to Ovo/Svb and/or Grh and located the position of the CRRs which showed enrichment in these motifs for each of the 12 genes, therefore indicating the presence of a potential enhancer element.

3.2.4 RNAi screen and trichome analysis

A RNAi screen was performed to test if any of the focal genes individually had an effect on trichome morphology. RNAi lines were ordered from VDRC and in most cases they were crossed to the driver line VT057077-GAL4 apart from CrebA which was crossed to a HS-Gal4 driver line (Kittelmann et al., 2018) This line contains a *svb* enhancer region that drives expression throughout the second leg in the time window when trichomes are specified. To measure the length of trichomes, T2 legs were dissected from adult flies and mounted in Hoyer's medium/lactic acid 1:1 and imaged under a Zeiss Axioplan microscope using ProgRes MF cool camera (Jenaoptik, Germany). The length of 10 distal trichomes and 10 trichomes proximal to the NV from 10 different femurs were measured and analysed using ImageJ (Fiji version) software (Schindelin et al., 2012). Subsequent statistical analyses were performed in R-Studio version 1.2.1335 (RStudio Team, 2015).

3.3 Results

3.3.1 Selection of twelve genes expressed in T2 legs, which are putative targets of *miR-92a* and predicted to be dependent on Svb

To investigate the cis-regulatory logic of genes involved in the switch that decides whether leg trichomes or naked cuticle will be generated, a list of 12 candidate genes were identified for further analysis*: shavenoid (sha), CG14395, tarsal-less (tal), CrebA, Rab23, CG5742, CG8303, neyo (neo), forked, shavenbaby (svb), CG4678, multiple wing-hairs (mwh)* (Fig. 21 Table 3) by filtering the predicted *miR-92a* target genes (generated by TargetScan (Agarwal et al., 2018; Ruby et al., 2007) by both leg RNA-seq data (Kittelmann et al., 2018) and known Svb-dependent genes for the production of larval trichomes (Menoret et al., 2013),

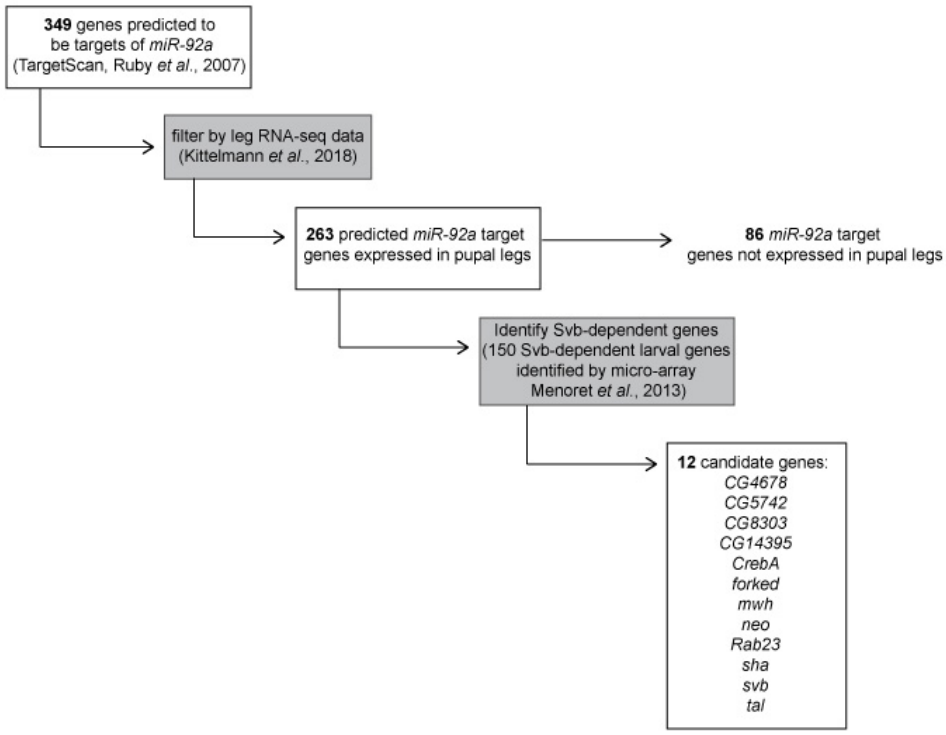


Fig. 21. Process of identifying candidate genes for motif analysis. There were 349 predicted target genes of miR-92a from TargetScan (Ruby et al., 2007;Agarwal et al., 2018) (Table S1) this number was reduced to 263 when filtered by RNA-seq data for the T2 pupal legs (Kittelmann, Buffry et al., 2018). By using data from Menoret et al., (2013) of 150 validated Svb-dependent genes in embryos I selected a list of 12 candidate genes that are both potential targets of *miR-92a* and dependent on *svb*.

Table. 3. List of the 12 genes identified in Fig. 21. The first column corresponds to the gene name/FlyBaseID, the second column identifies a broad function of the gene as defined by FlyBase, the third column shows the coordinates of the gene (dm6 genome annotation), the fourth column shows the coordinates that have been used in this investigation for the (dm6 genome annotation) and finally the last column has references to the papers which first identified the gene as Svb-dependent and/or exhibiting a role in trichome development

Gene name FlyBaseID	Function (FlyBase)	Coordinates (dm6)	Extended coordinates used in analysis (dm6)	Reference predicting interaction with Svb/trichome development
CG4678 FBgn0030778	-Peptide metabolic process -Protein processing	X:16696131 ..16711145	X:16694067 ..16710999	Menoret et al. (2013)
CG5742 FBgn0034304	-Signalling receptor binding -Protein retention in ER lumen -Neurogenesis	2R:18148096 ..18151410	2R:18148000 ..18151700	Menoret et al. (2013)
CG8303 FBgn0034143	-Metabolic processes	2R:16595028 ..16604373	2R:12482505 ..12495505	Menoret et al., (2013)
CG14395 FBgn0038073	ND	3R:12662831 ..12673959	3R:12661000 ..12675000	Menoret et al., (2013); Kittelmann et al., (2018)
CrebA FBgn0004396	-Roles in the secretory pathway -Salivary gland development	3L:15535395 ..15553214	3L:15532000 ..15554000	Abrams and Andrew (2005); Andrew et al. (1997)
forked FBgn0262111	-Actin filament binding -Cuticle pattern formation -Epidermal cell differentiation -Wing hair organisation	X:17232942 ..17280964	X:17227000 ..17283000	Chanut-Delalande et al. (2006); Dickinson and Thatcher (1997); Menoret et al. (2013)
mwh FBgn0264272	-Inhibits actin polymerisation and bundles F-actin -Imaginal disc wing hair organisation, selection and morphogenesis	3L:1200786 ..1232700	3L:1199000 ..1233000	Kaplan et al. (2011); Menoret et al. (2013)

neo FBgn0039704	-Apical constriction -Regulation of embryonic cell shape	3R:29819131 ..29825800	3R:29817800 ..29827300	Menoret et al. (2013)
Rab23 FBgn0037364	-Restriction of actin accumulation -Regulation of number and orientation of hairs on leg and abdominal cuticle	3R:5680054 ..5685434	3R:5679900 ..5685900	Menoret et al. (2013)
sha FBgn0003382	-Promotion of trichome formation -Cuticle pattern formation	2R:11324314 ..11336463	2R:11321000 ..11346000	Arif et al. (2013); Chanut-Delalande et al. (2006)
svb FBgn0003028	-Epidermis differentiation -Cuticle pattern formation -Germ line sex differentiation -Non-sensory hair organisation	X:5048273 ..5070929	X:5038000 ..5085000	Delon, Chanut-Delalande and Payre (2003)
tal FBgn0259730 (1A transcript)	-Actin filament organisation -Regulation of proteolysis -Imaginal disc-derived leg morphogenesis	3R:13813109 ..13814648	3R:13803000 ..13825000	Kondo et al. (2010)

3.3.2 RNAi against potential trichome genes

For the 12 genes identified (Fig. 21), I first decided to do RNAi against as many as possible to assess if the genes by themselves have an effect on trichome growth. A *svb* enhancer (VT057077) identified in (Kittelmann et al., 2018) that drives expression throughout T2 legs at the time when trichomes are specified was used as the GAL4 driver (Stern, 1998) or alternatively HS-GAL4. RNAi was performed for the following

genes*: CG4678, CG14395* (cross and data analysis performed by F. Franke) *CrebA* (cross performed by F. Franke), *forked, mwh, neo* and *Rab23*. The RNAi stock for *CG8303* exhibited a high degree of male sterility and so it was not possible to perform the cross. For the genes *tal* and *CG5742*, no suitable RNAi lines were available. RNAi against *svb* or *sha* was not performed as it was already known that these genes are involved in leg trichome development (Arif et al., 2013; Kittelmann et al., 2018). Out of the 7 genes tested with RNAi, three showed an effect on the length of trichomes. *CG14395* and *CrebA* showed a significant decrease in the length of the trichomes proximal to the NV when crossed to the VT057077 or HS-GAL4 driver line respectively (Fig. 22, 24). A significant decrease in the length of the trichomes both proximal to the NV and on the more distal part of the femur was seen with RNAi against *mwh* using the VT057077 driver (Fig. 26).

3.3.2.1 CG14395 RNAi

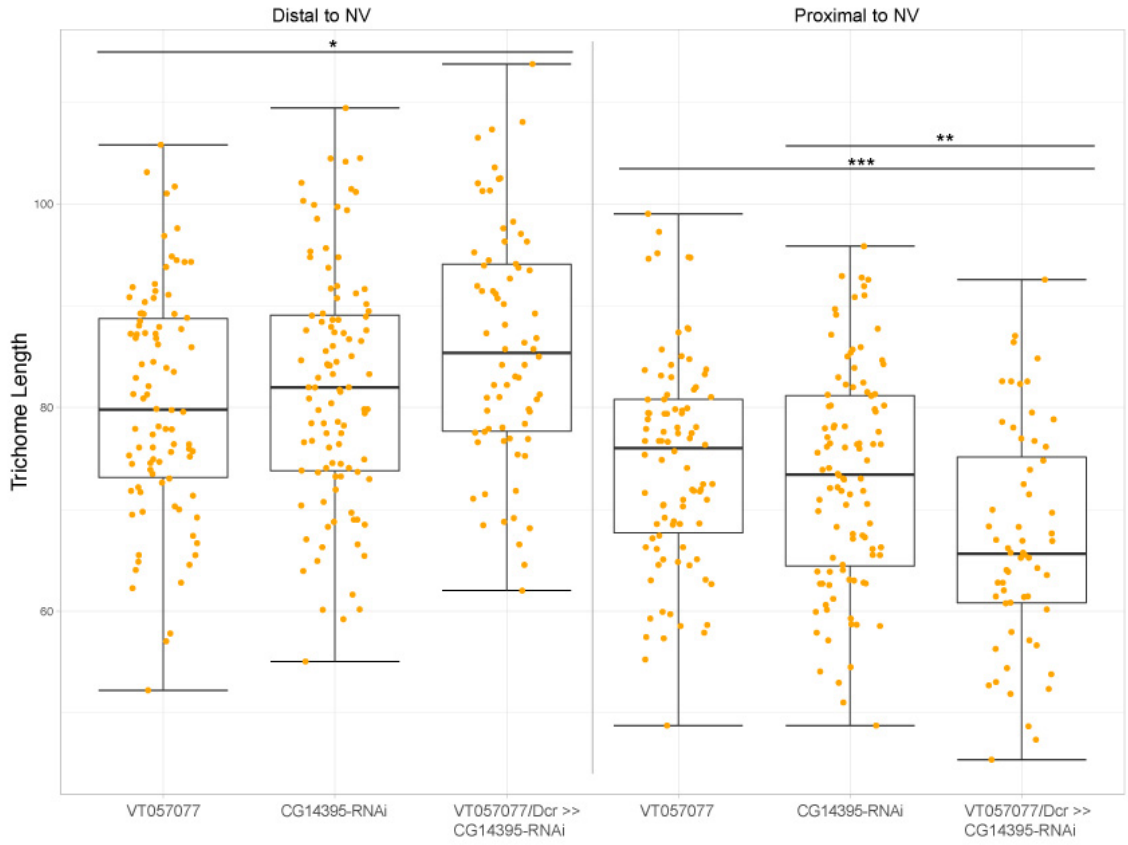


Fig. 22. Analysis of trichome length after RNAi knockdown of *CG14395* in T2 legs using driver line VT057077/Dicer. Boxplots show measurements from the distal part of the femur (left) and near to the NV (right). On the distal part of the femur, there was no significant difference in the length of trichomes between VT057077 parental line and VT057077/Dcr>>*CG14395*. The trichomes for VT057077/Dcr>>*CG14395* were significantly longer than the CG14395-RNAi parental line (* $P < 0.05$). However, if there is an effect on trichome length we would expect the VT057077/Dcr>>*CG14395* F1s to be significantly bigger or smaller than both parental lines. For the trichomes close to the NV, VT057077/Dcr>>*CG14395* had significantly shorter trichomes than both the VT057077 (*** $P < 0.001$) and *CG14395*-RNAi parental lines (** $P < 0.01$). Data was checked for normality using Shapiro-Wilk followed by an ANOVA.Tukey's multiple comparison test was used to measure significance between groups.

Knockdown of *CG14395* in T2 legs showed a significant decrease in the length of the trichomes proximal to the NV in comparison to the parental controls (Fig. 22). For the trichomes on the distal part of the femur, there was a significant difference between the VT057077 driver line and VT057077/Dcr>>*CG14395*, but no difference was seen

between VT057077/Dcr>>*CG14395* and the *CG14395*-RNAi parental line. If there was an effect on trichome development we would expect to see a significant difference to both parental control lines (Fig. 22).

3.3.2.2 CG4678 RNAi

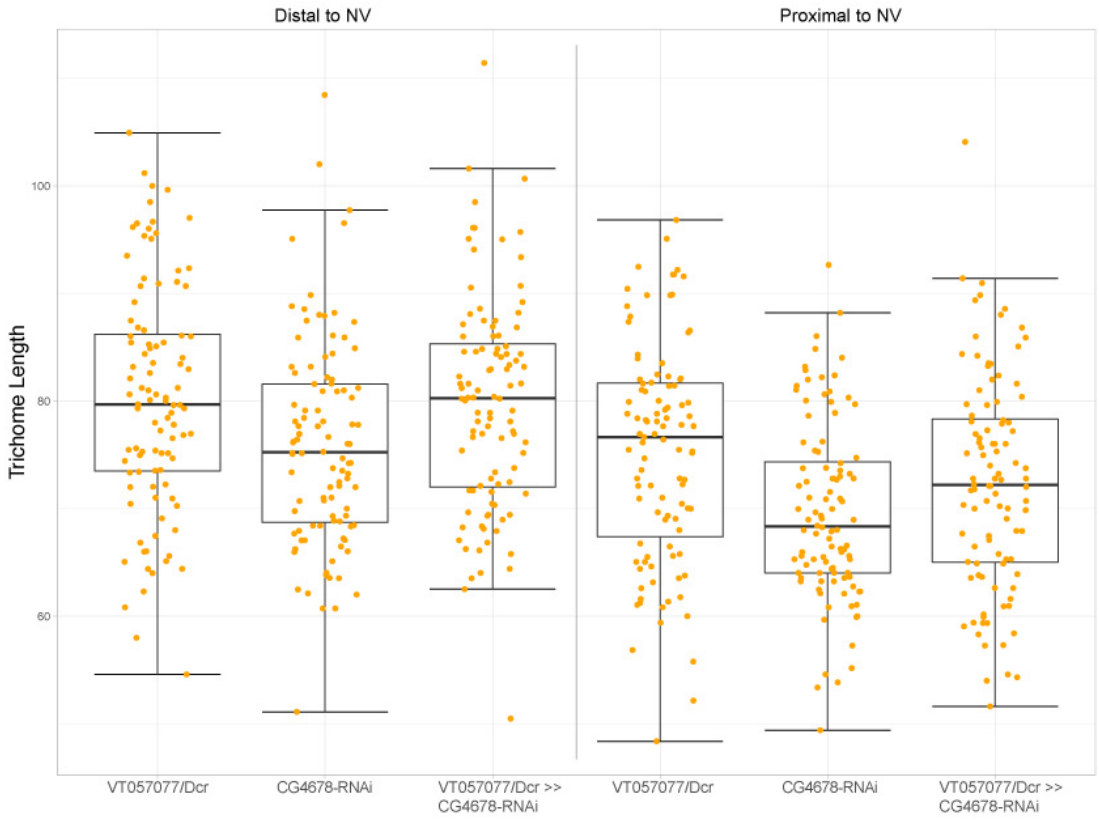


Fig. 23. Analysis of trichome length after knockdown of *CG4678* in T2 legs using VT057077 driver line. Boxplots shows measurements from the distal part of the femur (left) and near to the NV (right). There was no significant effect on the length of the trichomes at either the distal part of the femur or near to the NV. Data was checked for normality using Shapiro-Wilk followed by an ANOVA.Tukey's multiple comparison test was used to measure significance between groups.

RNAi knockdown of *CG4678* in T2 legs did not show any effect on the length of the trichomes either proximal to the NV or on the more distal part of the femur in comparison to the parental controls (Fig. 23)

3.3.2.3 CrebA RNAi

The cross to perform knockdown of *CrebA* was performed by F.Franke using a HS-GAL4 driver because the cross to the VT057077 driver was lethal. There was a highly significant difference in the length of the trichomes proximal to the NV compared to the control line (Fig. 24). No effect was seen on the trichomes on the more distal part of the femur (Fig. 24).

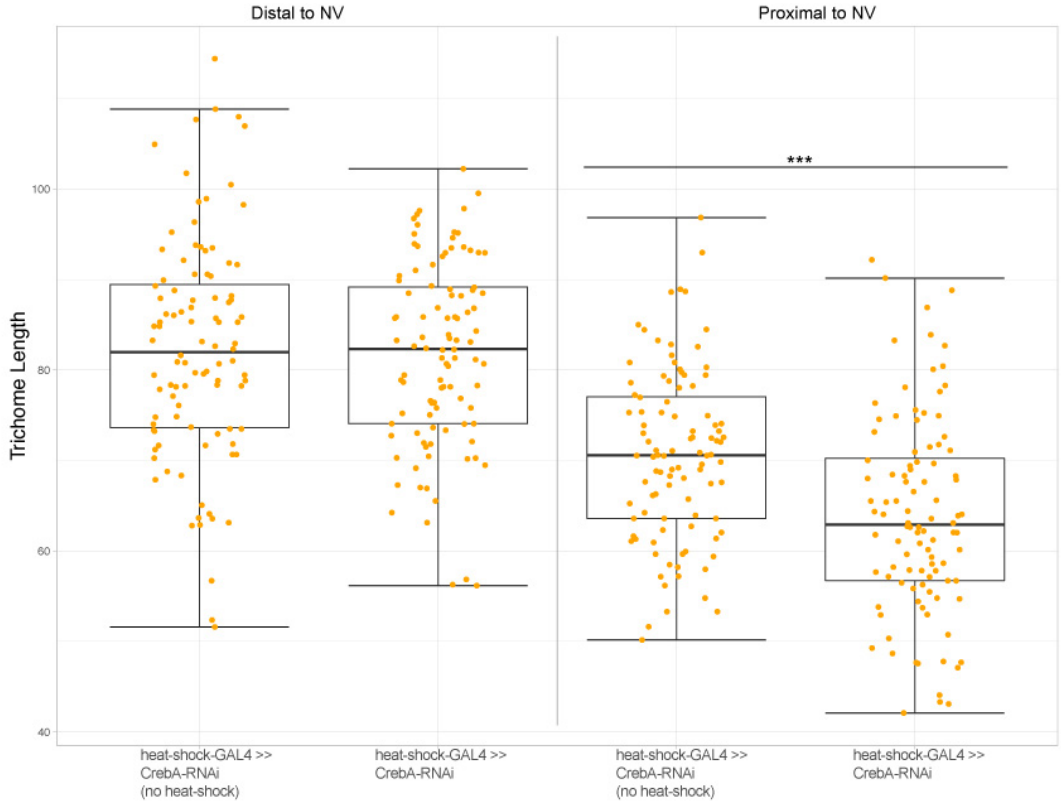


Fig. 24. Analysis of trichome length after RNAi knockdown of *CrebA* in T2 legs using heat-shock-GAL4 driver line, with heat shock performed at 24 hAPF. Boxplot shows measurements from the distal part of the femur (left) and near to the NV (right). There was no effect on the length of the trichomes on the distal part of the femur. The trichomes near to the NV are significantly reduced in length compared to the control line (*** $P < 0.001$). Data was checked for normality using Shapiro-Wilk followed by an ANOVA. Tukeys multiple comparison test was used to measure significance between groups.

3.3.2.4 forked RNAi

Despite evidence showing *forked* has an effect on the morphology of the trichomes on the larval epidermis (Chanut-Delalande et al., 2006) when it was knocked down in the developing leg I did not see any significant effect on the length of the trichomes either proximal to the NV or on the distal part of the femur (Fig. 25).

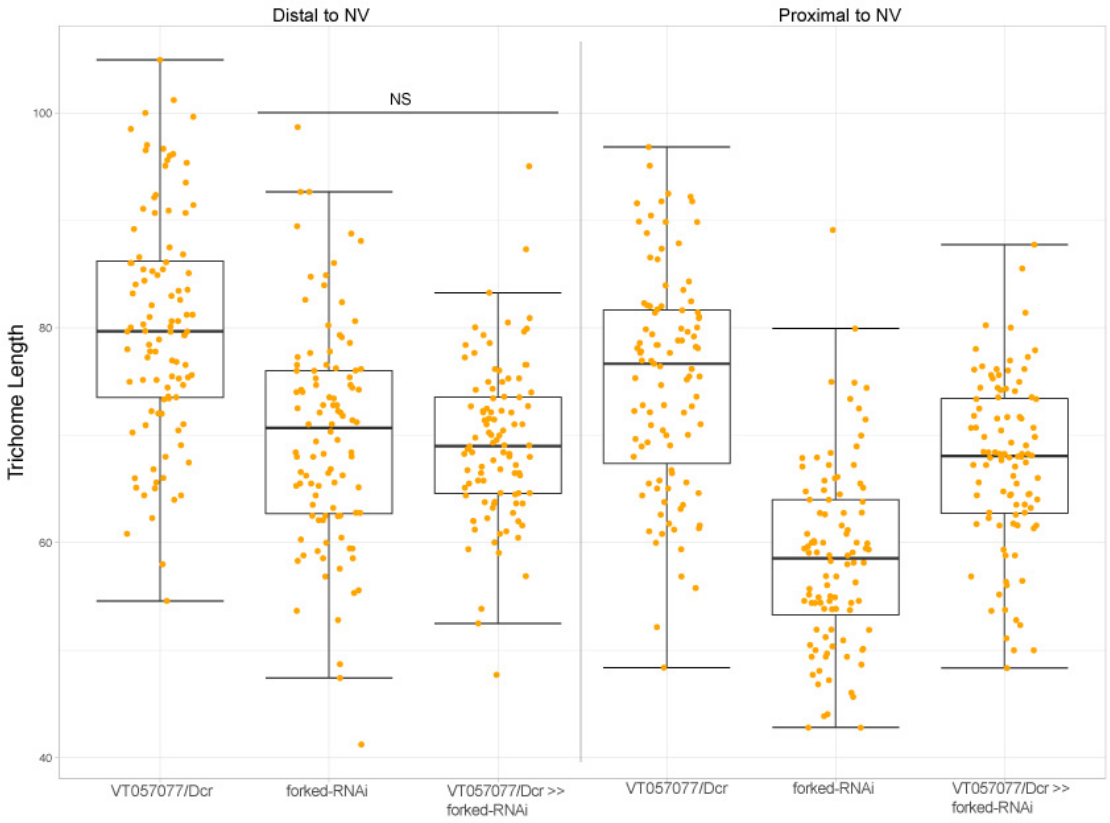


Fig. 25. Analysis of trichome length after RNAi knockdown of *forked* in T2 legs using driver line VT057077/Dicer. Boxplots shows measurements from the distal part of the femur (left) and near to the NV (right). There was no significant effect on the length of the trichomes at either the distal part of the femur or near to the NV. Data was checked for normality using Shapiro-Wilk followed by and ANOVA. Tukey's multiple comparison test was used to measure significance between groups.

3.3.2.5 mwh RNAi

Knockdown of *mwh* in the leg had a dramatic effect on the length of the trichomes both proximal to the NV and on the more distal part of the femur. The length of the trichomes was significantly smaller in comparison to both control lines in both areas of the leg (Fig. 26)

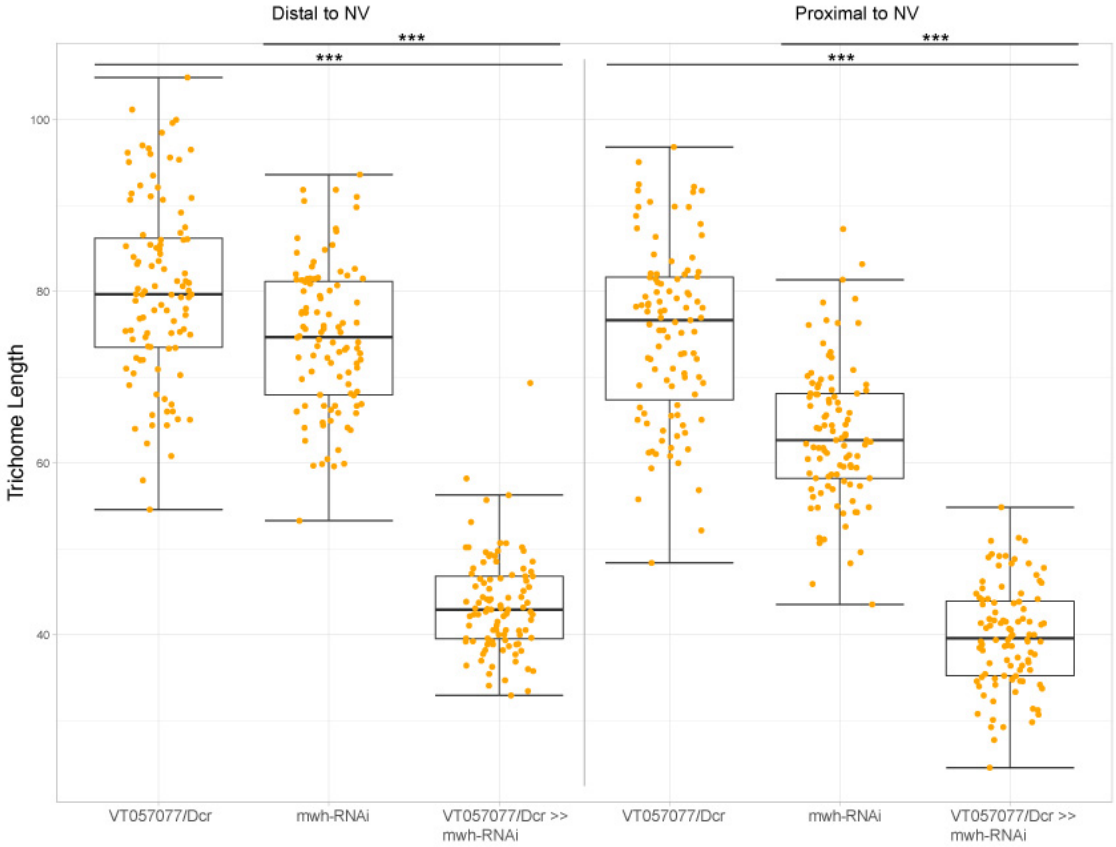


Fig. 26. Analysis of trichome length after RNAi knockdown of *mwh* in T2 legs using driver line VT057077/Dicer. Boxplots show measurements from the distal part of the femur (left) and near to the NV (right). There was a significant decrease in the length of the trichomes on both the distal part of the femur and close to the NV when *mwh* was knocked down (*** $p < 0.001$) Data was tested for normality using Shaprio-Wilk followed by an ANOVA.Tukey's multiple comparison test was used to measure significance between groups.

There was also a strong effect on the polarity of the trichomes all over the T2 leg after knockdown of *mwh* (Fig 27 A-B). Predictably, given the name of the gene, this effect also extended to the trichomes on the wing, where there also appears to be an increase in the number of trichomes per cell (Fig 27 C-D)

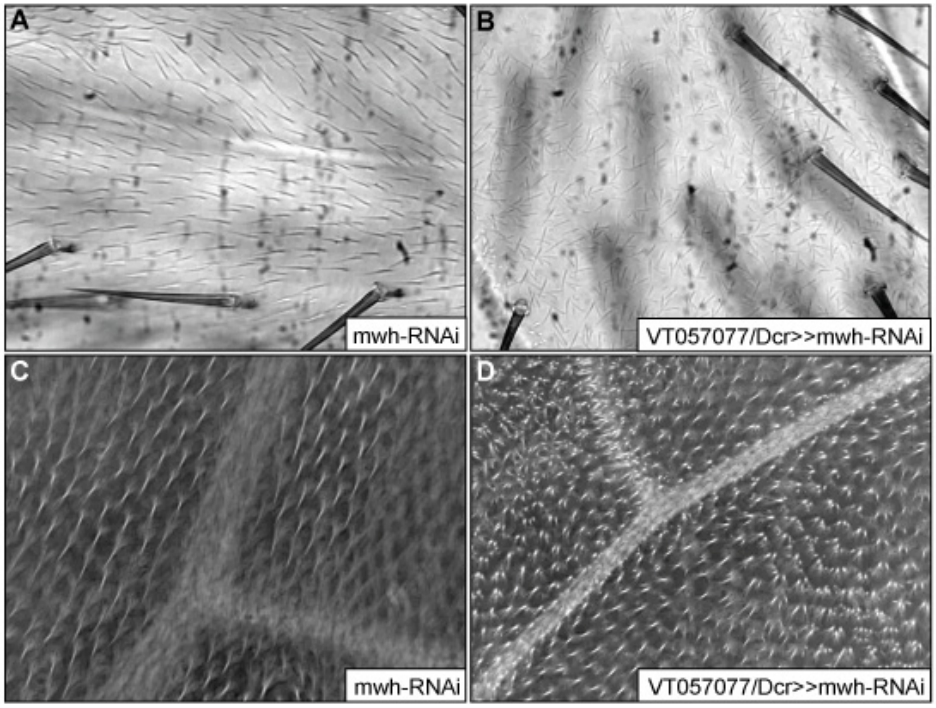


Fig. 27. Effect on the morphology of trichomes both on the T2 leg and the wing after knockdown of *mwh*. (A) Trichomes on the distal part of the T2 femur for *mwh*-RNAi parental line (B) trichomes on the distal part of the T2 femur for VT057077/Dcr>>*mwh*-RNAi where the trichomes are significant smaller in size (Fig. 26) and the polarity is also affected. (C) wing trichomes from the *mwh*-RNAi parental line (D) wing trichomes for the VT057077/Dcr>>*mwh*-RNAi line where again the trichomes appear much smaller and the polarity has been affected, and it also appears that there was more than one trichome per cell.

3.3.2.6 neo RNAi

RNAi knockdown of *neo* in T2 legs did not show any effect on the length of the trichomes either proximal to the NV or on the more distal part of the femur in comparison to the parental controls (Fig. 28).

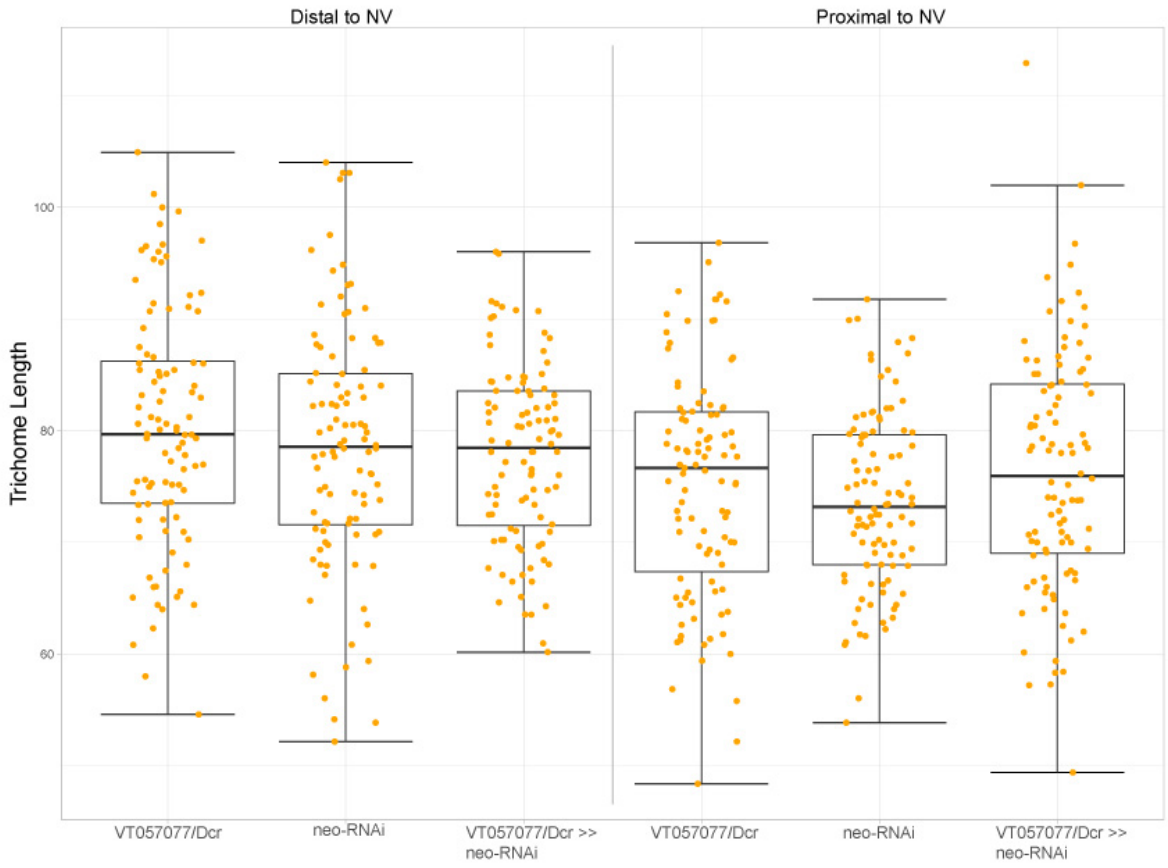


Fig. 28. Analysis of trichome length after RNAi knockdown of *neo* in T2 legs using driver line VT057077/Dicer. Boxplot shows measurements from the distal part of the femur (left) and near to the NV (right). There was no significant effect on the length of the trichomes on either the distal part of the femur or near to the NV. Data was checked for normality using Shaprio-Wilk followed by an ANOVA.Tukey's multiple comparison test was used to assess significance between groups.

3.3.2.7 Rab23 RNAi

Knockdown of *Rab23* did not show any effect on the length of the trichomes either on the distal part of the femur or proximal to the NV in comparison to parental controls (Fig. 29).

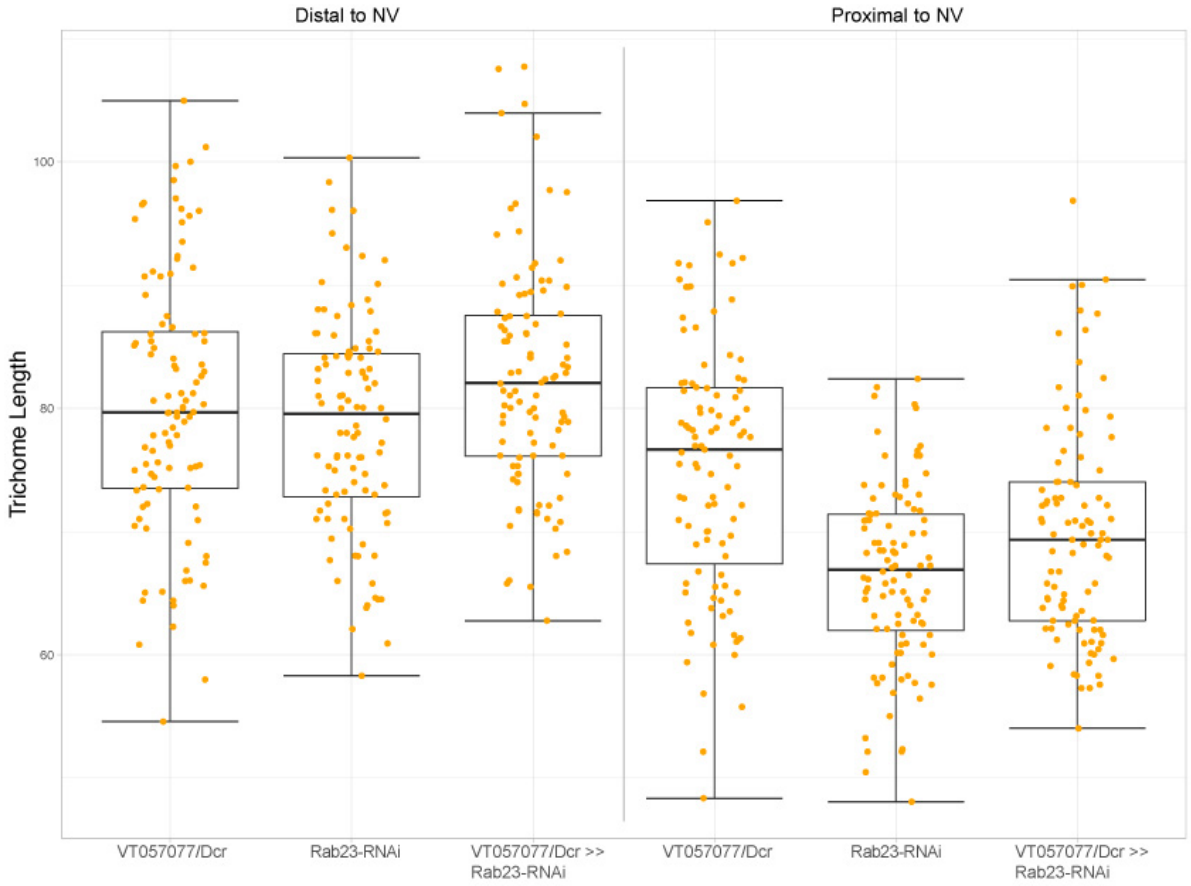


Fig. 29. Analysis of trichome length after RNAi knockdown of *Rab23* in T2 legs using driver line VT057077/Dicer. Boxplots show measurements from the distal part of the femur (left) and around the NV (right). There was no effect on the length of the trichomes on either the distal part of the femur or near to the NV. Data was tested for normality using Shapiro-Wilk followed by an ANOVA. Tukey's multiple comparison test was used to measure significance between groups.

While three (*CG14395*, *crebA* and *mwh*) of the seven focal genes tested showed an effect on the morphology of the trichomes during the RNAi analysis, (Figs 22, 24, 27) it was already known that *sha*, *svb* and *tal* affect the development of the leg trichomes (Arif et al., 2013; Kittelmann et al., 2018). Moreover, the genes tested with RNAi that

did not show any effect on the trichomes cannot necessarily be ruled out as not involved in the GRN controlling leg trichome development, as it is likely that the genes act together to ultimately build trichomes and therefore knocking down just one of the genes on its own might not show a discernable effect on trichome morphology or a very subtle effect that could not be detected. It is also possible that RNAi simply did not work for these genes.

3.3.3 Motif enrichment

Given that the 12 genes identified (Fig. 21) are predicted targets of *miR-92a* and Svb-dependent in embryos, if they are in the leg trichome network downstream of Svb it is likely that the genes contain Svb-responsive enhancers. In a study of Svb-dependent genes in the GRN controlling larval trichome development, Menoret et al. (2013), successfully predicted and identified enhancers in their set of Svb-dependent genes using motif analysis for Svb binding sites. Therefore, I took a similar approach for the genes identified in this analysis by using the ATAC-seq data from Kittelmann et al. (2018) to look for enrichment of motifs in accessible chromatin in the window of time when leg trichomes are specified.

Sequences corresponding to the peaks of open chromatin for these 12 genes were run through i-*cis*Target (Herrmann et al., 2012; Imrichová et al., 2015) to determine which TF motifs are enriched in my dataset. The program also outputs a normalized enrichment score (NES) for each motif which then allows the motifs to be ranked in order of enrichment, default threshold for minimum NES is 3.0, anything below this is not reported (Herrmann et al., 2012; Imrichová et al., 2015). As control sequences I used adjacent regions of closed chromatin in T2 legs at 24 hAPF according

to the ATAC-seq data for all 12 genes of interests an additional control I checked the enrichment of motifs in epidermal genes that are not dependent on the TF Svb (Menoret et al., 2013) I then compared the top 5 motifs for each dataset.

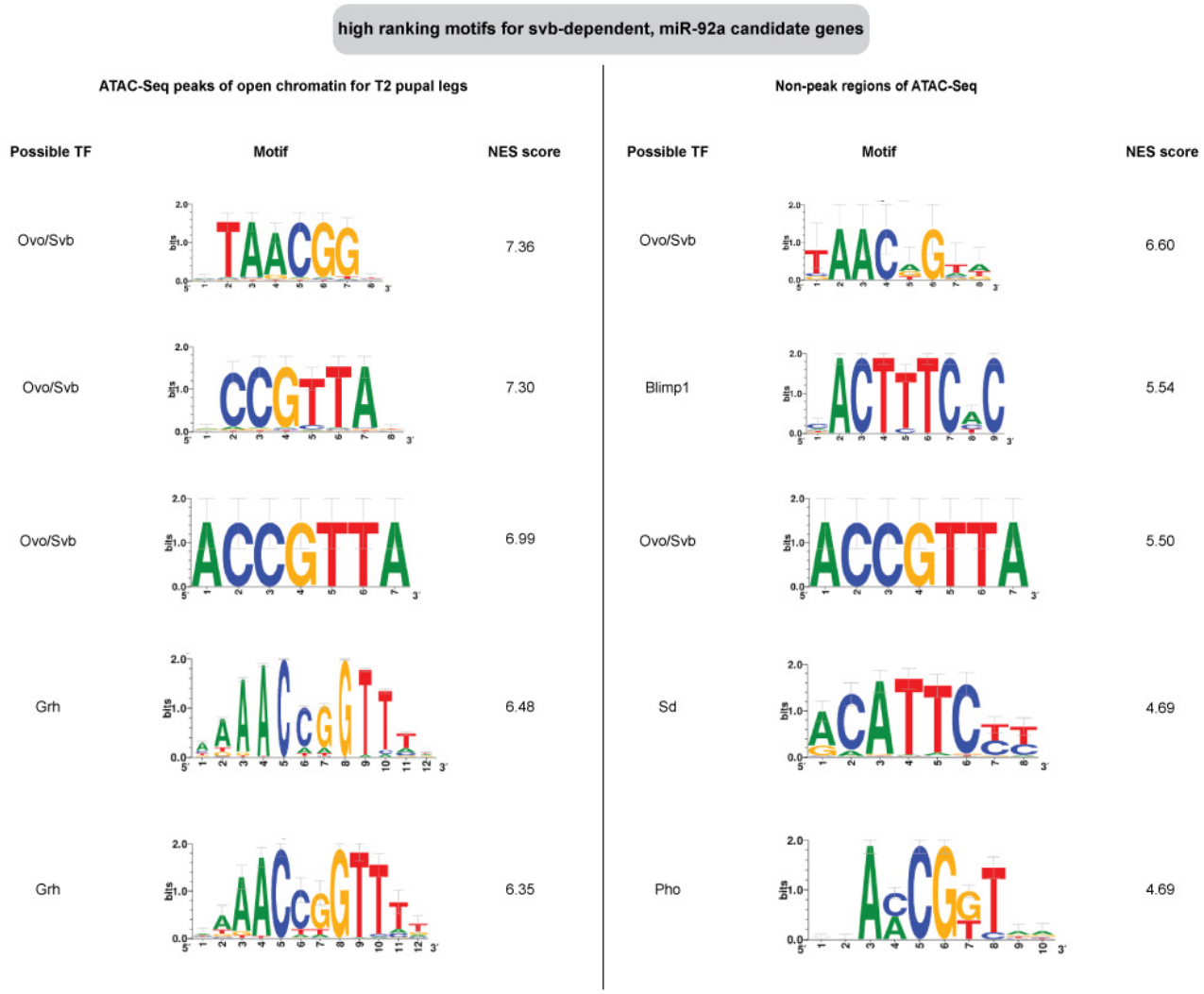


Fig. 30. Top 5 motifs found enriched in Svb-dependent, miR-92a candidate genes. Motifs on the left are enriched in regions that have been called as peaks in ATAC-seq data for T2 pupal legs by MACS (Kittelmann et al., 2018). Motifs on the right are enriched in the same genes but for regions that are not called as peaks in the ATAC-seq data. In the open chromatin the top three motifs correspond to Svb/Ovo motifs and all contain the consensus CnGTT (Menoret et al., 2013), motifs 4 and 5 match to Grh motifs. In the regions that are 'inaccessible' chromatin, the top motif also corresponds to Ovo/Svb but it not one of the sites found enriched in the open chromatin. The third site also corresponds to Ovo/Svb and matches to the site also ranking third in the open chromatin. Note also that the NES is lower for all the motifs in the closed chromatin. There is no enrichment of Grh found in the top 5 motifs for 'inaccessible' chromatin, Grh motifs were present further down in the list of enriched motifs but with a much lower NES.

For the peaks of open chromatin the top three motifs all match to the Ovo/Svb site (consensus CnGTT, Menoret et al., 2013), which is consistent with this set of genes being directly dependent on the Svb TF for activation. The first motif seems to be a reverse complementation of the second motif, perhaps explaining the similarity between their NES scores. The third motif seems to be a slight extension of the second motif with an extra 'A' at the beginning of the motif. The motifs that rank four and five match to a TF called Grainy-head (Grh) (Fig. 30). Interestingly Grh has been shown in previous studies to be necessary for epithelial enhancer activation (Jacobs et al., 2018) it acts as a pioneer factor by displacing nucleosomes, therefore making the enhancer accessible to TF binding (Jacobs et al., 2018).

In the motif search of 'inaccessible' DNA (i.e. that which is not called by MACS as an ATAC-seq peak) the top five ranked motifs (Fig. 30) showed a number of differences in comparison to the open chromatin and all had lower NES (Fig 30). While the top site found was actually a site corresponding to Ovo/Svb binding, this site was not found in the top enriched motifs for open chromatin. However it does seem to partially overlap in sequence with the top result from the open chromatin motifs. The second highest ranked site for the 'inaccessible' chromatin matched to a TF called Blimp-1, which is linked to the ecdysone pathway (Agawa et al., 2007), and has roles in prepupal development (Akagi and Ueda, 2011) and tracheal development (Ng, Yu and Roy, 2006; Ozturk-Colak et al., 2018). The third motif found enriched was 'ACCGTTA' and it was also found third in the open chromatin, but with a much higher NES. This is again consistent with a higher enrichment for Svb binding sites in the open chromatin for these 12 genes in the leg. The fourth motif found in the 'inaccessible' chromatin was for the TF Scalloped (Sd), which is a member of the hippo

pathway (Gokhale and Pfleger, 2019) and, amongst other processes, plays a role in the development of wings (Halder et al., 1998), the mushroom-bodies (Rohith and Shyamala, 2017) and the legs (Garg et al., 2007). The fifth motif found to be enriched in the 'inaccessible' chromatin matches to Pleiohomeotic (pho), which is a polycomb-group protein and has been shown to have several roles in development including regulation of *engrailed* and *hedgehog* in imaginal discs (Randsholt, Maschat and Santamaria, 2000) and working with Hox genes to repress alternative cell fates (Domsch et al., 2019).

For the top five motifs for 'inaccessible' chromatin I did not find enrichment for Grh motifs. This fits with Grh's proposed role as a pioneer factor that allows the DNA to be accessible to TFs (Jacobs et al., 2018) and therefore as a potential marker for active leg enhancers in this case (Fig. 30).

The top five motifs in epidermal genes that are not dependent on Svb revealed enrichment for IRF-1 motifs (Fig. S3). These motifs are very similar in sequence to the motif for the TF Blimp-1 which was also found in the 'inaccessible' chromatin (Fig. S3). I did not see enrichment for either Svb/Ovo or Grh in this set of genes.

From my analysis of the overall motif enrichment for these datasets, I hypothesise that leg enhancers for these 12 genes, which drive their expression to generate leg trichomes, will correspond to regions of open chromatin that contain both Svb/Ovo and Grh motifs. In addition to assessing the overall motif enrichment for the sequences provided to i-*cis*Target, it is possible to identify the location of the regions of DNA (the CRRs) that show enrichment for chosen motifs. Therefore, I next analysed the open chromatin for the 12 genes to identify potential leg enhancers among the CRRs that contain Svb/Ovo and/or Grh motifs.

3.3.4 Enrichment of Svb/Ovo and/or Grh motifs in ATAC-seq peaks

3.3.4.1 CG4678

Micro-array results from Menoret et al. (2013) suggest that *CG4678* is regulated by Svb in embryos (Menoret et al., 2013).I found seven peaks of accessible chromatin in T2 legs in the region analysed for *CG4678*. When the peaks were submitted to i-*cis*Target one region (encompassing peak 1) was found that overlaps with a CRR enriched for both Svb and Grh binding sites (Fig. 31 - orange box) and therefore may indicate a putative leg enhancer.

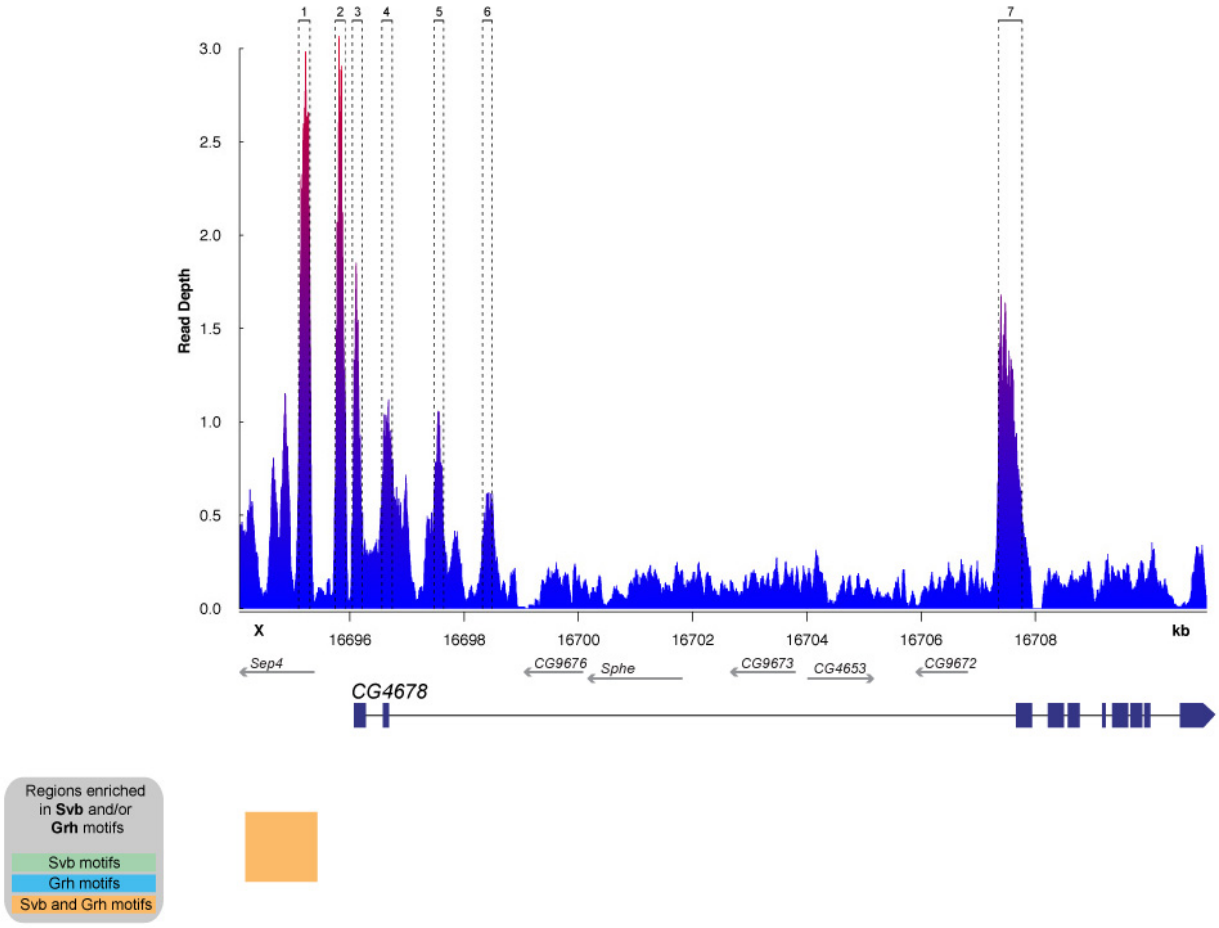


Fig. 31. ATAC-Seq profile for *CG4678* showing a region enriched for Svb and Grh motifs. There are seven peaks of open chromatin as determined by MACS (Kittlemann et al., 2018) for the region encompassing *CG4678*. When analysed with i-*cis*Target only one region of enrichment is found for the motifs of interest. This region is enriched for both Svb and Grh and overlaps with the first ATAC-Seq peak (orange box).

3.3.4.2 CG5742

CG5742 was also listed as Svb-dependent based on the microarray results from Menoret et al. (2013). However, I did not find any regions of open chromatin for this gene that overlap with CRRs that are significantly enriched in either Svb or Grh motifs (Fig. 32).

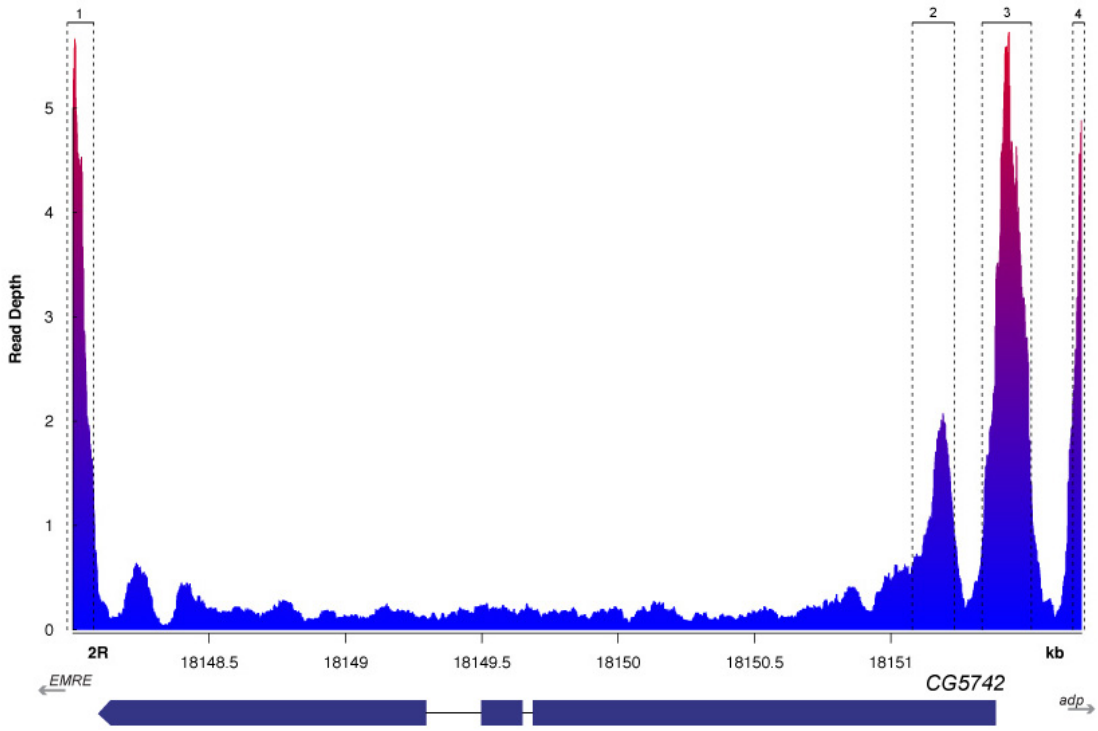


Fig. 32. ATAC-seq profile for *CG5742* in T2 pupal legs. There are four defined peaks of open chromatin as determined by MACS (Kittelmann et al., 2018). Despite being expressed in the legs at this developmental stage and containing open chromain, there were no regions of enrichment for either Svb or Grh motifs in regions of open chromatin.

3.3.4.3 CG8303

CG8303 has previously been shown to be dependent on Svb in embryos using in-situ hybridization and microarray profiling (Menoret et al., 2013). In my analysis of *CG8303* I found seven peaks of accessible chromatin in T2 pupal legs. Analysis from i-*cis*Target showed three regions in this locus overlapping with peaks 1, 3 and 4, which are significantly enriched for Grh motifs but not Svb (Fig. 33). From the known role

of Grh as a pioneer factor for epithelial enhancers (Jacobs et al., 2018) this suggests that these regions are in fact open and accessible, and may be bound by other TFs that are active at this developmental stage in legs.

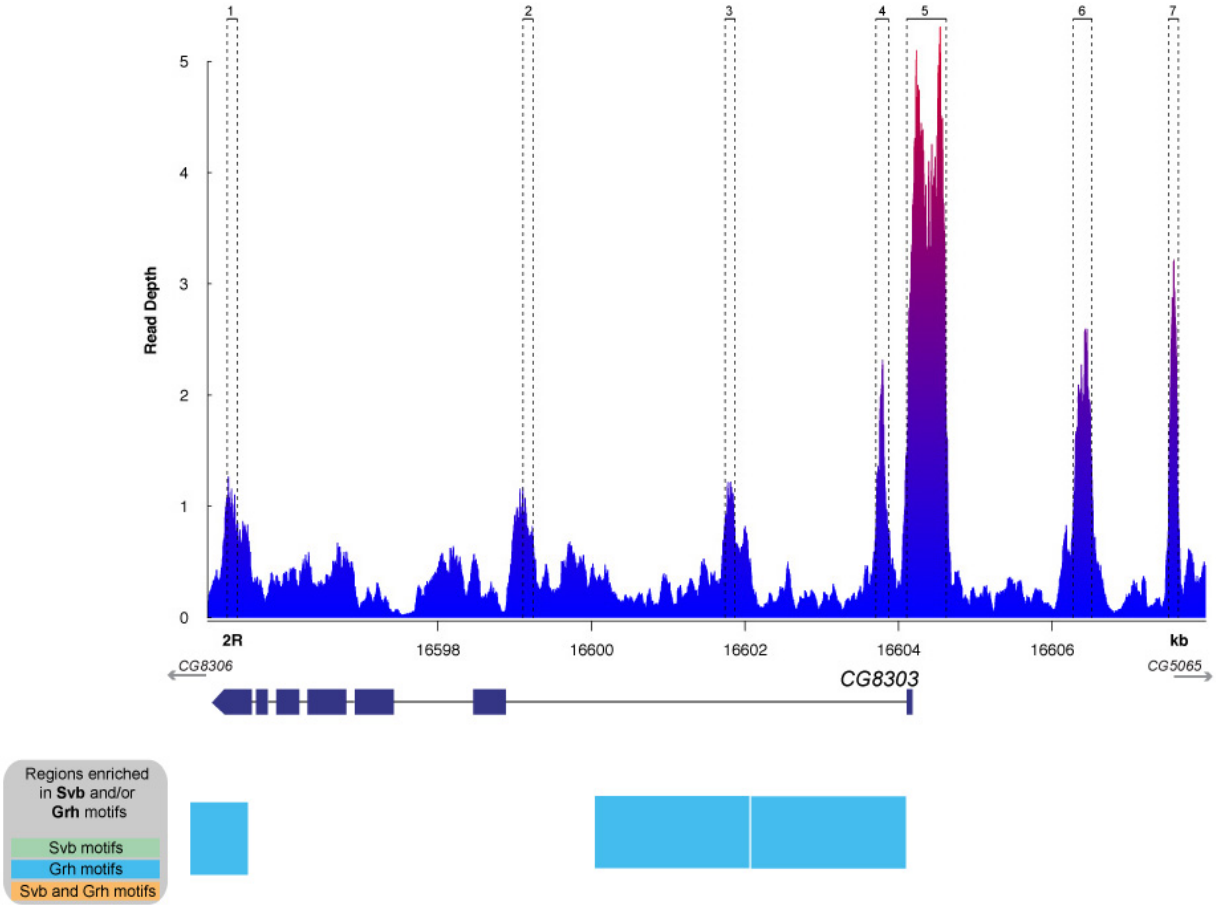


Fig. 33. ATAC-Seq profile for *CG8303* in T2 pupal legs. There are seven regions of accessible chromatin as determined by MACS (Kittelmann et al., 2018). In *CG8303* there are three locations with enrichment for the Grh motifs (blue boxes), which overlap with peaks 1, 3 and 4. No significant enrichment for Svb motifs was found by i-cisTarget.

3.3.4.4 CG14395

Expression of *CG14395* in the embryonic epidermis was shown by (Menoret et al., 2013) to be dependent on Svb. We previously also observed that knockdown of *CG14395* in T2 legs has a significant effect on trichome length (Fig. 22) (Kittelmann et al., 2018). In the ATAC-seq profile for the *CG14395* locus there are seven peaks of accessible chromatin, i-*cis*Target identified two CRRs that are significantly enriched in both Svb and Grh motifs (Fig. 34). These regions of enrichment overlap with peaks 1, 2 and 4 (Fig. 34). These represent good candidates for Svb-dependent leg enhancers of *CG14395*.

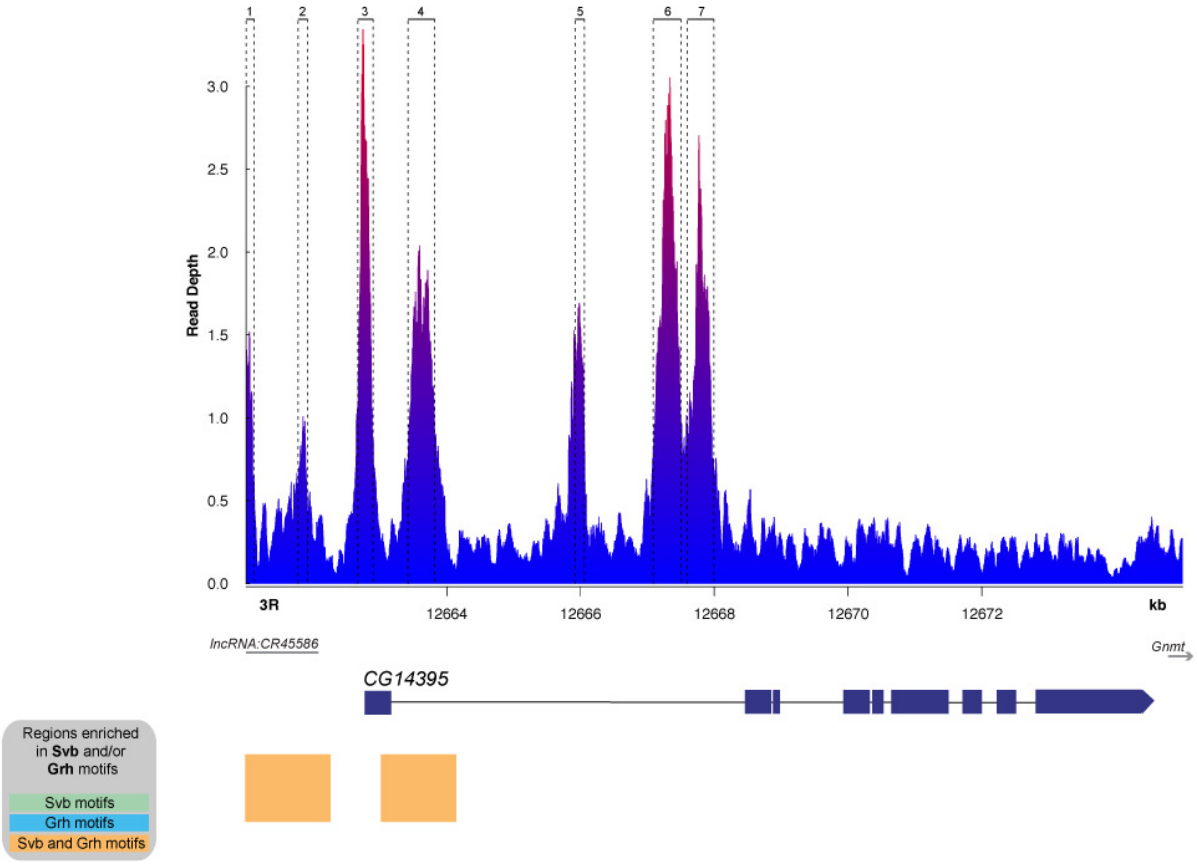


Fig. 34. ATAC-seq profile for *CG14395* in T2 pupal legs. There are seven regions of open chromatin as determined by MACS (Kittelmann et al., 2018). i-*cis*Target revealed two regions of enrichment for both Svb and Grh motifs that overlap with peaks 1, 2 and 4 (orange boxes).

3.3.4.5 CrebA

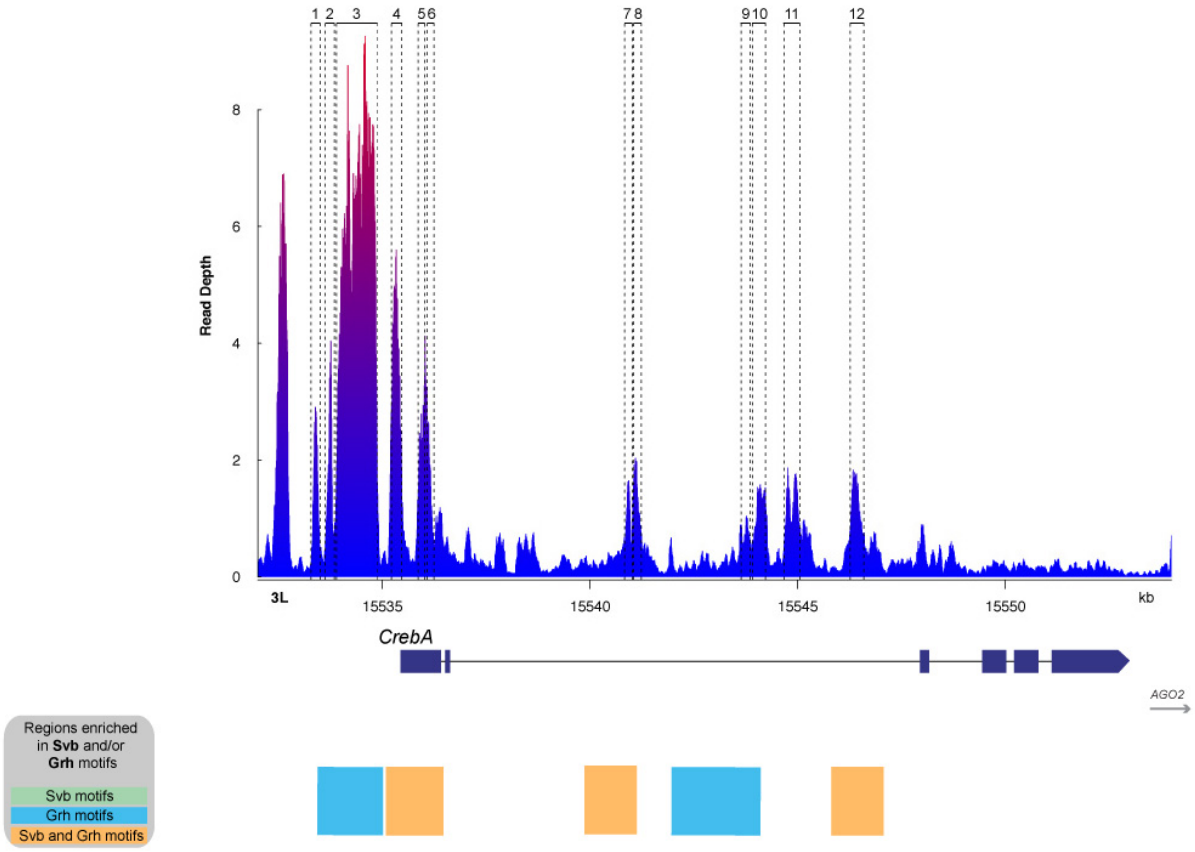


Fig. 35. ATAC-seq profile for *CrebA* in T2 pupal legs. Peaks called by MACS (Kittelmann et al., 2018) are outlined in dashed boxes and numbered. Two regions show Grh motif enrichment (blue boxes) and overlap with peaks 1, 2, 9 and 10. There are 3 regions with enrichment of both Svb and Grh motifs (orange boxes), which overlap with peaks 4, 5, 6, 7 and 12.

CrebA was first indicated to have a role in trichome formation by Andrew et al. (1997). Additionally, I saw a significant effect on the length of the trichomes proximal to the NV upon knockdown of *CrebA* (Fig. 24). Analysis of ATAC-seq data for the region containing, and surrounding, *CrebA* revealed 12 peaks of accessible chromatin as determined by MACS (Kittelmann et al., 2018). Analysis from i-*cis*Target shows three regions with Svb and Grh motif enrichment and two with only Grh motifs (Fig. 35) This suggests there are multiple candidate regulatory regions in *CrebA*, those which are enriched in both Svb and Grh, to be further tested to determine if they are Svb-dependent enhancers involved in the development of trichomes in T2 legs. The

regions which only show enrichment for Grh, if active, may be bound by different TFs or are perhaps even used in other tissues.

3.3.4.6 forked

Previous studies have predicted that *forked* acts downstream of Svb and *forked* alleles can affect the morphology of trichomes (Chanut-Delalande et al., 2006; Dickinson and Thatcher, 1997). The ATAC-seq profile for *forked* in T2 pupal legs revealed a large number of peaks of open chromatin - 37 peaks in total (Fig. 36). i-*cis*Target identified eight regions that overlap with CRRs and show enrichment for Svb and/or Grh motifs. Three regions have enrichment for only Grh motifs (blue boxes) and overlap with peaks 1, 2 and 19. Four regions were enriched for only Svb motifs (green boxes) and these regions overlap with peaks 15, 21-25, and 32-35. In addition, only one region was significantly enriched in both Svb and Grh motifs and this overlaps with peak 26 (Fig. 36). These results suggest that regulation of *forked* is complex, but my analysis has revealed at least one region that represents an excellent candidate for a leg enhancer of this gene.

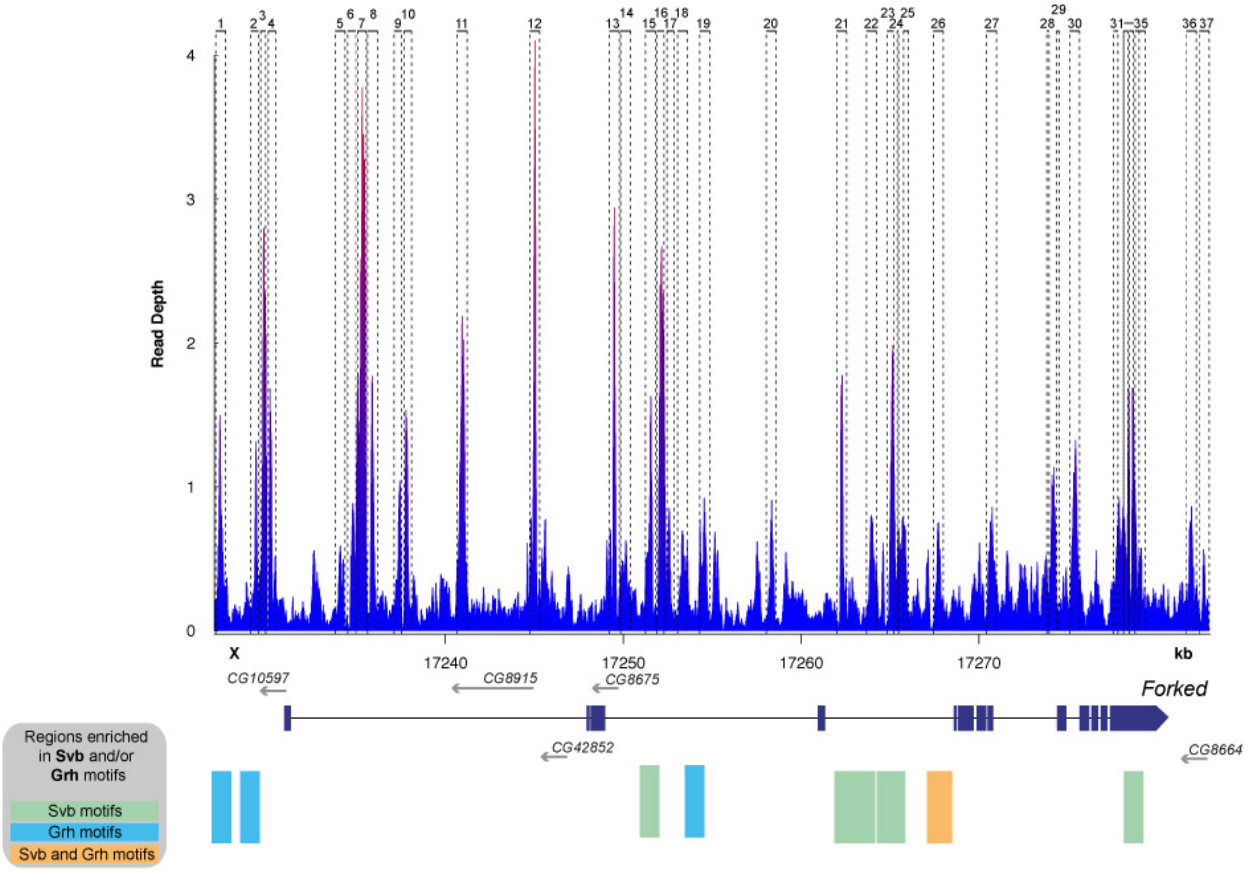


Fig. 36. ATAC-seq profile for *forked* in T2 pupal legs. There are 37 peaks of accessible chromatin as determined by MACS (Kittelmann et al., 2018). i-*cis*Target identified seven regions with enrichment for the motifs for Svb and/or Grh. Three regions (shown in blue boxes) are enriched for just Grh, four regions (shown in green boxes) are enriched for just Svb and one region is significantly enriched for both Svb and Grh motifs (orange box).

3.3.4.7 mwh

mwh has been previously been validated as Svb dependent in the embryonic epidermis (Menoret et al., 2013). I also saw a dramatic effect on the morphology and polarity of the 2nd leg trichomes when knocking down *mwh* (Fig. 27). I identified 18 peaks of accessible chromatin as determined by MACS (Kittelmann et al., 2018) in the ATAC-seq profile for *mwh* in T2 pupal legs (Fig. 37). i-*cis*Target highlighted seven regions corresponding to CRRs and with significant enrichment in Svb and/or Grh motifs (Fig. 37). One region showed enrichment for just Grh motifs (Fig. 37 - blue box) and this overlaps with peak 4, two regions show significant enrichment for both Svb and Grh

motifs (Fig. 37 - orange boxes) and overlap with peaks 5, 9 and 10. Finally, two regions show enrichment for just Svb motifs (Fig. 37 - green boxes) and these overlap with peaks 6 and 15-18. Based on the hypothesis, the regions containing both Svb/Ovo and Grh motifs may be leg enhancers active in the trichome GRN and can be further tested. The region enriched for just Grh, if active, could be a potential enhancer for another tissue or is bound by other TFs. The locus for *mwh* also contains regions that are only enriched in Svb motifs and not Grh motifs, suggesting perhaps that these regions are not active enhancers at this stage of development.

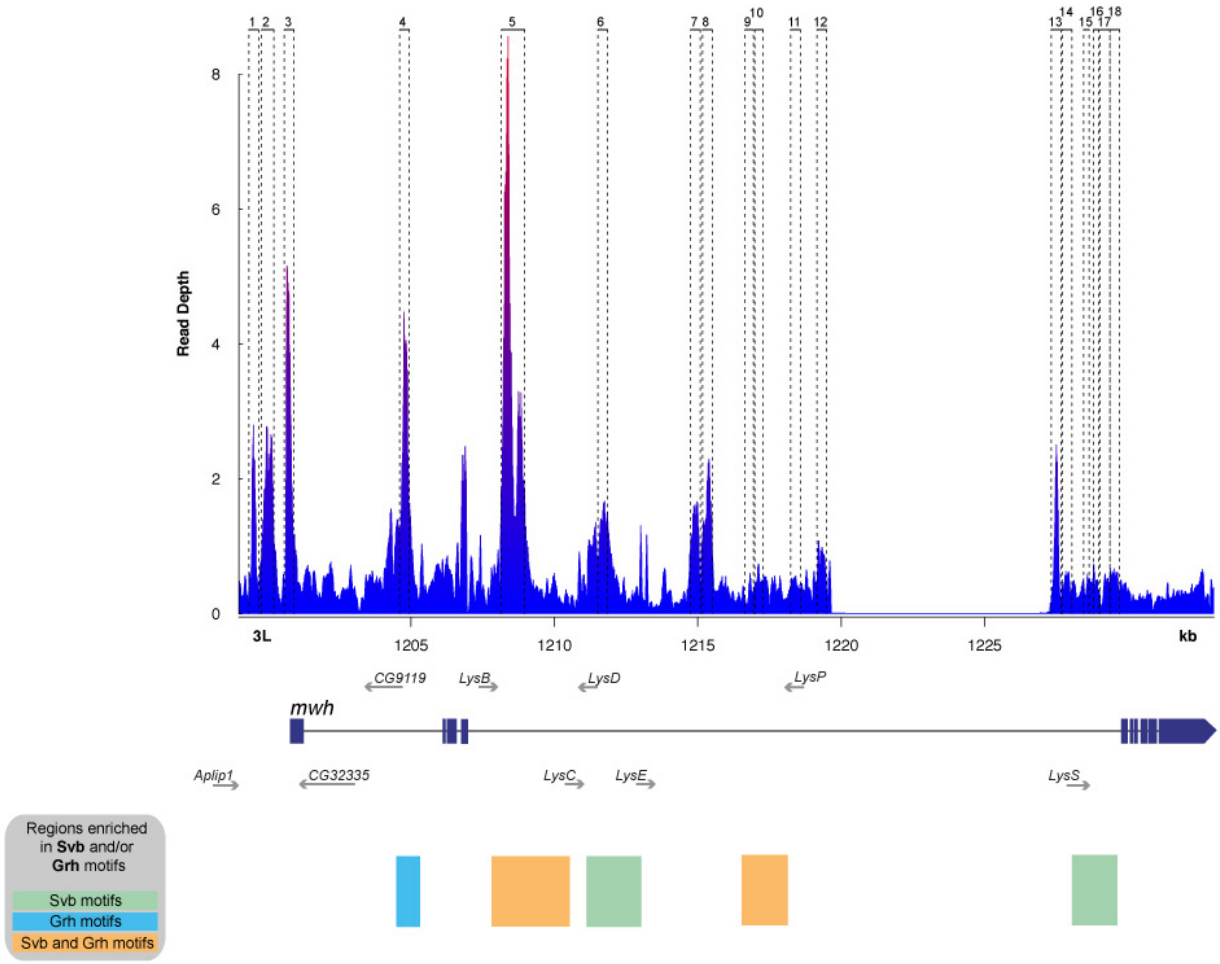


Fig. 37. ATAC-seq profile for the region containing *mwh* for T2 pupal legs. There are 18 peaks in total called for this locus (Kittelmann et al., 2018). Peaks are numbered and outlined by dashed boxes. i-*cis*Target identified five regions that show enrichment for Svb and/or Grh motifs. One region has enrichment for just Grh (blue box), two regions have enrichment for just Svb motifs (green boxes).Two regions show enrichment for both Svb and Grh motifs (orange boxes). Note there appears to be a gap in the profile for *mwh* which represents a region that did not produce any (or extremely low) signal in the ATAC-seq data (Kittelmann et al., 2018).

3.3.4.8 neo

neo has been validated as being Svb-dependent in the embryonic epidermis (Menoret et al., 2013). The ATAC-seq profile for the region spanning *neo* for T2 pupal legs shows there are four peaks of accessible chromatin (Kittelmann et al., 2018) (Fig. 38). The analysis of peaks using i-*cis*Target identified just one region that corresponds to a CRR and showed significant enrichment only for the Grh motifs (Fig. 38- blue box), and overlaps with peaks 1 and 2. As with the other profiles, this enrichment of just Grh motifs may suggest this is an enhancer in other epithelial tissues, or bound by different TFs.

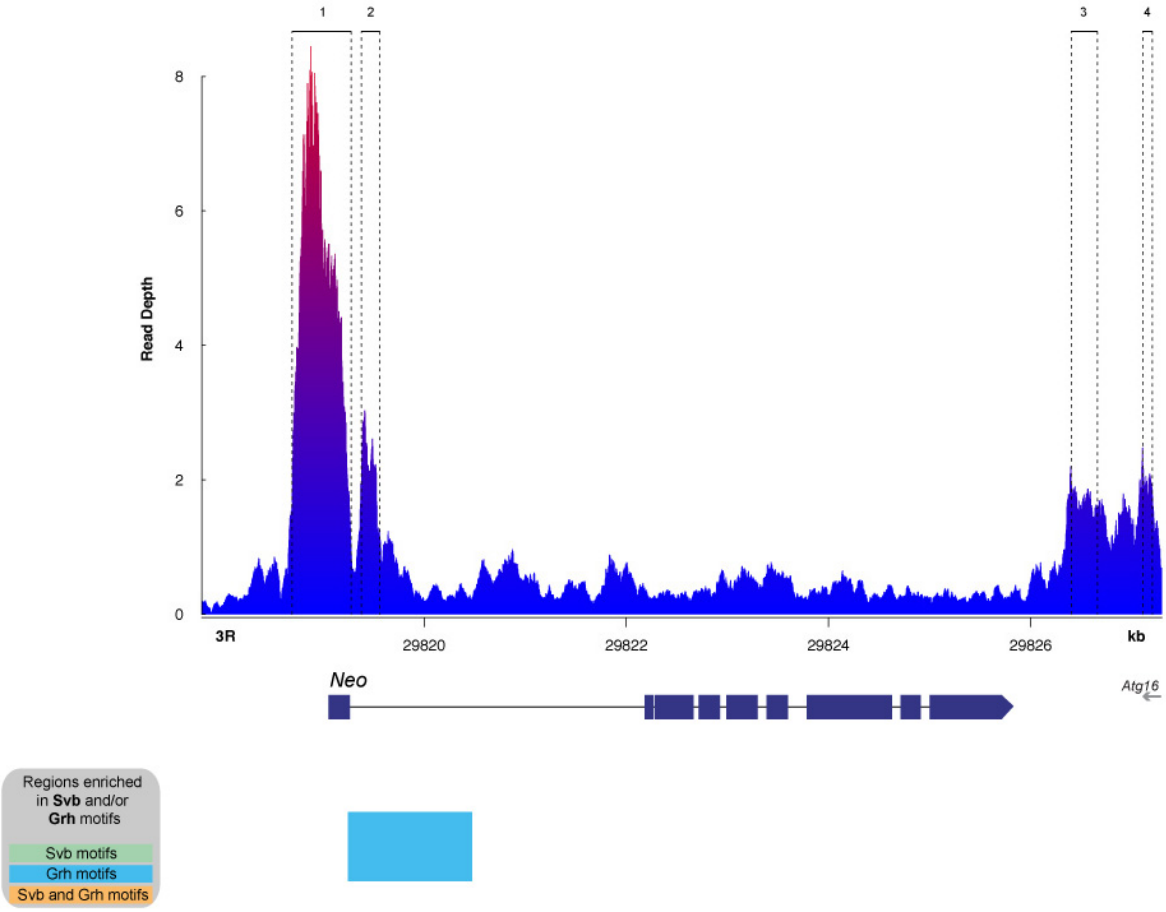


Fig. 38. ATAC-Seq profile for *neo* in T2 pupal legs. There are four peaks of open chromatin in this region as determined by MACS (Kittelmann et al., 2018). Peaks are numbered and outlined with dashed boxes. The i-*cis*Target analysis revealed one region which shows enrichment for Grh motifs (blue box), which overlaps with peaks 1 and 2. No regions were identified that showed enrichment of Svb motifs.

3.3.4.9 Rab23

Rab23 expression has been shown to be dependent on Svb in the embryonic epidermis (Menoret et al., 2013). The ATAC-seq data for T2 legs in the region analysed for *Rab23* (Fig. 39) shows there are six peaks of accessible chromatin as determined by MACS (Kittelmann et al., 2018) (Fig. 39). Further analysis of the peaks of accessible chromatin using i-*cis*Target revealed two regions that are enriched in Svb and/or Grh motifs. One region is only enriched for Grh motifs (blue box) and overlaps with peak 1. The second region is enriched for both Svb and Grh motifs and overlaps with peak 3 (Fig.39) representing an excellent candidate region for a *Rab23* leg enhancer.

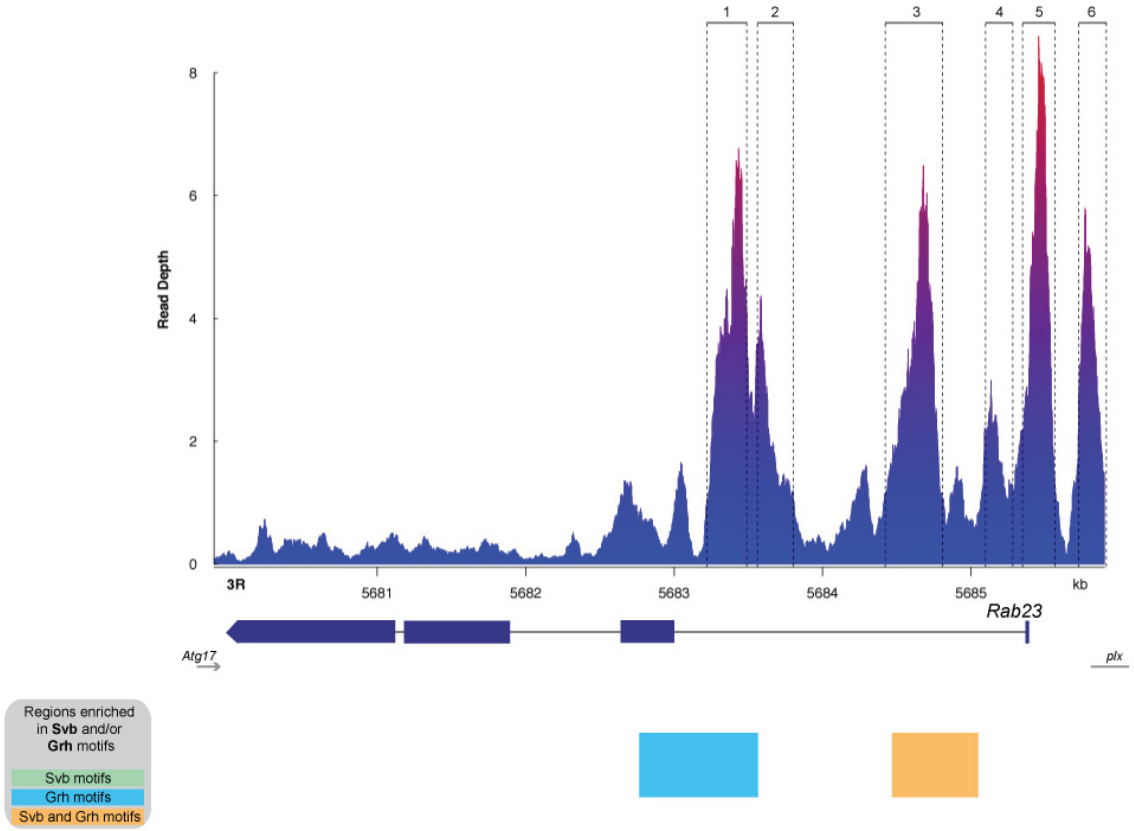


Fig. 39. ATAC-seq profile for *Rab23* in T2 pupal legs. There are six peaks of accesible chromatin as called by MACS (Kittelmann et al., 2018). Peaks are numbered and outlined in dashed boxes. Analysis of the peaks with i-*cis*Target shows two regions that have significant enrichment of motifs of interest. There is one region enriched for Grh motifs (blue box) that overlaps with peak 1, and a second region which shows enrichment for both Grh and Svb motifs (orange box). This second region overlaps with peak 3.

3.3.4.10 sha

sha is known to be involved in the development of both larval and leg trichomes and is a validated target of Svb in the embryonic epidermis (Chanut-Delalande et al., 2006; Menoret et al., 2013). Plotting the ATAC-seq profile for *sha* in T2 legs revealed 14 peaks of accessible chromatin as determined by MACS (Kittelmann et al., 2018) (Fig. 40). Further analysis of the peaks of accessible chromatin using i-*cis*Target revealed several regions significantly enriched in Svb and/or Grh motifs. There is one region enriched in only Svb motifs (green box) that overlaps with peak 3. In addition, there are three CRR regions that show enrichment for both Svb and Grh (orange boxes), these three regions overlap with peaks 1, 2 and 4-8 (Fig. 40). Interestingly there are three known enhancers of *sha,* two of which drive embryonic epidermal expression, identified in (Menoret et al., 2013) and these overlap with regions of Svb and Grh motif enrichment and suggest they may regulate Svb dependent expression of *sha* in legs as well as in embryos (Fig. 40).

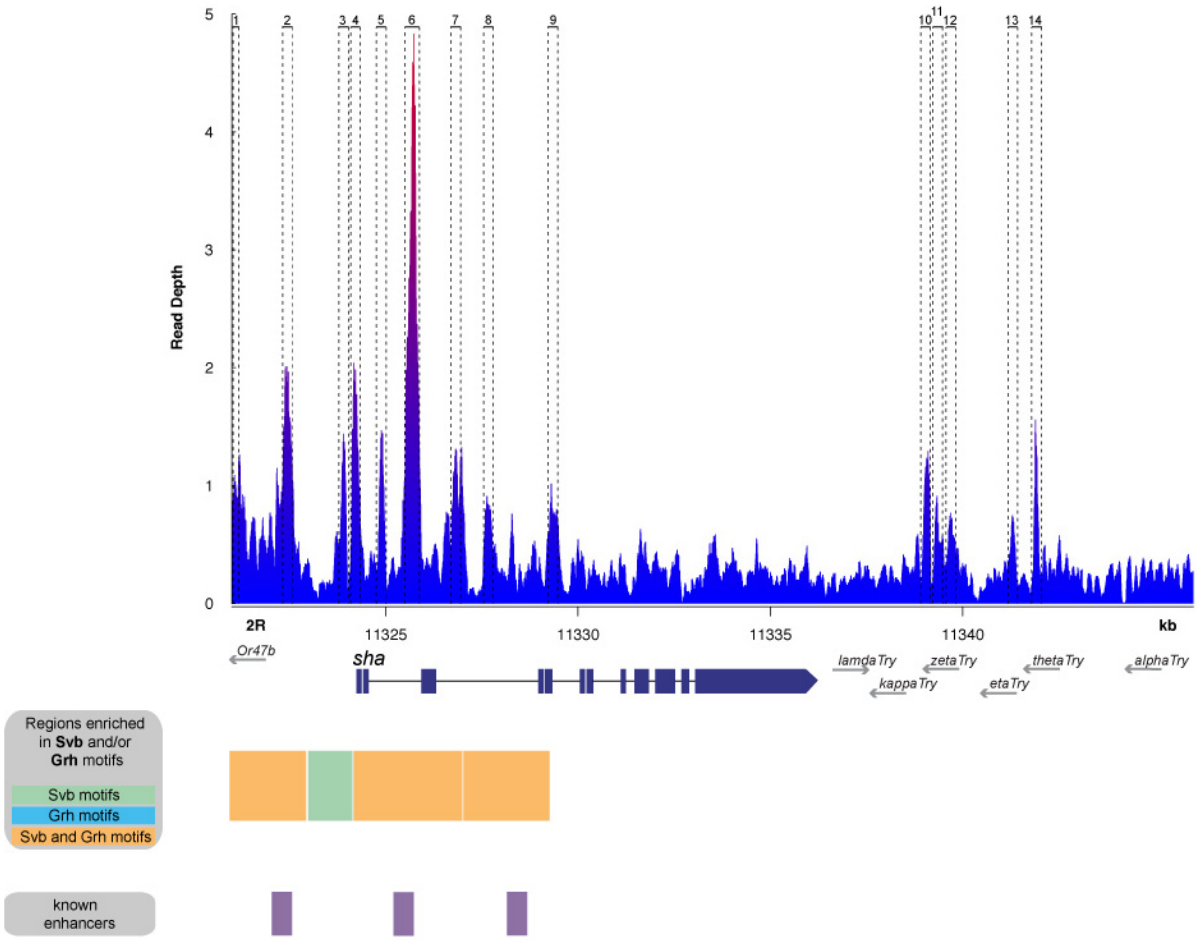


Fig. 40. ATAC-seq profile for *sha* in T2 pupal legs. There are 14 peaks of accessible chromatin for this locus as determined by MACS (Kittelmann et al., 2018). Peaks are numbered and outlined in dashed boxes. Analysis of peaks with i-*cis*Target revealed several regions of enrichment for the motifs of interest. One region shows enrichment for just Svb motifs (green box) and three regions that show enrichment for both Svb and Grh (orange boxes). Also shown are known enhancers of *sha* (purple boxes), named *sha* 1-3 (left to right) Menoret et al., 2018. *sha* 1 and 3 were shown to drive expression in embryonic epidermal cells that bear trichomes (Menoret et al., 2013).

3.3.4.11 svb

svb is a key regulator of trichome development (Chanut-Delalande et al., 2006; Delon, Chanut-Delalande and Payre, 2003; Menoret et al., 2013; Sucena et al., 2003; Sucena and Stern, 2000). Analysis of the ATAC-seq profile for Svb in T2 legs showed 24 peaks of open chromatin as called by MACS (Kittelmann et al., 2018) (Fig. 41). i-*cis*Target revealed seven regions that overlap with CRRs and show significant enrichment for the motifs of interest. There are three regions that show significant enrichment for

just Svb motifs (green boxes) these overlap with peaks 3, 9 and 10. There are also three regions that are significantly enriched for Grh motifs (blue boxes), which overlap with peaks 8, 12, 13 and 14. Additionally, there is one region that is enriched for both Svb and Grh motifs (orange box), which overlaps with peak 4 (Fig. 41). In the analysed region there is one known enhancer of Svb that drives in T2 legs (Kittelmann et al., 2018) (Fig. 41 - purple box), which overlaps with CRRs that show enrichment for Svb and Grh, and Grh alone. This suggests that Svb might feedback to regulate its own expression in legs. There are several regions in this locus that have been tested in reporter assays and also express in T2 pupal legs (Kittelmann et al., 2018) (Fig. 41 - peach coloured boxes), which overlap with regions enriched in just Svb or just Grh. This may be because although these regions express in the leg they are not active Svb-dependent enhancers, but this requires further testing.

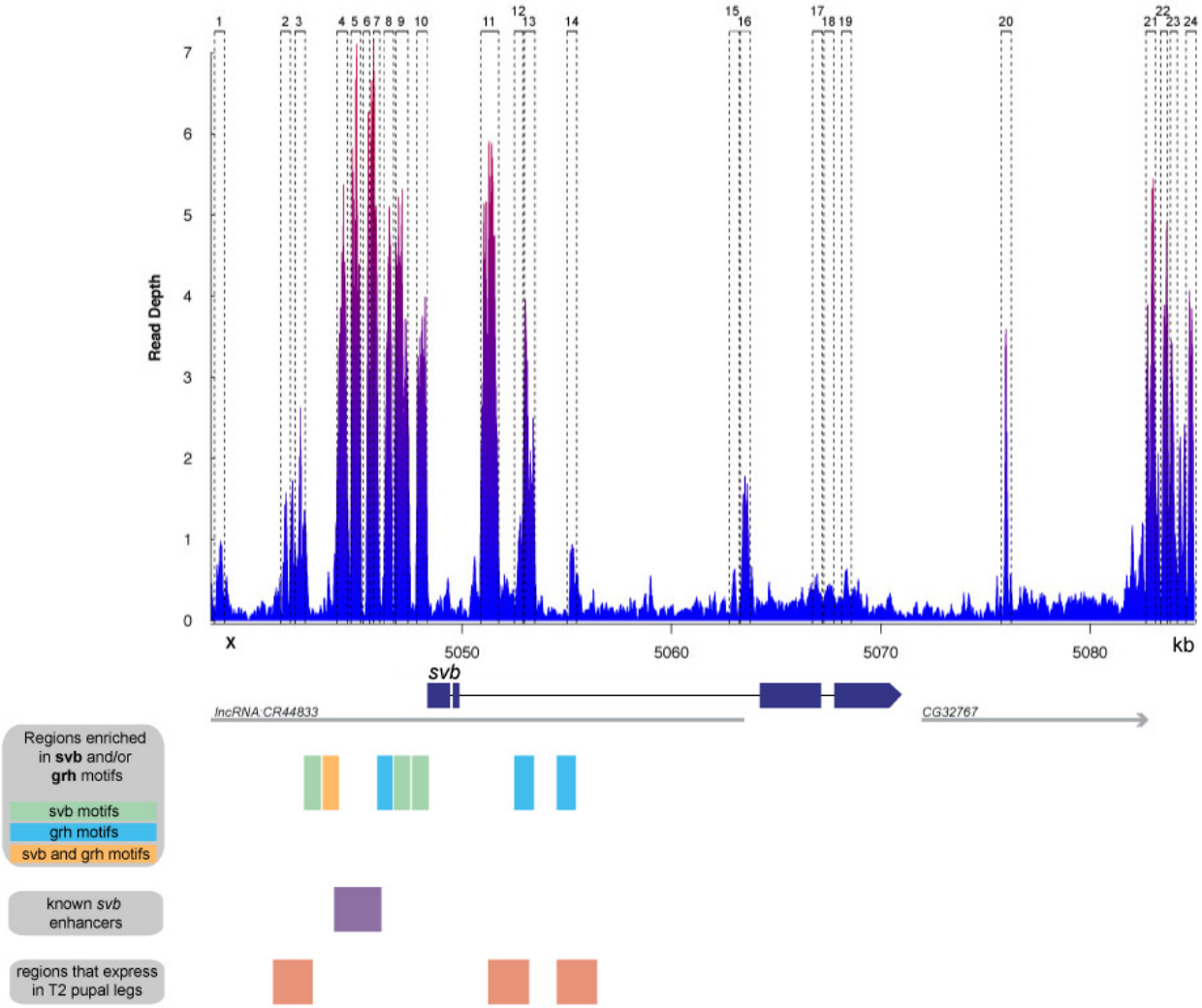


Fig. 41. ATAC-seq profile for *svb* in T2 pupal legs. There are 24 peaks in the analysed region as called by MACS (Kittelmann et al., 2018). Peaks are numbered and outlined in dashed boxes. Analysis with i-*cis*Target revealed seven regions that show enrichment for Svb and/or Grh motifs. Green boxes show the three regions that are enriched for Svb motifs, blue boxes show the regions enriched for Grh motifs and the orange box shows the region enriched for both Svb and Grh motifs. In this region there is one known enhancer of svb (purple box) that drives expression in T2 pupal legs (Kittelmann et al., 2018). There are also three regions that have been tested using reporter assays and have been shown to drive expression in T2 pupal legs (Kittelmann et al., 2018).

3.3.4.12 tal

tal encodes peptides that control the post-transcriptional activation of Svb in embryos (Kondo et al., 2010; Payre, 2004). My analysis of the ATAC-seq profile for *tal* in T2 legs revealed eight peaks of open chromatin as determined by MACS (Kittelmann et al., 2018) (Fig. 42). Analysis of the peaks of accessible chromatin using the i-*cis*Target database showed two regions of enrichment for the motif Svb. These CRRs overlap with peaks 4, 6 and 7 (Fig. 42). This suggests there are regions that could be potential Svb-dependent enhancers independent of Grh or that are employed during another developmental time-point.

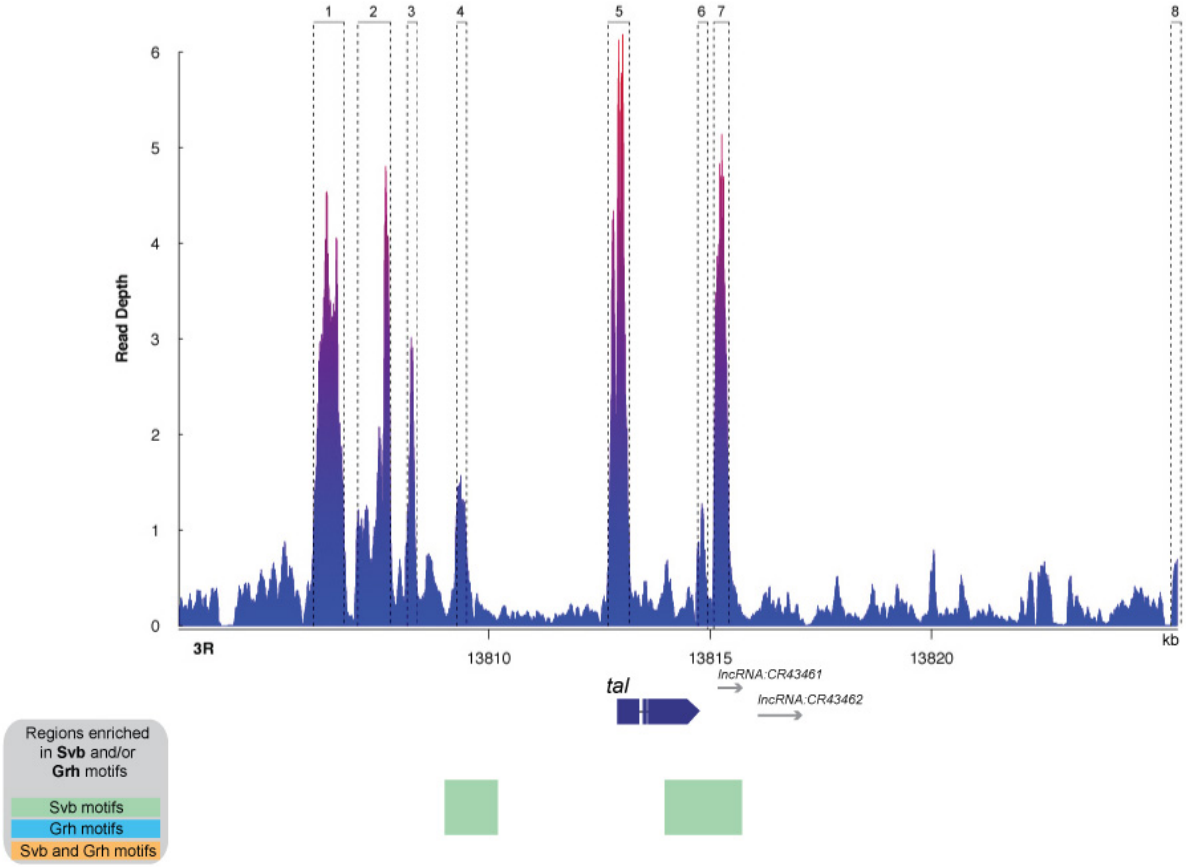


Fig. 42. ATAC-seq profile for *tal* analysis region in T2 pupal legs. There are eight peaks in this region as determined by MACS (Kittelmann et al., 2018). Analysis of the peaks with i-*cis*Target revealed two regions that show enrichment for Svb motifs (green boxes).

3.4 Discussion

3.4.1 A set of 12 genes that may form a finely-tuned developmental switch for trichomes

This investigation has identified a set of 12 genes that are predicted to be activated by Svb in embryos but repressed by *miR-92a*, and since they are also expressed in legs they are strong candidates for being regulated by Svb in the leg. These genes may therefore form a switch to decide if trichomes are made or not. In the NV these genes are likely activated by Svb, but in turn repressed by *miR-92a* to ultimately block the development of trichomes. On the more distal part of the femur, the genes are again

activated by Svb, but miR-92a is absent leading to the growth of trichomes (see Kittelmann and McGregor, 2019). This analysis identified new candidates in the leg trichome GRN and also defined a unique and functionally related set of genes to investigate cis-regulatory logic. To begin to investigate this potential switch further, I tested the function of some of these genes in the leg and explored their cis-regulatory logic. Ongoing experiments by F. Franke, including luciferase assays which aim to functionally verify that these genes are also directly regulated by *miR-92a*.

3.4.2 RNAi analysis reveals several of the candidate genes affect the growth of trichomes

To first assess if the genes by themselves have an effect on trichome development, I performed RNAi knockdown of each gene using the *svb* enhancer identified in (Kittelmann et al., 2018). This line, also referred to as VT057077, drives throughout the second leg, as well as other tissues, in the window of developmental time when trichomes are specified. For *CrebA*, RNAi knockdown using VT057077 was lethal, therefore a HS-GAL4 was used (cross performed by F.Franke). Out of the 7 genes tested three had a significant detectable effect on the length of the trichomes, *CG14395*, *CrebA* and *mwh*. In the case of *CG14395* and *CrebA* the effect on the trichomes was only seen in the trichomes proximal to the NV. Since different drivers were used it is unlikely that both drivers are stronger proximal to the NV, and so perhaps the trichomes proximal to the NV are more susceptible to the level of these genes than more distal trichomes. This could perhaps be tested further using additional drivers.

There is evidence to suggest that mutations in the gene *CrebA* produce patterning defects in the denticle belts on the *Drosophila* larval cuticle (Abrams and Andrew, 2005; Andrew et al., 1997). Taken together with the significant effect *CrebA* RNAi had on the trichomes on the adult leg, this strongly suggests that *CrebA* is also involved in the GRN for leg trichome development. However, further experiments are required to test where *CrebA* is positioned in this GRN. Furthermore, *CrebA* encodes a TF and perhaps it has a detectable effect on trichomes because it regulates multiple trichome genes and so it would be interesting to identify its direct targets.

Knockdown of *mwh* had a more severe effect on the trichomes, with defects in length and polarity observed all over the second leg and the wing. This suggests that *mwh* is a key gene in the development and patterning of trichomes on multiple parts of the adult cuticle. The role of *mwh* in trichome organization in the Drosophila pupal wing has been studied before and it has been determined that *mwh* acts downstream of the planar cell polarity effector (PPE) genes (Wong and Adler, 1993) Similar to what was seen in the leg and the wing with *mwh*-RNAi, mutations in *mwh* display aberrant trichome polarity and multiple trichomes per wing cell (three or more) (Strutt and Warrington, 2008; Wong and Adler, 1993; Yan et al., 2008). To my knowledge, this is the first time an effect of *mwh* loss has been studied on the leg and it will be interesting to see where this gene fits in to the leg trichome GRN and how it is regulated in comparison to the wing.

The remaining genes that were analysed using RNAi in this study did not show any significant effect on the length of the trichomes, and there was also no discernable effect on the polarity or number of trichomes, however, this does not mean that these genes are not involved in the patterning of trichomes on the second leg. It is possible

that when tested individually these genes are not capable of altering the morphology of the trichomes or that the effects are too subtle for detection, but perhaps knocking down combinations of these genes would show an effect on the trichomes. Moreover, for some of the lines the RNAi may not have been effective enough to elicit a phenotypic effect or simply not have worked. Therefore in the future to test the role of these genes it may be more effective to use loss-of-function mutations or clonal analysis.

I did expect to see an effect for some of the genes tested, for example, *Rab23* has previously been shown to belong to the planar cell polarity genes (PCP) along with *mwh* (Pataki et al., 2010). Mutant alleles of *Rab23* have shown similar phenotypes to the one seen for *mwh* RNAi in this investigation including aberrant trichome orientation and number, on various adult cuticular structures including the wing, abdomen and leg (Pataki et al., 2010). Perhaps my result can be explained by the RNAi not being effective enough in the case of this gene.

I also expected to see a phenotypic effect with the knockdown of *forked*, as this gene has previously been implicated in the development and morphology of the trichomes in the larval stage (Chanut-Delalande et al., 2006). Further testing of the genes with different leg drivers and in combination with other genes, as suggested above, is required to rule out any involvement in trichome development.

3.4.3 Overall motif enrichment reveals different motif combinations in open vs closed chromatin

The motif enrichment search for the 12 genes found some intriguing results for 'open' vs 'closed' chromatin. The key finding was that the open chromatin was highly enriched in Svb/Ovo and Grh motifs consistent with it being likely they are regulated by these TFs.

The TF Grh has been shown to have a general role in epidermis differentiation (Mace, Pearson and McGinnis, 2005). In particular, Jacobs et al. (2018) discovered a role for Grh in mediating chromatin accessibility, specifically in epidermal genes in *Drosophila*. By using a combination of ATAC-seq and DNA footprinting analysis it was determined that wherever a region that contains a Grh motif is accessible, Grh is stably bound there (Jacobs et al., 2018). Moreover, mutations in the Grh binding sites abolished enhancer activity in reporter constructs, which was determined to be a result of a change in the accessibility of the chromatin (Jacobs et al., 2018). It appears that Grh is able to change the chromatin landscape by displacing nucleosomes making it a pioneer TF for the *Drosophila* epithelia. Therefore, it is promising that I also find high enrichment of Grh in the peaks of open chromatin for the 12 epithelial genes tested, as presumably without Grh these regions would not be accessible. To further bioinformatically determine if Grh is bound *in vivo* to these regions, a search for di- and tri- nucleotide repeat sequences could be indicative of a functional Grh site, as these have shown to create a 'favourable' sequence environment for *in vivo* Grh binding (Jacobs et al., 2018).

3.4.4 Prediction of epidermal trichome enhancers

I found putative Svb-dependent binding sites in most of the 12 genes analysed often in regions of open chromatin also enriched for Grh motifs. Such regions are candidates for enhancers that drive leg expression and perhaps as well in other tissues. Therefore, I used the results from i-*cis*Target to assess the presence of Svb and/or Grh motifs in specific peaks for each of the 12 genes. It should be noted that due to the nature of the i-*cis*Target program, the presence of motifs is analysed for CRRs that overlap by 40% to the user-supplied data (Herrmann et al., 2012; Imrichová et al., 2015) for further information). This may mean that some of the regions identified as separate CRRs may not actually represent separate enhancers *in vivo*, but this can be investigated further with functional testing either at the endogenous locus using CRISPR or with reporter constructs.

For several of the genes I found some excellent candidate enhancers that are enriched in both Svb and Grh to be followed up. For example, there is one region for *Rab23* that shows high enrichment for both Svb and Grh (Fig. 39) and there are two regions in *mwh* that would definitely be interesting to functionally test given the extreme phenotype produced by the RNAi. There is an excellent candidate region in *forked* (Fig. 36), and several regions that can be tested for *CrebA*, again a gene that had an effect on the growth of the trichomes. *CG14395* also had a significant effect on the growth of trichomes and there are two regions in this gene that are good candidates for enhancers (Fig. 34). Finally there is also one candidate region in *CG4678* that should be functionally tested.

As well as regions enriched for both Svb/Ovo and Grh, I also identified regions that are only enriched for Svb/Ovo, it would be interesting to test these regions to

determine if they can drive in the legs and/or if they are Svb-dependent embryonic enhancers. Along with testing putative enhancers containing Svb and Grh motifs, this would test the hypothesis that Grh is only required to induce the accessibility of epidermal enhancers later in development and perhaps Svb alone is sufficient to activate these trichome genes in embryos. One piece of evidence that supports this hypothesis already comes from Menoret et al., (2013) who did not find Grh enriched in their set of embryonic genes downstream of Svb, in contrast to my analysis of genes expressed in pupal legs.

I also identified regions that are only enriched for Grh motifs. If these represent active Grh sites it may be that these are enhances are bound by other TFs for Svb-independent regulation of these genes during epidermal differentiation or involved in the regulation of expression in other tissues. Given that most of the 12 genes tested are highly pleiotropic (Table 2) it would not be surprising that they have other enhancers involved in the regulation of other aspects of development. This theory could be tested further by first assessing where these regions drive expression of reporter constructs, then further analyzing which other TFs might be bound at these regions perhaps using a yeast-one-hybrid assay (Hens, Feuz and Deplancke, 2012). The results of this analysis may determine if the presence of Grh is generally a reliable signature of epidermal enhancers and could therefore be used in future predictions.

Chapter 4

Exploring the evolution of hunchback enhancers

4.1 Introduction

Recently there has been growing interest in a more in depth analysis of individual enhancers in order to uncover their molecular signatures and activities (reviewed in Borok et al., 2010) . Key to understanding how individual enhancers work and more broadly the general principles of enhancer is to investigate details such as which TFBS of what affinities are responsible for functional activity, as well as how the TFBS are organised – their syntax (Buffry et al., 2016).

However, despite these studies detailed understanding of functionality remains elusive, for the majority of enhancers we do not fully understand the required syntax, nor how turnover results in different combinations of binding sites that evolve to drive the same expression patterns (see Chapter 1). For example, for one of the best studied enhancers to date, *eve* stripe 2, (see Chapter 1) no synthetic enhancers generated have been able to recapitulate the tightly controlled spatial and temporal expression driven by the endogenous enhancer, even when integrating all of the current knowledge of TFBS function, syntax and evolutionary conservation (Barr et al., 2017; Crocker and Ilsley, 2017; Vincent, Estrada and DePace, 2016). This is also the case for other enhancers (Johnson et al., 2008). This strongly suggests that there are gaps in our knowledge of enhancer functionality and it is clear that part of the solution is that we must thoroughly investigate the functionality of individual enhancers – even those we think we understand - to decipher all their components that work together to control

gene expression.

With the development of tools, including CRISPR/Cas9 and live-cell imaging, there are now many more precise and quantitative ways to investigate enhancer functionality. In this investigation I aimed to apply existing tools and develop new approaches to investigate enhancer functionality by utilising the natural variation produced by evolution to understand the crucial elements in a given enhancer, and where flexibilities and constraints lie within enhancer sequences.

Due to the vast amount of information already available, I chose to focus on the P2 enhancer of *hunchback* (*hb*). In *D. melanogaster* and other higher dipterans zygotic expression of *hb* is required for the development of anterior segments and A7/A8 (Lehmann and Nusslein-Volhard, 1987; Tautz et al., 1987). *hb* mRNA is initially deposited maternally and is distributed throughout the early egg (Lehmann and Nusslein-Volhard, 1987). However, *hb* translation is inhibited in the posterior by *nanos*, creating a gradient of expression from anterior to posterior (Irish, Lehmann and Akam, 1989). After fertilisation, the maternal *hb* input is replaced by zygotic expression which reaches from the anterior to about 50% egg length (EL) and also includes some expression in the posterior cap of the embryo, and subsequently this expression resolves to an anterior and a posterior stripe (Schroder et al., 1988; Tautz, 1988; Tautz and Pfeifle, 1989).

The *hb* locus contains three distinct enhancers each responsible for mediating parts of the expression pattern (see Chapter 1 – Fig. 2) (Perry et al., 2012). The proximal and distal enhancers mediate activation in response to the Bicoid gradient (Bcd) and are responsible for driving the anterior expression through nuclear cycles (nc) 12-13 through the P2 promoter (Driever and Nusslein-Volhard, 1989; Perry,

Boettiger and Levine, 2011; Struhl, Struhl and Macdonald, 1989). Later in development, the anterior and posterior stripes of expression are regulated by the stripe enhancer through both the P1 and the P2 promoters during nc14 (Perry, Boettiger and Levine, 2011).

Observations using live-quantitative imaging have revealed that the primary and distal enhancer do not function in a simple additive manner, but in fact the functional integration of the two enhancers to drive expression of *hb* is more complex (Bothma et al., 2015). In the anterior region of the embryo where the Bcd concentration is high, the primary and distal enhancer function sub-additively, in contrast, where the level of Bcd is lower, further towards the posterior, the enhancers appear to function additively (Bothma et al., 2015). Furthermore, it appears that loss of the distal enhancer does not affect the level of transcription in the anterior region of the embryo, however, the proximal enhancer alone cannot produce endogenous levels of expression in the absence of the distal enhancer at low Bcd levels (Bothma et al., 2015). This suggests that one of the roles of the distal enhancer is to contribute additional sensitivity to lower concentrations of Bcd as found in the centre of the embryo (Bothma et al., 2015). These findings demonstrate that the relationship between the proximal and distal enhancers in the *hb* locus is complex. To understand the functionality of individual enhancers perhaps we must experimentally investigate their activity at the endogenous locus, where enhancers combine to drive gene expression that may not be captured by examining individual enhancers driving reporter genes at exogenous locations.

To functionally analyse natural variation within an enhancer I chose to further study the P2 proximal enhancer (referred to throughout this Chapter as P2). Our

current knowledge on the P2 suggest that it is dependent on the TF Bcd and seven Bcd sites have been functionally verified in *D. melanogaster* (Driever and Nusslein-Volhard, 1989; Ma et al., 1996; Struhl, Struhl and Macdonald, 1989) (Fig. 44). The P2 also contains two sites for the pioneer factor Zld, one of which is situated in the minimal enhancer region of ~252 bp and one placed slightly more towards the promoter (Xu et al., 2014) (Fig. 44).

Previous studies have used reporter constructs and genetic methods to examine the effect of deleting individual Bcd binding sites or using different combinations of individual or clusters of binding sites (Driever and Nusslein-Volhard, 1989; Ma et al., 1996; Driever, Thoma and Nusslein-Volhard, 1989; Struhl, Struhl and Macdonald, 1989). It is thought that the function of the P2 enhancer relies on cooperative binding of Bcd to both high affinity and low affinity binding sites (Burz et al., 1998; Driever, Thoma and Nusslein-Volhard, 1989; Hanes et al., 1994; Yuan, Ma and Ma, 1996; Yuan, Ma and Ma, 1999; Zhao et al., 2000b). However, recent analysis suggests there are other mechanisms required such as higher order cooperativity or some form of energy expenditure that explains the expression profile driven by the P2 and the influence of the Bcd binding sites (Park et al., 2019).

There has been extensive evolutionary turnover of the Bcd-binding sites among *Drosophila* species and other higher dipterans (Lukowitz et al., 1994; McGregor et al., 2001) (Fig. 3). However, these variant sequences with differences in the binding site nucleotides, binding site spacing and number of sites are generally still able to drive the same expression patterns (Lukowitz et al., 1994; McGregor et al., 2007). Furthermore, the P2 sequence from *D.virilis* has been shown to rescue *hb* expression in *D. melanogaster*, suggesting these enhancers are functionally equivalent (Lukowitz

et al., 1994). Given that in *Drosophila* species the Bcd homeodomain has remained highly conserved, suggesting there is no co-evolution between the P2 and its TF (McGregor et al., 2001), the question remains as to how changes in these enhancers can be tolerated to maintain the same expression pattern and function.

Despite our current understanding of the P2 proximal enhancer we still do not fully understand the functionality of this cis-regulatory element and as discussed previously, attempts to construct synthetic P2 elements have failed to recapitulate the expression driven by the endogenous enhancer (Driever, Thoma and Nusslein-Volhard, 1989; Park et al., 2019; Struhl, Struhl and Macdonald, 1989). In this investigation I aimed to exploit the natural variation found in the P2 enhancer and new technologies to create tools to further decipher and analyse the functionality and evolution of the P2.

4.2 Methods

4.2.1 Sequence alignment

Genomes for the species used in the alignment were downloaded from FlyBase (*D. melanogaster* r6.18, *D. yakuba* r1.05, *D. erecta* r1.05, *D. eugracilis* r2.0, *D. biarmpies* r2.0, *D. elegans* r2.0, *D. ficusphila* r2.0, *D. ananassae* r1.05, *D. pseudoobscura* r3.04, *D. virilis* r1.06). The sequence for *D. melanogaster hb* P2 was extracted from FlyBase (r6 3R:88694400..8694900). BLAST+ (Camacho et al., 2009) was used to search P2 enhancers in other species using *D. melanogaster* as the query sequence. Coordinates for the P2 enhancer for each species were then extended to cover ~2 kb upstream and ~2 kb downstream, giving a ~4 kb fragment, this included the first exon of the *hb* gene to be included in the alignment to anchor the sequences and produce a more

reliable alignment of upstream non-coding DNA. Sequences were extracted using Bedtools (Quinlan and Hall, 2010) and aligned in UGENE (Okonechnikov, Golosova and Fursov, 2012) using MAFFT algorithm (Katoh et al., 2002). The alignment was then trimmed to include just the *hb* P2 region and binding sites A1, X1, X2, X3, A2, A3 and X4 were identified in *D. melanogaster* using sequences identified by (Driever, Thoma and Nusslein-Volhard, 1989; Ma et al., 1996; Struhl, Struhl and Macdonald, 1989). The Zelda sites were identified using data from (Xu et al., 2014).

4.2.2 Fly stocks

All stocks used were kept on standard yeast extract-sucrose medium at 25˚C. Fly stocks for different wild-type *Drosophila* species were either ordered from Kyoto DGGR or in-house stocks were used: *D. melanogaster* – OreR (in-house), *D. virilis* – MT2011 (in-house), *D. yakuba* – 0261.01 (DGRC 921201) and *D. pseudoobscura* – 0121.94 (DGRC 920401). The yw;His2Av-mRFP;Pnos-MCP-EGFP stock was ordered from Bloomington stock centre (#60340). The P2 promoter and enhancer deletion line was made by J.Ling and kindly provided by Steve Small (NYU).

4.2.3 Molecular cloning

Three species were initially chosen to determine if their P2 could functionally replace the *D. melanogaster* P2 and for live imaging analysis: *D. melanogaster* was chosen as a control, *D. yakuba* was chosen due to exhibiting several mutations in TFBS despite being evolutionary close to *D. melanogaster* and finally *D. virilis* was chosen as the most distantly related species with the most divergent P2 sequence. Primers were designed for each species to PCR the P2 enhancer and the P2 promoter (based on the

alignment) from gDNA. To clone the P2 sequences into the rescue construct a restriction enzyme approach was taken using *Hin*dIII restriction sites, replacing the GFP in the plasmid backbone with the P2 enhancer + promoter flanked by two attB sites. To clone the same regions into the MS2 plasmid (kindly provided by Thomas Gregor, Princeton) a restriction enzyme approach was also taken, using unique restriction sites to insert the P2 enhancer and promoter upstream of the 24xMS2 stem loops. The integrity of the loops was checked using both sequencing and diagnostic restriction digests. Constructs were sent to Cambridge Injection Facility for phiC31 mediated injection into the P2 enhancer + promoter deletion line (kindly provided by Steve Small, NYC) (Fig. 43).

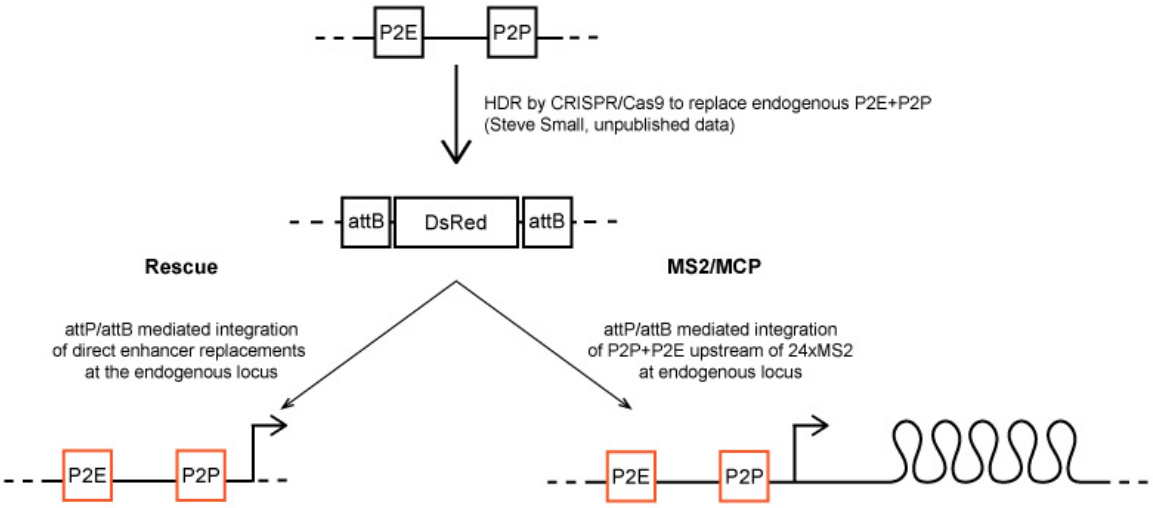


Fig. 43. Experimental design for both the rescue and the MS2 experiments. HDR CRISPR/Cas9 was performed on to replace the endogenous P2 enhancer + P2 promoter with DsRed flanked by two attB sites (Steve Small, unpublished data). This allows any cassette flanked by two attP sites to be exchanged with the dsRed. For the rescue experiment, the P2E+P2P from other species (outlined in red) can be exchanged into the endogenous locus using phiC31 mediated integration. For the live imaging, the P2E+P2P from other species can also be integrated into the endogenous locus but this time with the addition of 24x MS2 stem loops to give a quantitative read out of live *hb* transcription driven by the P2.

4.2.4 MS2 system

To investigate the level and timing of transcription driven by the P2 I used the MS2-MCP system (Bertrand et al., 1998). This method involves the tagging of nascent transcripts of mRNA with multiple repeats of a stem loop sequence that is recognised by a cognate binding protein. The binding protein is constitutively expressed and bound to a fluorescent protein that can be visualised using confocal or two-photon microscopy (Bertrand et al., 1998; Lionnet et al., 2011; Yunger et al., 2010). The stem loops (MS2) bind to the fluorescent tagged binding protein (MCP-GFP), which results in spatially localised fluorescent expression at the locus of interest. This is further enhanced by each additional polymerase that engages in transcriptional elongation (Garcia et al., 2013) .This method has already been adopted in the fly embryo for the study of various processes, including the rate of transcription driven by enhancers (Bothma et al., 2014; Bothma et al., 2015; Ferraro et al., 2016; Forrest and Gavis, 2003; Fukaya, Lim and Levine, 2016; Garcia et al., 2013; Lucas et al., 2013). This system allows the real-time dynamics of transcription to be observed as well as giving quantification of the absolute number of transcribing Pol II molecules and numbers of produced mRNA molecules (Gregor, Garcia and Little, 2014)

4.2.5 Sample preparation and live-imaging

Male flies containing the MS2 step loops downstream of the P2 enhancer and promoter were crossed to females containing His-RFP;MCP-GFP. Flies were allowed to lay in cages with apple juice agar plates for 90 mins, and embryos were then collected from the plates and dechorionated in 50% household bleach for 2 mins.

Specialist slides for live imaging (based on the protocol by Garcia and Gregor (2018) were prepared by inserting a breathable Lumox Film (Sarstedt 94.6077.317) into a custom made slide holder. Heptane glue (made in house by rolling double-sided tape with heptane) was applied the film to allow the embryos to stick. Dechorionated embryos were placed on the film in anterior-posterior orientation. Embryos were then covered with a few drops of halocarbon 10S oil, and finally a cover slip was placed on top of the embryos. Live imaging was carried out at Oxford Brookes University Bioimaging facility on a Zeiss 880 confocal microscope using 488 nm and 561 nm laser excitation lines and 40x objective. Z-stacks were performed on expressing embryos with ~ 0.5um separating each slice. In the cases where whole embryos were imaged, this was achieved using tile scan with a 5% overlap between the tiles. Videos were taken using z-stack continuous acquisition, and the embryos were monitored regularly to account for drift. For the purpose of this thesis, still images have been extracted from the videos.

4.3 Results

I generated a multiple species alignment of the P2 from 10 *Drosophila* species encompassing approximately 50 MYA of evolution. This revealed extensive turnover in the P2 proximal enhancer with respect to the seven Bcd binding sites and the Zld binding site in the *D. melanogaster* enhancer.

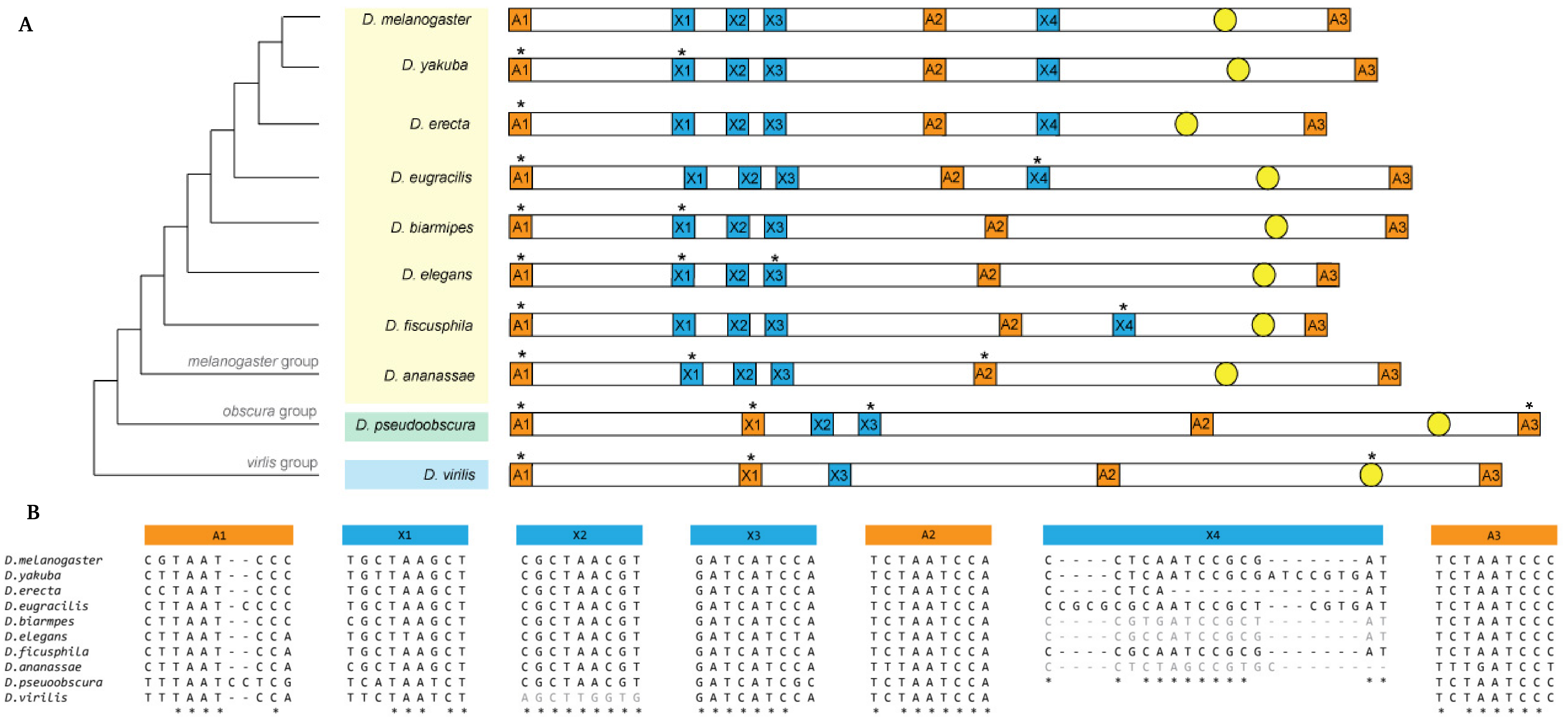


Fig.44. (A) Scematic showing multiple species alignment of the P2 enhancer of *hb*. Species are shown on the left with evolutionary relationships from Seetharam and Stuart, (2013). Binding sits shown are for the TF Bcd with respect to *D. melanogaster*, orange sites also labelled as 'A' are high affinity Bcd sites, the blue sites also labelled as 'X' represent low affinity Bcd sites. The Zelda site is shown in yellow. If there has been a mutation in a TF binding site this is represented by an asterix on top of the binding site. The length of the enhancers are to scale with changes in spacing between binding sites representing changes in the number of nucleotides seen between the sites in the alignment. (B) A closer examination of the mutation within individual binding sites in the P2 enhancer. Again high affinity sites are represented in orange and low affinity sites are represented in blue. Sequences that are grey or missing highlight that this binding site is likely not present in this species. As explained in the Methods, all sequences are taken from a much larger multiple species alignment and trimmed down to just focus on the P2 region.

I33 focused on the sites that have been confirmed by functional assays in previous studies (Driever, Thoma and Nusslein-Volhard, 1989; Ma et al., 1996; Struhl, Struhl and Macdonald, 1989). However, note that preliminary data from Thomas Gregor (unpublished) suggests that there are actually more Bcd binding sites in the *D. melanogaster* P2 and when analysing the P2 sequence from other species I was also able to recognise additional occurrences of the Bcd consensus motifs, suggesting the presence of additional sites in other species. These additional predicted sites first require functional validation so therefore, I have concentrated on the known 7 Bcd binding sites for this analysis.

The A1 Bcd binding site was recognisable in all the species analysed, however in comparison to *D. melanogaster*, the site has accumulated several mutations in some species (Fig. 44). Despite these mutations the recognised Bcd consensus binding motif, 'TAATCC' (Noyes et al., 2008; Treisman et al., 1989) is still maintained in all species – even the distantly related *D. virilis*. The nucleotides between A1 and the next site X1 do not change much in terms of length, until *D. pseudoobscura* and *D. virilis* where there is a substantial increase in the number of nucleotides separating these two sites (~15 nucelotides).

Binding site X1 is also recognisable in all the species used for the alignment, however it has accumulated sequence changes with respect to *D. melanogaster* in several species, and in the case of *D. pseudoobscura* and *D. virilis* (TAAT) the differences would be predicted to make this a higher affinity site than in *D. melanogaster* (TAAG) (Fig. 44) based on what is known about Bcd consensus binding motifs (Noyes et al., 2008; Treisman et al., 1989) although this remains to be functionally tested.

Strikingly, binding site X2 has maintained the same sequence throughout all of

these species, except *D. virilis* where it appears to either have been lost or it evolved after the divergence of this species from the others (Fig. 44). Binding site X3 is also conserved in the species analysed, and has only accumulated a couple of mutations in total (only in *D. elegans* and *D. pseudoobscura*) (Fig. 44).

Apart from *D. virilis*, the most distantly related species used, the configuration the three low affinity sites, X1, X2 and X3 is strongly conserved. These three sites always seem to appear in the same order and with very similar spacing between the sites, perhaps highlighting the importance of these core sites for the function of the P2 (Burz et al., 1998).

The next binding site, A2, is conserved throughout the focal species in the alignment and is easily recognisable as a high-affinity Bcd site (Driever, Thoma and Nusslein-Volhard, 1989). In fact, the A2 site has only accumulated one mutation according to the alignment and this is in the species *D. ananassae*.

Binding site X4 is often not included in studies on the P2 enhancer of *hb*, but it was shown by (Ma et al., 1996) to be a functional Bcd site later than the other *D. melanogaster* sites, so I have included it this analysis. X4 is perhaps the least well-defined site and this is reflected in the alignment, as it is clearly missing in several species (Fig. 44). Interestingly, the spacing between X4 and A3 has changed substantially, for example between *D. ananassae* and *D. pseudoobscura* the distance between A3 and X4 differs by ~40 nucleotides).

In between X4 and A3 is a Zld site, which is important for ensuring that the P2 is accessible (Xu et al., 2014), and therefore not surprisingly this site is highly conserved. Indeed only one species, *D. virilis*, shows a different canonical Zld site (also see Fig. 45).

The final site in the P2 is the A3 binding site and this is strikingly conserved in all the species in the alignment, with the exception of *D. ananassae* where it appears to have changed from a high affinity site to low affinity (Driever, Thoma and Nusslein-Volhard, 1989). This conservation of A3 highlights that this site is important to the function of the P2 enhancer as indicated by previous functional studies (Hanes et al., 1994).

Finally, when examining other nucleotides in the alignment data (Fig. S4) it is clear that it is not only the characterised *D. melanogaster* binding sites that have accumulated mutations, but there is also extensive turnover in the inter-binding site nucleotides, which may also affect P2 activity through the gain and loss of other uncharacterised Bcd binding sites or binding sites potentially for other TFs. However, understanding the consequences of this turnover first requires detailed analysis and comparison of the activities of this enhancer from different species in vivo.

4.3.1 Analysis of Zelda sites in the P2 enhancer

It has previously been shown that Zld is required for Bcd to bind to Bcd-dependent enhancers, such as the *hb* P2, and the presence of Zld sites allows for the correct activation of target genes leading to coordinated development (Xu et al., 2014). Given the importance of Zld for the P2 enhancer, and the different Zld sites present in the P2 of *D. virilis*, I decided to look at more *Drosophila* species to investigate if there is further variants of the eight characterised canonical Zld motifs (Liang et al., 2008; Nien et al., 2011; ten Bosch, Benavides and Cline, 2006) (Fig. 45).

As mentioned above, there are two Zld sites within the P2 region of *D. melanogaster*, the first is located between binding sites A2 and A3 (Fig. 44) and the

second is located after binding site A3, further towards to minimal promoter. Note that this region is not shown in Fig. 44, but it is included in the MS2 constructs presented below. The majority of the species analysed contain the same two Zld sites, CAGGTAG and CAGGTAC (Fig. 45). However there are a few species where the sites differ. *D. persimillis*, and *D. willistoni* as well as *D. virilis*, all also have CAGGTAG, but the other site is CAGGTAT (Fig. 45). In addition, for two of the species studied, *D. mojavensis* and *D. grimshawi*, I could only identify one Zld site matching to the CAGGTAT consensus (Fig. 45). The consequences of different Zld sites on transcription have not yet been investigated.

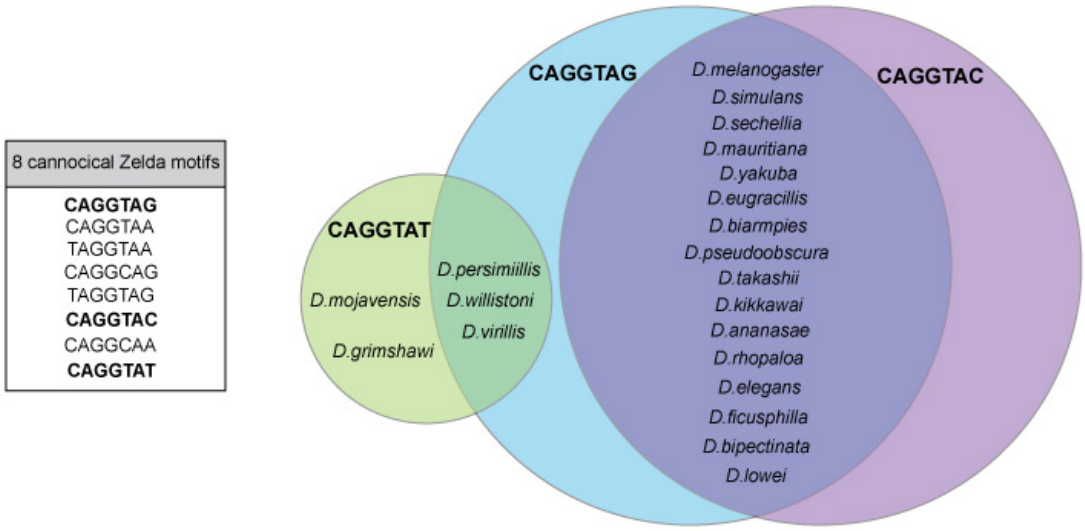


Fig. 45. Analysis of the usage of canonical Zld sites in the P2 enhancer of *hb* in 21 different *Drosophila* species. (A) 8 canoncial Zld motifs (ten Bosch et al., 2006; Liang et al., 2008 ; Nien et al., 2011) with sites in bold those found in the P2 enhancer of *hb*. (B) Diagram showing usage of the three sites found for each species. Most species seem to use CAGGTAG and CAGGTAC, with exceptions to this show in the green circle.

4.3.2 Testing the functional equivalence of P2 enhancers

I next investigated whether the P2 enhancers from the other species assessed in the alignment analysis can functionally replace the endogenous P2 enhancer in *D. melanogaster*. The endogenous P2 deletion line from Steve Small is homozygous lethal, displaying a classic *hb* mutant larval cuticle phenotype (Nusslein-Volhard and Wieschaus, 1980). In this line the *hb* P2 enhancer and promoter were replaced with two attP sites flanking a dsRed marker cassette (Fig. 43). Therefore, I decided to clone the P2 enhancer and promoter from *D. melanogaster*, *D. yakuba*, *D. virilis* and *D. pseudoobscura* from the multiple sequence alignment and use phiC31 mediated transgenesis to insert them into the endogenous deletion line. If the rescue was sufficient the F1 progeny should be viable without a balancer chromosome when crossed together.

However, the results of this rescue assay did not show what was expected. Given that we already know that *D. virilis* P2 should be functionally equivalent to that of *D. melanogaster* (Lukowitz et al., 1994), I expected to see a full rescue for the other species. I did not observe the loss of balancer for any of the insertions and after careful consideration, I belatedly discovered that the plasmid used for the rescue assay, piB-GFP, is unsuitable for this experiment. The cloning strategy for this construct involves removing the GFP and replacing it with the P2 enhancer + promoter using restriction enzymes. When this is inserted into the fly using attP/attB integration it places the insert just upstream of the endogenous *hb* start codon (Fig. 46). However, there is a portion of plasmid backbone now placed between the P2 promoter and the start codon of *hb*, and analysis of this sequence revealed there is a plasmid derived start codon in this sequence (Fig. 46). Therefore, the lack of rescue is probably because this sequence

potentially interferes with *hb* expression. Therefore unfortunately it remains unclear from this experiment whether the P2 from *D. melanogaster*, *D. yakbua*, *D. pseudoobscura* and *D. virilis* are functional in the endogenous location in *D. melanogaster*. These experiments should be repeated with a better designed construct or using CRISPR/Cas9 mediated homologous recombination.

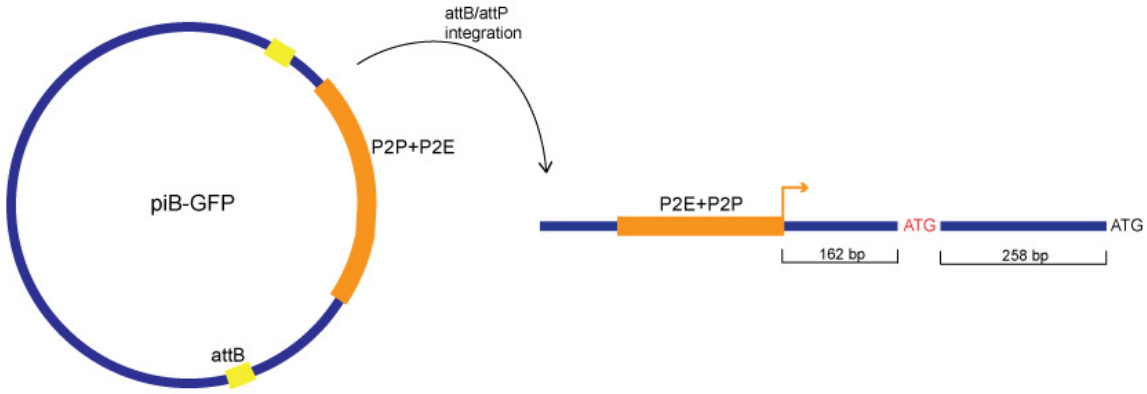


Fig. 46. Schematic of the rescue experiment strategy. When the P2E+P2P from other *Drosophila* species is inserted into the endogenous *hb* locus some of the plasmid backbone is also integrated. This part of the plasmid backbone contains a start codon (red ATG) which becomes situated between the inserted promoter and the endogenous TSS (black ATG), which likely interferes with *hb* expression.

4.3.3 Development and optimisation of tools to study enhancer turnover in P2

We already know that in drosophilids, *hb* has maintained its function and expression pattern but we do not know if the mutations accumulated through evolution have changed the timing of transcription, the levels of mRNA produced and/or the transcriptional boundaries. This information can be provided by using the MS2/MCP system to assess the live transcriptional dynamics driven from P2 variants. Therefore, to be able to fully and thoroughly investigate the evolution and functionality of the P2 enhancer, I decided to generate fly lines to test if the natural variants of the P2

could drive similar expression from the endogenous location in *D. melanogaster*. On one hand, if they could drive similar expression this would suggest that the conserved characterised binding sites as well cis-cis coevolutionary turnover had maintained focal enhancer function and facilitates future exploration of necessary and sufficient components in these enhancers. On the other hand if the natural enhancer variants could not drive similar expression to *D. melanogaster,* this could be a consequence of species-specific changes in the focal enhancer and perhaps co-evolution with Bcd and/or with the distal enhancer.

I chose the P2 enhancers from three different species to clone downstream of the MS2 loops and insert these constructs into the endogenous location of the P2 in *D. melanogaster* (Fig. 43): *D. melanogaster*, *D. yakuba* and *D. virilis*. *D. melanogaster* was chosen as a control while the other species were chosen for their evolutionary distance and the changes seen in the P2 enhancers (Fig. 44). Here I explore the utility of this approach for the questions above and report my initial qualitative analysis of these flies.

4.3.4 *D.melanogaster hb* P2 enhancer and promoter

Flies containing the *D. melanogaster* P2E+P2P upstream of 24xMS2 loops were crossed to MCP-GFP flies and the embryos from the cross were prepared as described in the Methods. Expression was observed in the anterior of the embryo as expected for *hb* (Fig. 47). This expression extended to 40-50% EL which is within the range of endogenous zygotic hb expression at this stage (Perry et al., 2012). Faint expression was also detected in the posterior cap, again as expected (Fig. 47). These results show that this construct is suitable for further future analyse with pipelines (developed by

Thomas Gregor's lab) to quantify the levels of transcription. Unfortunately this was not possible in the time frame of this PhD project.

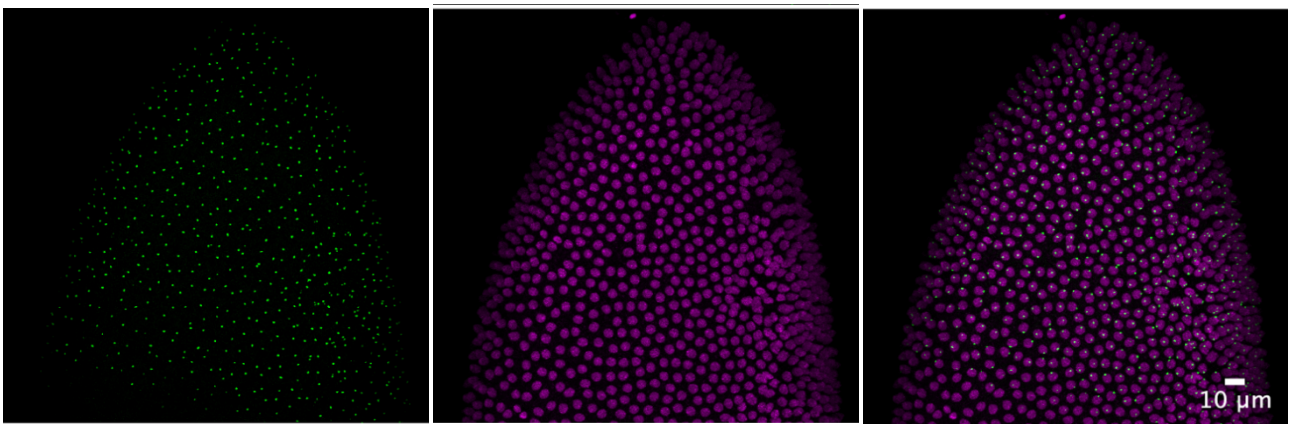


Fig. 47. The anterior of an embryo expressing GFP detected by MS2 loops driven by the *D. melanogaster* P2E+P2P (GFP) (left hand panel) as well as histone RFP (middle panel) at ~nc 13 (the right hand panel shows the overlay). The expression driven by *D. melanogaster* was observed in most, if not all, cells in the anterior of the embryo and extends to approximately 40-50% EL. Anterior is to the top.

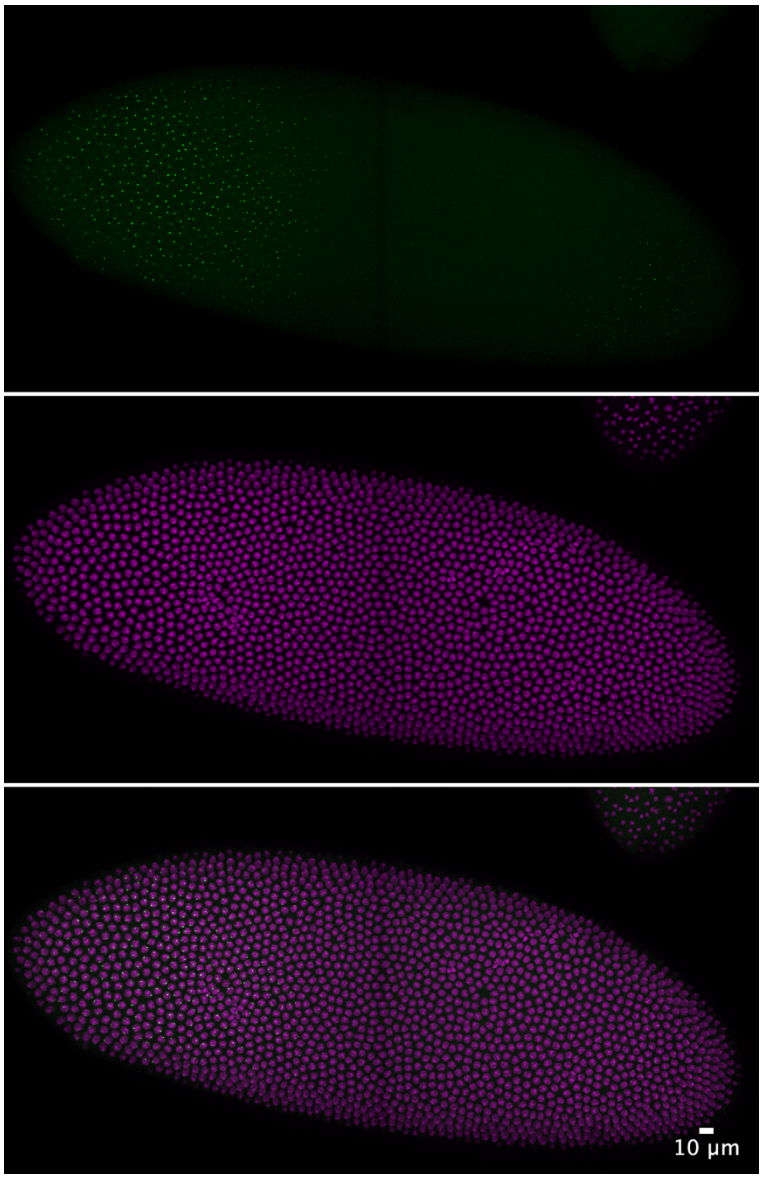


Fig. 48. Whole embryo expressing GFP detected by MS2 loops driven by the *D. melanogaster* P2E+P2P (GFP) (top panel) as well as histone RFP (middle panel) at ~nc 14 (the lower panel shows the overlay).. Expression driven by this construct is consistent we would expect from the endogenous *D. melanogaster* P2: expression in the anterior of the embryo that extends to ~40-50% EL. additionally, there is faint, expression in the posterior cap. Anterior is on the left. Note, the faint line down the centre of the image is due to tile stitching issues in image processing.

4.3.5 *D.yakuba hb* P2 enhancer and promoter

Analysis of *D. melanogaster* embryos expressing the *D. yakuba* P2E+P2P-MS2 and MCP-GFP revealed that the MS2 integration worked successfully and this line can be used in further studies to assess the quantitative level of transcription driven by the *D. yakuba* P2. Expression was observed in the anterior of the embryo as expected. However, qualitative assessment suggests there is less expression both at the anterior and in the posterior cap in comparison to *D. melanogaster* control at this stage (Fig. 49). Furthermore the expression from the *D. yakuba* construct appeared to be more anteriorly restricted than for *D. melanogaster* (Fig. 49). I was able to take longer movies of the embryos for *D. yakuba* P2 and this demonstrated how dynamic the expression pattern is between and within each nc (Fig. 49). Fig 49C shows the capture of the stripe pattern at nc15 that is driven by the stripe enhancer, which makes sense as this enhancer still acts through the P2 promoter.

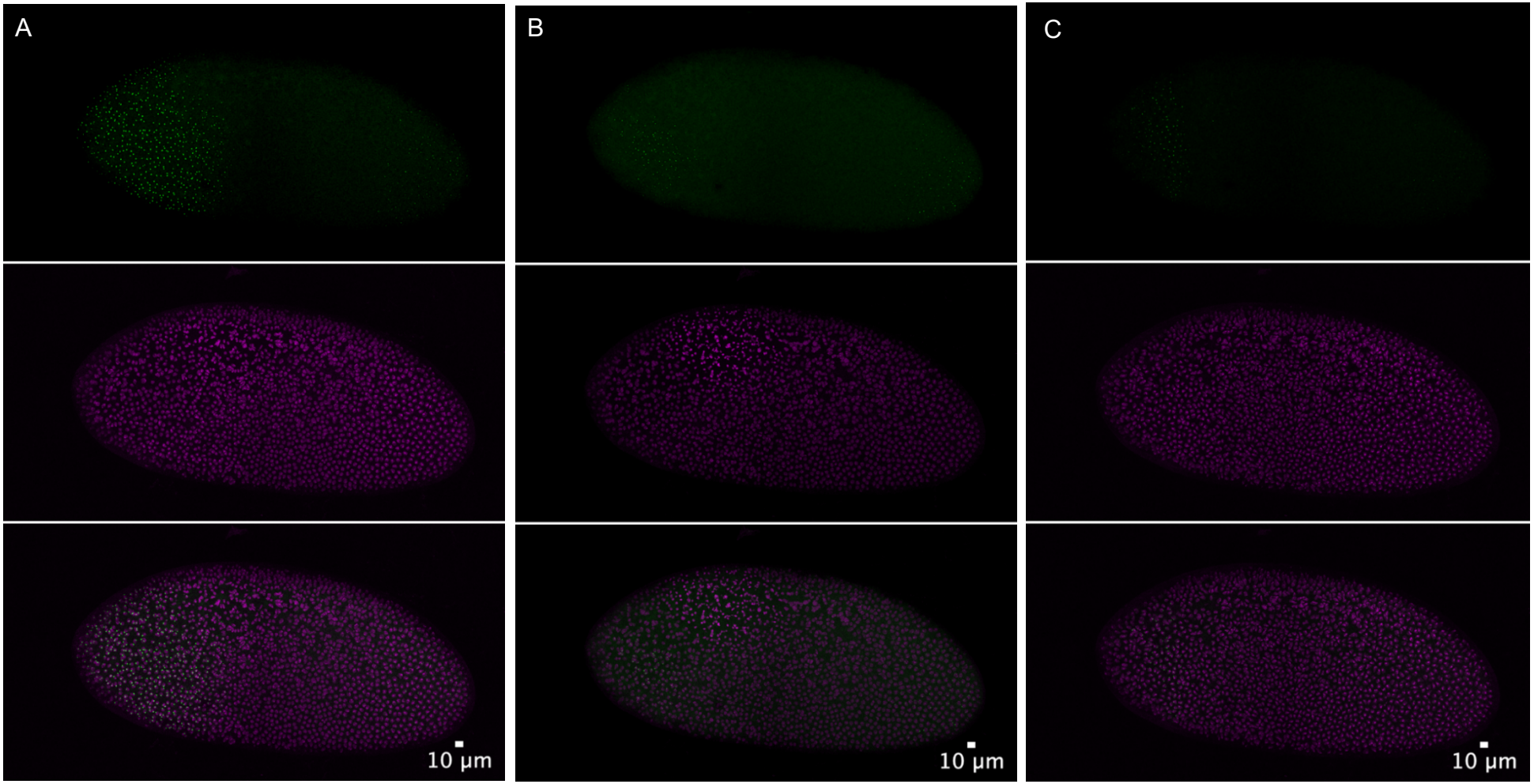


Fig. 49. Whole embryo expressing GFP detected by MS2 loops driven by the *D. yakuba* P2E+P2P (GFP) (top row) as well as histone RFP (middle row) (the lower row shows the overlay) at nc 14 (A) mid-cycle (B) late cycle and (C) nc15. (A) Expression can be seen in the anterior of the embryo as would be expected from the P2 enhancer, but more anteriorly restricted, and there is also faint expression in the posterior. (B) Later nc14 the expression is starting to diminish, both in the anterior and the posterior. (C) Moving into nc 15, the anterior expression seems to be resolving into a stripe.

4.3.6 *D. virilis hb* P2 enhancer and promoter

D. virilis is the most evolutionary distant species chosen for this analysis, but its P2 enhancer is still known to be functionally equivalent to the *D. melanogaster* P2 (Lukowitz et al., 1994). Analysis of embryos expressing the *D. virilis* P2E+P2P-MS2 revealed that the MS2 integration was successful and this has provided another species enhancer tagged with MS2 that can be used in future investigations. Again the expression of *D. virilis* P2 was seen in the anterior of the embryo as we would expect (Fig 51-52) and at later in nc14 expression appears to resolve into a stripe with relatively sharp boundaries (Fig. 52). However, again it appears that expression is more anteriorly restricted than seen for the *D. melanogaster* control. Unfortunately, due to issues with the embryo preparation, I was unable to generate an image of a whole embryos at a later stage from the *D. virilis* construct and therefore cannot say if the posterior cap expression is present (Fig. 50).

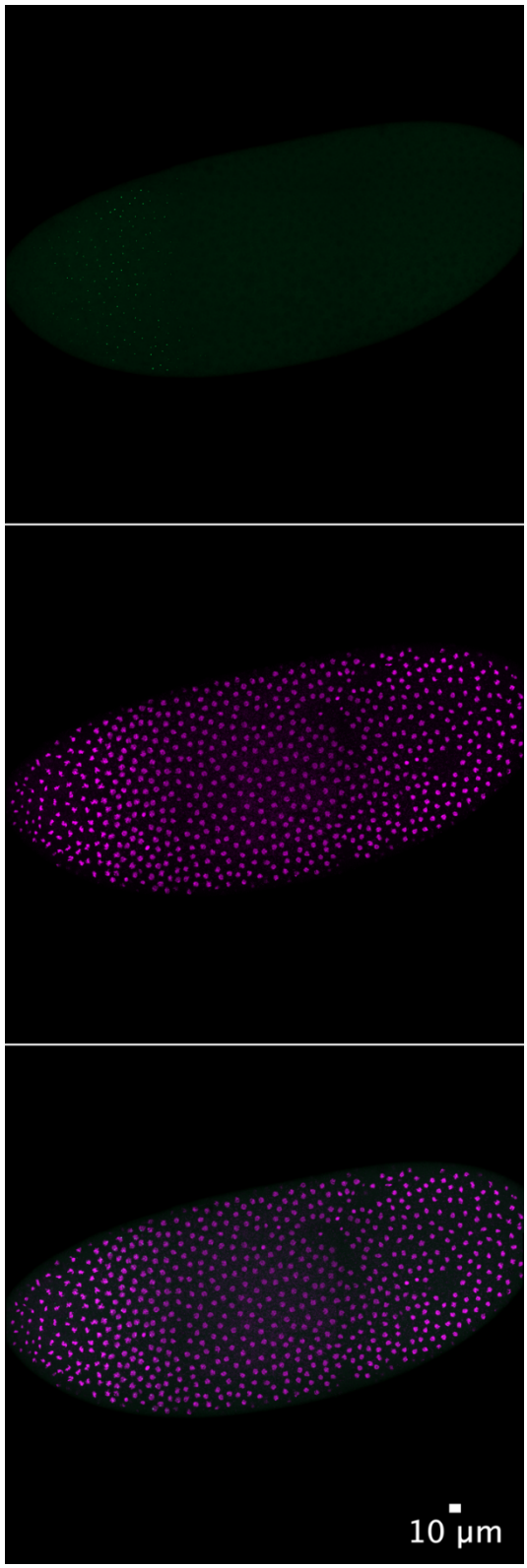


Fig. 50. Whole embryos expressing GFP detected by MS2 loops driven by the *D. virilis* P2E+P2P (GFP) (top panel) as well as histone RFP (middle panel) at nc 10-11 (the bottom panel shows the overlay). Anterior is to the top. Expression driven by the P2 enhancer is just turning on at this stage of embryogenesis with faint expression detected only in the anterior of the embryo.

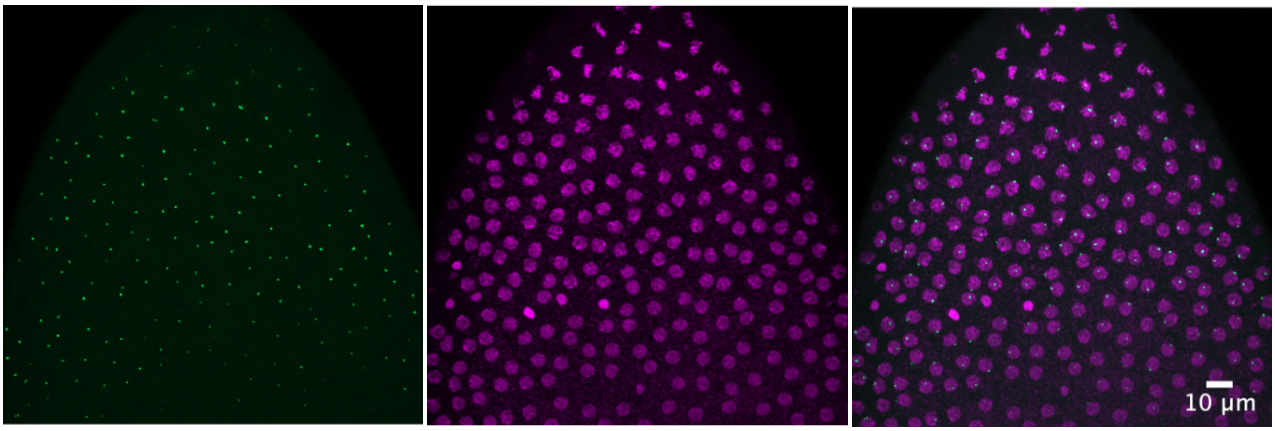


Fig. 51. The anterior of an embryo expressing GFP detected by MS2 loops driven by the *D. virilis* P2E+P2P (GFP) (left hand panel) as well as histone RFP (middle panel) at ~nc 12 (the right hand panel shows the overlay). Expression was observed in most cells in the anterior of the embryo and extends to approximately 40% EL. Anterior is to the top.

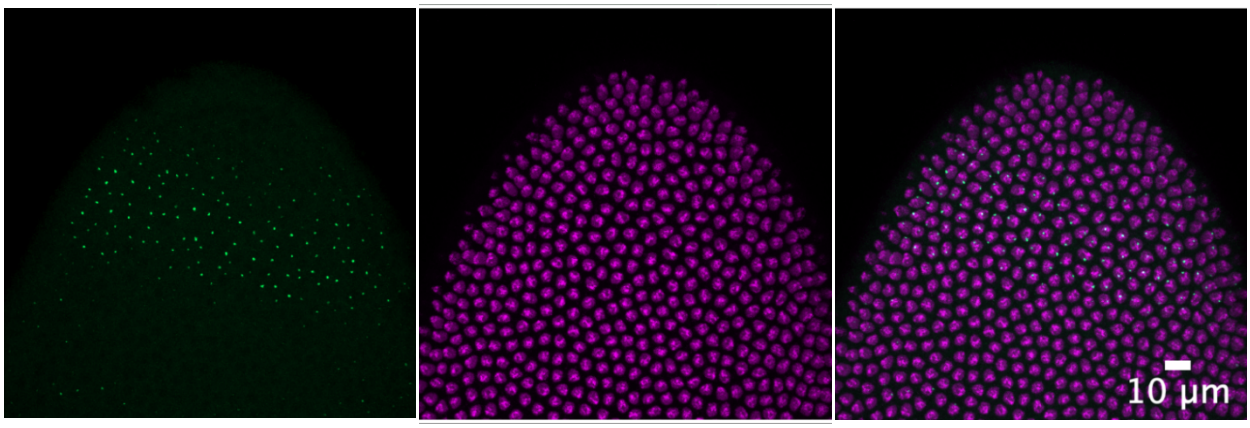


Fig. 52. The anterior of an embryo expressing GFP detected by MS2 loops driven by the *D. virilis* P2E+P2P (GFP) (left hand panel) as well as histone RFP (middle panel) at ~nc 15 (the right hand panel shows the overlay). Anterior is to the top. Again there is expression in the anterior of the embryo, however at this time expression is observed in a sharp stripe as I would expect.

4.4 Discussion

Evaluation of the turnover in the P2 including Bcd and Zelda binding sites

The multiple species alignment for the P2 enhancer revealed a large amount of turnover (Fig. 44), not just in the Bcd binding sites themselves but also in the inter-binding site nucleotides (supplementary fig. 4). Previous studies into the distal enhancer of *hb* have shown that mutations in the known Bcd and Zld binding sites have little effect on the expression pattern of *hb*, however further analysis showed that mutations in the short sequences flanking Bcd binding sites have stronger effects than mutations in the binding sites themselves (Li and Eisen, 2018). The reason for this effect has been explored by (Le et al., 2018) who hypothesised that mutations in flanking sequences change the overall binding energy of TFs by an amount equal to, or greater than consensus site mutations. This suggests that the mutations I observed in the P2 are likely to have an effect on transcriptional output. However, I hypothesise that these changes are somehow being buffered by the system to maintain function and expression of the P2 throughout evolution perhaps by co-evolutionary changes in the P2 itself or in the distal enhancer.

The multiple species alignment also highlighted conserved blocks of DNA in the P2 that are not characterised binding sites for either Bcd or Zld, suggesting, in agreement with other studies (Park et al., 2019), that there are unidentified features in the proximal P2 that are likely to be important for enhancer function. The multiple species alignment created in this investigation provided a platform to subsequently functionally assess the mutations in the characterised P2 binding sites as well as turnover of intervening regions.

Due to unforeseen issues with the rescue assay constructs, the results of this

investigation cannot conclude if the P2 from different species are actually functionally equivalent to *D. melanogaster*, although previous a experiment indicated that the P2 from *D. virilis* can replace the *D. melanogaster* P2 (Lukowitz et al., 1994). However, my preliminary qualitative comparison of the P2 enhancers from *D. melanogaster*, *D. yakuba* and *D. virilis* suggest that the enhancers of the latter two species do not drive expression as strongly or as posteriorly as the *D. melanogaster* P2 enhancer. This could indicate there are functional differences between these sequences caused by differences in binding sites that are compensated for by externally cis-sequence changes in the *hb* locus or in trans - potentially by differences in Bcd concentrations of these species. It could also be that the presence of species-specific promoters is contributing to the variation; this would require further testing of species-specific enhancers with the *D. melanogaster* promoter.

Given these qualitative differences in expression detected for P2 variants (Figs 47-52), it would be important to design new rescue assays perhaps using CRISPR/Cas9 to introduce variants at the endogenous P2 location and test their ability to rescue. I hypothesise however, that all of the species analysed in the alignment (Fig. 44) could be able to functionally replace *hb* P2 from *D. melanogaster* and that this could provide an excellent system to expand on my qualitative analysis using MS2 to investigate enhancer turnover and identify motifs that are necessary and sufficient for endogenous expression.

To my knowledge, this is the first study to apply the MS2 system to compare the homologous enhancer from different species at the endogenous enhancer location. The flies I have generated are valuable tools to study the functionality of the P2 enhancer much further using established pipelines to quantify and compare the the

expression between variants (Bothma et al., 2014; Bothma et al., 2015; Garcia and Gregor, 2018; Garcia et al., 2013; Gregor, Garcia and Little, 2014). In addition, although it was not possible within the time limits of my PhD, in the future I plan to insert the same MS2 constructs containing enhancers from *D. melanogaster, D. yakuba* and *D. virilis* into a non-endogenous attP integration site in the *D. melanogaster* genome. This is an important and interesting experiment as it would allow quantification of the expression of the P2 proximal enhancer without any predicted buffering effects from the distal enhancer. Furthermore my analysis using the MS2 could be extended to other species with interesting P2 variants and even to properly test synthetic P2s designed using quantitative measurements from analysis natural P2 variants (Park et al., 2019).

Chapter 5

General Discussion

The aim of my PhD work was to investigate three important aspects of enhancers to further our understanding of these important cis-regulatory elements: enhancer identification, enhancer functionality and enhancer evolution. I have discussed specific aspects of my results at the end of the three results chapters. Here I expand more broadly on my findings to present more general perspectives on each of these three themes.

5.1 Chapter two

5.1.1 Approaches to identify post-embryonic enhancers

Chapter two of this thesis aimed to identify new enhancers through integrating both bioinformatics and functional testing. The combination of ATAC-seq and reporter assays proved a successful method for identifying a new enhancer of *Ubx* in the post-embryonic T2 that had been previously predicted from evo devo data (Davis et al., 2007). This combination of methods has also proven successful for other studies e.g. (Davie et al., 2015; Quillien et al., 2017). Therefore, the application of ATAC-seq is an excellent approach to guide the discovery of tissue and stage specific enhancers. Furthermore, despite a number of shortcomings, reporter constructs remain an essential step to efficiently verify predicted enhancers (Kvon, 2015). The expression driven by the reporter constructs, especially VT42733, in this investigation appeared to be quite promiscuous – with strong expression being observed in several tissues in

the *Drosophila* pupae (Fig. 6). At first glance, the Ubx protein trap expression suggests that the reporter gene expression in many of the tissues is ectopic. However, this protein trap is not strong enough to provide an accurate assessment (Fig. 18). While this could be followed up using an antibody staining against GFP on specific tissues, some of the tissues in the *Drosophila* pupae, for example the legs, are difficult to dissect and perform antibody stains on. Therefore, it remains a challenging task to assess the actual expression pattern of Ubx in the pupae.

One successful approach I took to help deal with the complex expression patterns driven by reporter constructs, was to carry out a functional assay in parallel to test enhancers. In this case, despite the complex expression pattern of VT42733, the effect of this region on the trichome pattern in the T2 leg was clear, and it allowed me to identify this fragment as a *Ubx* NV enhancer. It must be said, however, that this approach relies on having a reliable phenotypic assay, which may not be available for all genes in all post embryonic tissues of interest.

5.1.2 Deciphering post-embryonic GRNs

Identification of Hox gene post-embryonic enhancers, like the recently discovered *Dll* responsive enhancer of *Scr* (Eksi et al., 2018) and the *Ubx* enhancer I have identified, will allow us to start to uncover how Hox gene expression is regulated throughout development and better understand their post-embryonic roles.

The next step is to decipher how these post-embryonic Hox enhancers fit into the GRNs that determine fine scale morphology and further uncover the TFs that bind to them. In this investigation I have discovered some good candidate TFs for postembryonic regulation of *Ubx*, including Dll and Pho (Fig. 19), which can now be

functionally tested. In case there are still large numbers of candidate TFs that bind validated enhancers of interest, they could be further filtered through the use of the ATAC-seq data directly because there have been recent developments in both the preparation and analysis of ATAC-seq data which now allow the discovery of TFBS within the ATAC-seq profile (Karabacak Calviello et al., 2019; Li et al., 2019). One approach to then test TF candidates for *Ubx* regulation that can be applied to the enhancers of other genes, might be to make another reporter construct containing the NV enhancer (VT42733) directly driving GFP (instead of GAL4). Potential TFs could then be knocked down using RNAi (via the GAL4 system) and the effect on the GFP pattern can be assayed.

To further integrate *Ubx* and the NV enhancer into the leg trichome GRN the downstream factors that might be directly regulated by Ubx in T2 leg development need to be identified. This is a challenging task though because the eight Hox genes in *Drosophila* have very similar DNA binding preferences when assayed in vitro (reviewed in (Mann, Lelli and Joshi, 2009). Even with the knowledge that Hox proteins generally bind together with a co-factor in vivo, e.g. Exd or Hth in *Drosophila*, it is not sufficient to look for the longer Hox-cofactor TFBS because in some tissues, such as the developing halteres, Hox proteins can function in absence of their cofactors (Galant, Walsh and Carroll, 2002).

Recently it has been shown that there is a strong association between specific Hox protein binding and chromatin accessibility, and that Ubx binds almost exclusively to open chromatin (Porcelli et al., 2019). This could help narrow the search for the presence of Ubx binding sites. Indeed, such approaches could be used in combination with those shown in an elegant study by Pavlopoulos and Akam

(2011) that made the identification of high probability Ubx target genes in halters possible. However, this study showed that Hox genes may potentially regulate hundreds of genes in their post embryonic roles, and therefore functional validation to verify direct targets of Hox genes and thus to fully describe these networks may still not be straightforward.

5.2 Chapter three

5.2.1 Exploring a potential regulatory switch for trichome development

In chapter three, I sought to better understand enhancer functionality by using a specific set of genes to determine if there were rules to their cis-regulatory logic. I identified a set of genes that are predicted to be post-transcriptionally repressed by miR-92a and also directly dependent on Svb for activation (Fig. 21) and thus may combine to provide a regulatory switch between development of trichomes or naked cuticle. However, there are very likely to be other genes that help form this switch because the Svb targets were identified in embryos, and although checked against RNA-seq data for the pupal legs (Kittelmann et al., 2018), there are possibly genes that are also predicted targets of the miRNA and Svb-dependent but that are only expressed in the legs. These could be identified by generating new RNA-seq data for pupal legs from *svb* mutants and comparing the expression of genes to wild type legs at the same stage. These genes could then be tested individually and in combination using RNAi in the same approach taken in this thesis.

To further explore the evolution of potential leg trichome switch it would be interesting to also study the expression, regulation and function of Svb, miR-92a, and

their targets in the T2 pupal legs of *D. virilis* because this species does not have a naked valley (Stern, 1998). Therefore, there are presumably changes in leg trichome GRN of this species - perhaps even as simply differences in the regulation of miR-92a so that it is not expressed in the legs.

5.2.2 Investigating cis-regulatory logic

I then used the 'switch' genes as a system to explore transcriptional cis-regulatory logic. I hypothesised that these genes would have Svb-responsive enhancers. Therefore, I used the open chromatin for all 12 genes to perform a motif enrichment search in comparison to both 'closed' chromatin regions and epidermal genes that are independent of Svb (Menoret et al., 2013). This revealed the open chromatin of the 12 genes is enriched in motifs for both Svb and Grh. By then looking more closely at each individual gene, I was able to pinpoint potential enhancers of some of these genes predicted to be regulated by these two TFs.

Further analysis of the actual Svb/Ovo motifs present (as several different motifs were identified as enriched by i-cisTarget) could also allow an even more precise way of predicting enhancers. Indeed, the research by (Menoret et al., 2013) showed that Svb-enhancers in the embryo were generally composed of particular Svb motifs, and looking for this composition allowed them to more accurately predict enhancers (Menoret et al., 2013). For example, they determined that a combination of Svb-F7 motifs and either yellow or blue Svb motifs was a more accurate predictor of enhancers than enrichment of Ovo-Q6 sites (Menoret et al., 2013).

It now remains to test the predictive framework I applied by verifying if these sequences are actually enhancers and if they drive in different tissues based on the TF

binding motifs they contain. Again, this can now be carried out using reporter constructs with and without engineered changes in the Svb and Grh binding sites they contain. I predict that open chromatin regions enriched for both Svb and Grh will express in the developing pupal leg whereas the regions that are just enriched for Svb will not be pupal epidermal enhancers due to the lack of Grh. I also predict that regions with enrichment for Grh but not Svb, will be pupal epidermal enhancers but not-dependent on Svb and therefore may express in different epidermal tissues.

If the enhancers that I have predicted are functional then my study has also demonstrated that by analysing a set of functionally related genes, there can be a common pattern of motifs that allow you to identify enhancers. Although this approach has been successfully carried out in embryos (Menoret et al., 2013), importantly my study has examined it in a post-embryonic context, where we lack knowledge of post-embryonic enhancers, and therefore being able to use motifs as a predictive framework is a useful step.

5.3 Chapter four

Chapter four aimed to develop tools to better study and understand enhancer functionality and evolution using the *hunchback* P2 enhancer.

5.3.1 Turnover in enhancer sequences

In this investigation I was able to build on our current understanding of turnover in the *hb* P2 enhancer by using up-to-date genome releases, which allowed me to assess more species and get a better idea of the extent of turnover in the P2 enhancer. I

observed that the Bcd binding sites within the enhancer have accumulated mutations throughout evolution, and in some species some of the binding sites are not present (Fig. 44). The analysis of several species also highlighted the extent of turnover in the inter-binding site nucleotides, and changes in the spacing between, which several other studies have highlighted, can significantly alter the output of enhancers (Farley et al., 2015; Li and Eisen, 2018). Furthermore, from the alignment, I was able to identify blocks of conserved DNA that have not yet been classified as TFBS, but are likely important for enhancer function and can now be tested.

5.3.2 Combining natural variation and live imaging to understand enhancer function

By using the P2 enhancers from several species and a state-of-the-art live-imaging approach, I was able to create a set of flies that will now be invaluable in future studies of enhancer functionality. Specifically, I generated flies containing different *hb* P2 enhancers from *D. melanogaster*, *D. yakuba* and *D. virilis*, and placed the P2s into the endogenous locus in *D. melanogaster* upstream of MS2 stem loops to allow quantitative measurements of transcription.

My preliminary analysis of these flies suggests that while they give similar expression there are differences in the levels and spacing of expression. Further detailed comparison of these variants perhaps in combination with other species now offers an excellent opportunity to identify and test the sequences that are necessary and sufficient for *hb* expression as well as which sequence differences cause differences in expression.

Therefore, the future directions of this project are to image and quantify all of the constructs made so far but also to extend the number of species used to better capture the *Drosophila* phylogeny. It will also be interesting to insert the same constructs into a non-endogenous locus in *D. melanogaster* where the P2 cannot communicate with the distal *hb* enhancer. I hypothesise that the distal enhancer is somehow aiding the buffering of transcription to deal with the mutations accumulated in the P2, therefore it is crucial to monitor what happens to transcription in the absence of the distal enhancer.

In future studies, the same approach could be carried out for the distal enhancer of *hb:* different combinations of the proximal and distal enhancers could be tested and the impact of buffering between the two enhancers evaluated. This would give novel insights into whether there is compensatory evolution between the proximal and distal enhancers to maintain tight transcriptional control. Furthermore, the ultimate goal of this investigation would be to take the information gained in the enhancer analysis, i.e which parts of the enhancer are flexible and which parts are constrained, to engineer a synthetic P2 enhancer element. This would allow precise testing of which sequences are sufficient and which are necessary (Park et al., 2019).

5.4 Conclusions

My investigation has broadly contributed to our knowledge of enhancer organisation and functionality, and in my opinion has highlighted some key ideas for future enhancer studies:

- It is more efficient to combine ATAC-seq (or other chromatin analysis) with reporter assays and functional studies to find enhancers. Furthermore,

adding motif analysis to the above methods can greatly support enhancer identification.

- The above can be further aided by comparing potential enhancers of functionally related genes to better reveal cis-regulatory logic.
- To functionally analyses individual enhancers an evolutionary perspective can be powerful. This could also greatly improve our knowledge of enhancer logic.
- While reporter assays provide important information, it is now desirable to manipulate the endogenous locus with CRISPR/Cas9
- The ultimate test of our knowledge of individual enhancers is to make synthetic constructs and test them *in vivo*.

References

Abrams, E. W. and Andrew, D. J. (2005) 'CrebA regulates secretory activity in the Drosophila salivary gland and epidermis', *Development,* 132(12), pp. 2743-58. doi: 10.1242/dev.01863.

Agarwal, V., Subtelny, A. O., Thiru, P., Ulitsky, I. and Bartel, D. P. (2018) 'Predicting microRNA targeting efficacy in Drosophila', *Genome Biol,* 19(1), pp. 152. doi: 10.1186/s13059-018-1504-3.

Agawa, Y., Sarhan, M., Kageyama, Y., Akagi, K., Takai, M., Hashiyama, K., Wada, T., Handa, H., Iwamatsu, A., Hirose, S. and Ueda, H. (2007) 'Drosophila Blimp-1 is a transient transcriptional repressor that controls timing of the ecdysone-induced developmental pathway', *Mol Cell Biol,* 27(24), pp. 8739-47. doi: 10.1128/mcb.01304-07.

Akagi, K. and Ueda, H. (2011) 'Regulatory mechanisms of ecdysone-inducible Blimp-1 encoding a transcriptional repressor that is important for the prepupal development in Drosophila', *Dev Growth Differ,* 53(5), pp. 697-703. doi: 10.1111/j.1440-169X.2011.01276.x.

Alonso, C. R. and Wilkins, A. S. (2005) 'The molecular elements that underlie developmental evolution', *Nat Rev Genet: Vol. 9*. England, pp. 709-15.

Alonso, M. E., Pernaute, B., Crespo, M., Gomez-Skarmeta, J. L. and Manzanares, M. (2009) 'Understanding the regulatory genome', *Int J Dev Biol,* 53(8-10), pp. 1367-78. doi: 10.1387/ijdb.072428ma.

Andrew, D. J., Baig, A., Bhanot, P., Smolik, S. M. and Henderson, K. D. (1997) 'The Drosophila dCREB-A gene is required for dorsal/ventral patterning of the larval cuticle', *Development,* 124(1), pp. 181-93.

Arif, S., Kittelmann, S. and McGregor, A. P. (2015) 'From shavenbaby to the naked valley: trichome formation as a model for evolutionary developmental biology', *Evol Dev,* 17(1), pp. 120-6. doi: 10.1111/ede.12113.

Arif, S., Murat, S., Almudi, I., Nunes, M. D., Bortolamiol-Becet, D., McGregor, N. S., Currie, J. M., Hughes, H., Ronshaugen, M., Sucena, E., Lai, E. C., Schlotterer, C. and McGregor, A. P. (2013) 'Evolution of mir-92a underlies natural morphological variation in Drosophila melanogaster', *Curr Biol,* 23(6), pp. 523-8. doi: 10.1016/j.cub.2013.02.018.

Arnold, C. D., Gerlach, D., Spies, D., Matts, J. A., Sytnikova, Y. A., Pagani, M., Lau, N. C. and Stark, A. (2014) 'Quantitative genome-wide enhancer activity maps for five Drosophila species show functional enhancer conservation and turnover during cis-regulatory evolution', *Nat Genet,* 46(7), pp. 685-92. doi: 10.1038/ng.3009.

Arnone, M. I. and Davidson, E. H. (1997) 'The hardwiring of development: organization and function of genomic regulatory systems', *Development,* 124(10), pp. 1851-64.

Arnoult, L., Su, K. F., Manoel, D., Minervino, C., Magrina, J., Gompel, N. and Prud'homme, B. (2013) 'Emergence and diversification of fly pigmentation through evolution of a

gene regulatory module', *Science,* 339(6126), pp. 1423-6. doi: 10.1126/science.1233749.

Badis, G., Berger, M. F., Philippakis, A. A., Talukder, S., Gehrke, A. R., Jaeger, S. A., Chan, E. T., Metzler, G., Vedenko, A., Chen, X., Kuznetsov, H., Wang, C. F., Coburn, D., Newburger, D. E., Morris, Q., Hughes, T. R. and Bulyk, M. L. (2009) 'Diversity and complexity in DNA recognition by transcription factors', *Science,* 324(5935), pp. 1720-3. doi: 10.1126/science.1162327.

Baek, S. and Sung, M. H. (2016) 'Genome-Scale Analysis of Cell-Specific Regulatory Codes Using Nuclear Enzymes', *Methods Mol Biol,* 1418, pp. 225-40. doi: 10.1007/978-1-4939-3578-9_12.

Baeza, M., Viala, S., Heim, M., Dard, A., Hudry, B., Duffraisse, M., Rogulja-Ortmann, A., Brun, C. and Merabet, S. (2015) 'Inhibitory activities of short linear motifs underlie Hox interactome specificity in vivo', *Elife,* 4. doi: 10.7554/eLife.06034.

Bailey, T. L., Boden, M., Buske, F. A., Frith, M., Grant, C. E., Clementi, L., Ren, J., Li, W. W. and Noble, W. S. (2009) 'MEME SUITE: tools for motif discovery and searching', *Nucleic Acids Res,* 37(Web Server issue), pp. W202-8. doi: 10.1093/nar/gkp335.

Balmert, A., Florian Bohn, H., Ditsche-Kuru, P. and Barthlott, W. (2011) 'Dry under water: comparative morphology and functional aspects of air-retaining insect surfaces', *J Morphol,* 272(4), pp. 442-51. doi: 10.1002/jmor.10921.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) 'Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences', *Cell,* 27(2 Pt 1), pp. 299-308. doi: 10.1016/0092-8674(81)90413-x.

Barkoulas, M., Vargas Velazquez, A. M., Peluffo, A. E. and Félix, M. A. (2016) 'Evolution of New cis-Regulatory Motifs Required for Cell-Specific Gene Expression in Caenorhabditis', *PLoS Genet: Vol. 9*.

Barmina, O. and Kopp, A. (2007) 'Sex-specific expression of a HOX gene associated with rapid morphological evolution', *Dev Biol,* 311(2), pp. 277-86. doi: 10.1016/j.ydbio.2007.07.030.

Barolo, S. (2012) 'Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy', *Bioessays,* 34(2), pp. 135-41. doi: 10.1002/bies.201100121.

Barr, K. A., Martinez, C., Moran, J. R., Kim, A. R., Ramos, A. F. and Reinitz, J. (2017) 'Synthetic enhancer design by in silico compensatory evolution reveals flexibility and constraint in cis-regulation', *BMC Syst Biol,* 11(1), pp. 116. doi: 10.1186/s12918-017-0485-2.

Barriere, A. and Ruvinsky, I. (2014) 'Pervasive divergence of transcriptional gene regulation in Caenorhabditis nematodes', *PLoS Genet,* 10(6), pp. e1004435. doi: 10.1371/journal.pgen.1004435.

Belting, H. G., Shashikant, C. S. and Ruddle, F. H. (1998) 'Modification of expression and cis-regulation of Hoxc8 in the evolution of diverged axial morphology', *Proc Natl Acad Sci U S A,* 95(5), pp. 2355-60. doi: 10.1073/pnas.95.5.2355.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H. and Long, R. M. (1998) 'Localization of ASH1 mRNA particles in living yeast', *Mol Cell,* 2(4), pp. 437-45. doi: 10.1016/s1097-2765(00)80143-4.

Bischof, J., Duffraisse, M., Furger, E., Ajuria, L., Giraud, G., Vanderperre, S., Paul, R., Björklund, M., Ahr, D., Ahmed, A. W., Spinelli, L., Brun, C., Basler, K. and Merabet, S. (2018) 'Generation of a versatile BiFC ORFeome library for analyzing protein–protein interactions in live Drosophila', *eLife*.

Blythe, A., S. and Wieschaus, E. F. (2016) 'Establishment and maintenance of heritable chromatin structure during early Drosophila embryogenesis'. doi: doi:10.7554/eLife.20148.

Bokor, P. and DiNardo, S. (1996) 'The roles of hedgehog, wingless and lines in patterning the dorsal epidermis in Drosophila', *Development,* 122(4), pp. 1083-92.

Bonneton, F., Shaw, P. J., Fazakerley, C., Shi, M. and Dover, G. A. (1997) 'Comparison of bicoid-dependent regulation of hunchback between Musca domestica and Drosophila melanogaster', *Mech Dev,* 66(1-2), pp. 143-56.

Borok, M. J., Tran, D. A., Ho, M. C. W. and Drewell, R. A. (2010) 'Dissecting the regulatory switches of development: lessons from enhancer evolution in Drosophila', *Development,* 137(1), pp. 5-13. doi: 10.1242/dev.036160.

Bothma, J. P., Garcia, H. G., Esposito, E., Schlissel, G., Gregor, T. and Levine, M. (2014) 'Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos', *Proc Natl Acad Sci U S A,* 111(29), pp. 10598-603. doi: 10.1073/pnas.1410022111.

Bothma, J. P., Garcia, H. G., Ng, S., Perry, M. W., Gregor, T. and Levine, M. (2015) 'Enhancer additivity and non-additivity are determined by enhancer strength in the Drosophila embryo', *eLife*.

Brodu, V., Elstob, P. R. and Gould, A. P. (2002) 'abdominal A specifies one cell type in Drosophila by regulating one principal target gene', *Development,* 129(12), pp. 2957-63.

Bryantsev, A. L., Duong, S., Brunetti, T. M., Chechenova, M. B., Lovato, T. L., Nelson, C., Shaw, E., Uhl, J. D., Gebelein, B. and Cripps, R. M. (2012) 'Extradenticle and homothorax control adult muscle fiber identity in Drosophila', *Dev Cell,* 23(3), pp. 664-73. doi: 10.1016/j.devcel.2012.08.004.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013) 'Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position', *Nat Methods,* 10(12), pp. 1213-8. doi: 10.1038/nmeth.2688.

Burgess, D. and Freeling, M. (2014) 'The most deeply conserved noncoding sequences in plants serve similar functions to those in vertebrates despite large differences in evolutionary rates', *Plant Cell,* 26(3), pp. 946-61. doi: 10.1105/tpc.113.121905.

Burz, D. S., Rivera-Pomar, R., Jäckle, H. and Hanes, S. D. (1998) 'Cooperative DNA-binding by Bicoid provides a mechanism for threshold-dependent gene activation in the

Drosophila embryo', *EMBO J,* 17(20), pp. 5998-6009. doi: 10.1093/emboj/17.20.5998.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. and Madden, T. L. (2009) 'BLAST+: architecture and applications', *BMC Bioinformatics,* 10, pp. 421. doi: 10.1186/1471-2105-10-421.

Cannavo, E., Khoueiry, P., Garfield, D. A., Geeleher, P., Zichner, T., Gustafson, E. H., Ciglar, L., Korbel, J. O. and Furlong, E. E. (2016) 'Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks', *Curr Biol,* 26(1), pp. 38-51. doi: 10.1016/j.cub.2015.11.034.

Carl, S. H. and Russell, S. (2015) 'Common binding by redundant group B Sox proteins is evolutionarily conserved in Drosophila', *BMC Genomics,* 16, pp. 292. doi: 10.1186/s12864-015-1495-3.

Carroll, S. B. (2008) 'Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution', *Cell,* 134(1), pp. 25-36. doi: 10.1016/j.cell.2008.06.030.

Casas-Tintó, S., Arnés, M. and Ferrús, A. (2017) 'Drosophila enhancer-Gal4 lines show ectopic expression during development', *R Soc Open Sci: Vol. 3*.

Chan, Y. F., Marks, M. E., Jones, F. C., Villarreal, G., Jr., Shapiro, M. D., Brady, S. D., Southwick, A. M., Absher, D. M., Grimwood, J., Schmutz, J., Myers, R. M., Petrov, D., Jonsson, B., Schluter, D., Bell, M. A. and Kingsley, D. M. (2010) 'Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer', *Science,* 327(5963), pp. 302-5. doi: 10.1126/science.1182213.

Chanut-Delalande, H., Fernandes, I., Roch, F., Payre, F. and Plaza, S. (2006) 'Shavenbaby couples patterning to epidermal cell shape control', *PLoS Biol,* 4(9), pp. e290. doi: 10.1371/journal.pbio.0040290.

Clarke, S. L., VanderMeer, J. E., Wenger, A. M., Schaar, B. T., Ahituv, N. and Bejerano, G. (2012) 'Human developmental enhancers conserved between deuterostomes and protostomes', *PLoS Genet,* 8(8), pp. e1002852. doi: 10.1371/journal.pgen.1002852.

Costas, J., Vieira, C. P., Casares, F. and Vieira, J. (2003) 'Genomic characterization of a repetitive motif strongly associated with developmental genes in Drosophila', *BMC Genomics,* 4(1), pp. 52. doi: 10.1186/1471-2164-4-52.

Creyghton, M. P., Cheng, A. W., Welstead, G. G., Kooistra, T., Carey, B. W., Steine, E. J., Hanna, J., Lodato, M. A., Frampton, G. M., Sharp, P. A., Boyer, L. A., Young, R. A. and Jaenisch, R. (2010) 'Histone H3K27ac separates active from poised enhancers and predicts developmental state', *Proc Natl Acad Sci U S A,* 107(50), pp. 21931-6. doi: 10.1073/pnas.1016071107.

Crocker, J., Abe, N., Rinaldi, L., McGregor, A. P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F., Mann, R. S. and Stern, D. L. (2015) 'Low Affinity Binding Site Clusters Confer Hox Specificity and Regulatory Robustness', *Cell,* 160(0), pp. 191-203. doi: 10.1016/j.cell.2014.11.041.

Crocker, J. and Ilsley, G. R. (2017) 'Using synthetic biology to study gene regulatory evolution', *Curr Opin Genet Dev,* 47, pp. 91-101. doi: 10.1016/j.gde.2017.09.001.

Crocker, J., Noon, E. P. and Stern, D. L. (2016) 'The Soft Touch: Low-Affinity Transcription Factor Binding Sites in Development and Evolution', *Curr Top Dev Biol,* 117, pp. 455-69. doi: 10.1016/bs.ctdb.2015.11.018.

Cusanovich, D. A., Reddington, J. P., Garfield, D. A., Daza, R. M., Aghamirzaie, D., Marco-Ferreres, R., Pliner, H. A., Christiansen, L., Qiu, X., Steemers, F. J., Trapnell, C., Shendure, J. and Furlong, E. E. M. (2018) 'The cis-regulatory dynamics of embryonic development at single-cell resolution', *Nature,* 555(7697), pp. 538-542. doi: 10.1038/nature25981.

Dave, V., Zhao, C., Yang, F., Tung, C. S. and Ma, J. (2000) 'Reprogrammable recognition codes in bicoid homeodomain-DNA interaction', *Mol Cell Biol,* 20(20), pp. 7673-84. doi: 10.1128/mcb.20.20.7673-7684.2000.

Davie, K., Jacobs, J., Atkins, M., Potier, D., Christiaens, V., Halder, G. and Aerts, S. (2015) 'Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling', *PLoS Genet,* 11(2), pp. e1004994. doi: 10.1371/journal.pgen.1004994.

Davis, G. K., Srinivasan, D. G., Wittkopp, P. J. and Stern, D. L. (2007) 'The function and regulation of Ultrabithorax in the legs of Drosophila melanogaster', *Dev Biol,* 308(2), pp. 621-31. doi: 10.1016/j.ydbio.2007.06.002.

Delon, I., Chanut-Delalande, H. and Payre, F. (2003) 'The Ovo/Shavenbaby transcription factor specifies actin remodelling during epidermal differentiation in Drosophila', *Mech Dev,* 120(7), pp. 747-58.

Dickinson, W. J. and Thatcher, J. W. (1997) 'Morphogenesis of denticles and hairs in Drosophila embryos: involvement of actin-associated proteins that also affect adult structures', *Cell Motil Cytoskeleton,* 38(1), pp. 9-21. doi: 10.1002/(sici)1097-0169(1997)38:1<9::aid-cm2>3.0.co;2-4.

Ditsche-Kuru, P., Schneider, E. S., Melskotte, J. E., Brede, M., Leder, A. and Barthlott, W. (2011) 'Superhydrophobic surfaces of the water bug Notonecta glauca: a model for friction reduction and air retention', *Beilstein J Nanotechnol,* 2, pp. 137-44. doi: 10.3762/bjnano.2.17.

Domsch, K., Carnesecchi, J., Disela, V., Friedrich, J., Trost, N., Ermakova, O., Polychronidou, M. and Lohmann, I. (2019) 'The Hox transcription factor Ubx stabilizes lineage commitment by suppressing cellular plasticity in Drosophila', *Elife,* 8. doi: 10.7554/eLife.42675.

Dover, G. (2000) 'How genomic and developmental dynamics affect evolutionary processes', *Bioessays,* 22(12), pp. 1153-9. doi: 10.1002/1521-1878(200012)22:12<1153::aid-bies13>3.0.co;2-0.

Driever, W. and Nusslein-Volhard, C. (1989) 'The bicoid protein is a positive regulator of hunchback transcription in the early Drosophila embryo', *Nature,* 337(6203), pp. 138-43. doi: 10.1038/337138a0.

Driever, W., Thoma, G. and Nusslein-Volhard, C. (1989) 'Determination of spatial domains of zygotic gene expression in the Drosophila embryo by the affinity of binding sites for the bicoid morphogen', *Nature,* 340(6232), pp. 363-7. doi: 10.1038/340363a0.

Dunipace, L., Ozdemir, A. and Stathopoulos, A. (2011) 'Complex interactions between cis-regulatory modules in native conformation are critical for Drosophila snail expression', *Development,* 138(18), pp. 4075-84. doi: 10.1242/dev.069146.

Duque, T. and Sinha, S. (2015) 'What Does It Take to Evolve an Enhancer? A Simulation-Based Study of Factors Influencing the Emergence of Combinatorial Regulation', *Genome Biol Evol: Vol. 6*, pp. 1415-31.

Durrett, R. and Schmidt, D. (2008) 'Waiting for Two Mutations: With Applications to Regulatory Sequence Evolution and the Limits of Darwinian Evolution', *Genetics: Vol. 3*, pp. 1501-9.

Dynan, W. S. (1989) 'Modularity in promoters and enhancers', *Cell,* 58(1), pp. 1-4. doi: 10.1016/0092-8674(89)90393-0.

Eksi, S. E., Barmina, O., McCallough, C. L., Kopp, A. and Orenic, T. V. (2018) 'A Distalless-responsive enhancer of the Hox gene Sex combs reduced is required for segment- and sex-specific sensory organ development in Drosophila', *PLoS Genet,* 14(4), pp. e1007320. doi: 10.1371/journal.pgen.1007320.

El-Sherif, E. and Levine, M. (2016) 'Shadow Enhancers Mediate Dynamic Shifts of Gap Gene Expression in the Drosophila Embryo', *Curr Biol,* 26(9), pp. 1164-9. doi: 10.1016/j.cub.2016.02.054.

Erives, A. and Levine, M. (2004) 'Coordinate enhancers share common organizational features in the Drosophila genome', *Proc Natl Acad Sci U S A,* 101(11), pp. 3851-6. doi: 10.1073/pnas.0400611101.

Ernst, J., Kheradpour, P., Mikkelsen, T. S., Shoresh, N., Ward, L. D., Epstein, C. B., Zhang, X., Wang, L., Issner, R., Coyne, M., Ku, M., Durham, T., Kellis, M. and Bernstein, B. E. (2011) 'Systematic analysis of chromatin state dynamics in nine human cell types', *Nature,* 473(7345), pp. 43-9. doi: 10.1038/nature09906.

Estacio-Gomez, A., Moris-Sanz, M., Schafer, A. K., Perea, D., Herrero, P. and Diaz-Benjumea, F. J. (2013) 'Bithorax-complex genes sculpt the pattern of leucokinergic neurons in the Drosophila central nervous system', *Development,* 140(10), pp. 2139-48. doi: 10.1242/dev.090423.

Farley, E. K., Olson, K. M., Zhang, W., Brandt, A. J., Rokhsar, D. S. and Levine, M. S. (2015) 'Suboptimization of developmental enhancers', *Science,* 350(6258), pp. 325-8. doi: 10.1126/science.aac6948.

Ferraro, T., Esposito, E., Mancini, L., Ng, S., Lucas, T., Coppey, M., Dostatni, N., Walczak, A. M., Levine, M. and Lagha, M. (2016) 'Transcriptional Memory in the Drosophila Embryo', *Curr Biol,* 26(2), pp. 212-218. doi: 10.1016/j.cub.2015.11.058.

Forrest, K. M. and Gavis, E. R. (2003) 'Live imaging of endogenous RNA reveals a diffusion and entrapment mechanism for nanos mRNA localization in Drosophila', *Curr Biol,* 13(14), pp. 1159-68. doi: 10.1016/s0960-9822(03)00451-2.

Frankel, N., Davis, G. K., Vargas, D., Wang, S., Payre, F. and Stern, D. L. (2010) 'Phenotypic robustness conferred by apparently redundant transcriptional enhancers', *Nature,* 466(7305), pp. 490-3. doi: 10.1038/nature09158.

Frankel, N., Erezyilmaz, D. F., McGregor, A. P., Wang, S., Payre, F. and Stern, D. L. (2011) 'Morphological evolution caused by many subtle-effect substitutions in regulatory DNA', *Nature,* 474(7353), pp. 598-603. doi: 10.1038/nature10200.

Frankel, N., Wang, S. and Stern, D. L. (2012) 'Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution', *Proc Natl Acad Sci U S A,* 109(51), pp. 20975-9. doi: 10.1073/pnas.1207715109.

Fukaya, T., Lim, B. and Levine, M. (2016) 'Enhancer Control of Transcriptional Bursting', *Cell,* 166(2), pp. 358-368. doi: 10.1016/j.cell.2016.05.025.

Galant, R., Walsh, C. M. and Carroll, S. B. (2002) 'Hox repression of a target gene: extradenticle-independent, additive action through multiple monomer binding sites', *Development,* 129(13), pp. 3115-26.

Ganley, A. R. and Kobayashi, T. (2007) 'Phylogenetic footprinting to find functional DNA elements', *Methods Mol Biol,* 395, pp. 367-80. doi: 10.1007/978-1-59745-514-5_23.

Garcia, H. and Gregor, T. (2018) 'Live imaging of mRNA synthesis in Drosophila', *Methods Mol Biol,* 1649, pp. 349-57. doi: 10.1007/978-1-4939-7213-5_23.

Garcia, H. G., Tikhonov, M., Lin, A. and Gregor, T. (2013) 'Quantitative imaging of transcription in living Drosophila embryos links polymerase activity to patterning', *Curr Biol,* 23(21), pp. 2140-5. doi: 10.1016/j.cub.2013.08.054.

Garg, A., Srivastava, A., Davis, M. M., O'Keefe, S. L., Chow, L. and Bell, J. B. (2007) 'Antagonizing scalloped with a novel vestigial construct reveals an important role for scalloped in Drosophila melanogaster leg, eye and optic lobe development', *Genetics,* 175(2), pp. 659-69. doi: 10.1534/genetics.106.063966.

Giniger, E. and Ptashne, M. (1988) 'Cooperative DNA binding of the yeast transcriptional activator GAL4', *Proc Natl Acad Sci U S A,* 85(2), pp. 382-6. doi: 10.1073/pnas.85.2.382.

Gokhale, R. and Pfleger, C. M. (2019) 'The Power of Drosophila Genetics: The Discovery of the Hippo Pathway', *Methods Mol Biol,* 1893, pp. 3-26. doi: 10.1007/978-1-4939-8910-2_1.

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. and Carroll, S. B. (2005) 'Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila', *Nature,* 433(7025), pp. 481-7. doi: 10.1038/nature03235.

Goodwyn, P. J., Voigt, D. and Fujisaki, K. (2008) 'Skating and diving: Changes in functional morphology of the setal and microtrichial cover during ontogenesis in Aquarius paludum fabricius (Heteroptera, Gerridae)', *J Morphol,* 269(6), pp. 734-44. doi: 10.1002/jmor.10619.

Goodwyn, P. P., De Souza, E., Fujisaki, K. and Gorb, S. (2008) 'Moulding technique demonstrates the contribution of surface geometry to the super-hydrophobic properties of the surface of a water strider', *Acta Biomater,* 4(3), pp. 766-70. doi: 10.1016/j.actbio.2008.01.002.

Gordon, K. L., Arthur, R. K. and Ruvinsky, I. (2015) 'Phylum-Level Conservation of Regulatory Information in Nematodes despite Extensive Non-coding Sequence Divergence', *PLoS Genet,* 11(5), pp. e1005268. doi: 10.1371/journal.pgen.1005268.

Gregor, T., Garcia, H. G. and Little, S. C. (2014) 'The embryo as a laboratory: quantifying transcription in Drosophila', *Trends Genet,* 30(8), pp. 364-75. doi: 10.1016/j.tig.2014.06.002.

Gruzdeva, N., Kyrchanova, O., Parshikov, A., Kullyev, A. and Georgiev, P. (2005) 'The Mcp Element from the bithorax Complex Contains an Insulator That Is Capable of Pairwise Interactions and Can Facilitate Enhancer-Promoter Communication', *Mol Cell Biol: Vol. 9*, pp. 3682-9.

Gupta, S., Stamatoyannopoulos, J. A., Bailey, T. L. and Noble, W. S. (2007) 'Quantifying similarity between motifs', *Genome Biol,* 8(2), pp. R24. doi: 10.1186/gb-2007-8-2-r24.

Halder, G., Polaczyk, P., Kraus, M. E., Hudson, A., Kim, J., Laughon, A. and Carroll, S. (1998) 'The Vestigial and Scalloped proteins act together to directly regulate wing-specific gene expression in Drosophila', *Genes Dev,* 12(24), pp. 3900-9. doi: 10.1101/gad.12.24.3900.

Halfon, M. S. (2019) 'Studying Transcriptional Enhancers: The Founder Fallacy, Validation Creep, and Other Biases', *Trends Genet,* 35(2), pp. 93-103. doi: 10.1016/j.tig.2018.11.004.

Hancock, J. M., Shaw, P. J., Bonneton, F. and Dover, G. A. (1999) 'High sequence turnover in the regulatory regions of the developmental gene hunchback in insects', *Mol Biol Evol,* 16(2), pp. 253-65. doi: 10.1093/oxfordjournals.molbev.a026107.

Hanes, S. D., Riddihough, G., Ish-Horowicz, D. and Brent, R. (1994) 'Specific DNA recognition and intersite spacing are critical for action of the bicoid morphogen', *Mol Cell Biol,* 14(5), pp. 3364-75. doi: 10.1128/mcb.14.5.3364.

Hardison, R. C. and Taylor, J. (2012) 'Genomic approaches towards finding cis-regulatory modules in animals', *Nat Rev Genet,* 13(7), pp. 469-83. doi: 10.1038/nrg3242.

Hare, E. E., Peterson, B. K., Iyer, V. N., Meier, R. and Eisen, M. B. (2008) 'Sepsid even-skipped Enhancers Are Functionally Conserved in Drosophila Despite Lack of Sequence Conservation', *PLoS Genet: Vol. 6*.

Hasma, H. and Halfon, M. S. (2019) 'Computational enhancer prediction: evaluation and improvements', *BMC Bioinformatics,* 20(1), pp. 1-15. doi: doi:10.1186/s12859-019-2781-x.

He, X., Duque, T. S. and Sinha, S. (2012) 'Evolutionary origins of transcription factor binding site clusters', *Mol Biol Evol,* 29(3), pp. 1059-70. doi: 10.1093/molbev/msr277.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E. and Ren, B. (2007) 'Distinct and predictive chromatin signatures of transcriptional

promoters and enhancers in the human genome', *Nat Genet,* 39(3), pp. 311-8. doi: 10.1038/ng1966.

Hens, K., Feuz, J. D. and Deplancke, B. (2012) 'A high-throughput gateway-compatible yeast one-hybrid screen to detect protein-DNA interactions', *Methods Mol Biol,* 786, pp. 335-55. doi: 10.1007/978-1-61779-292-2_20.

Herrmann, C., Van de Sande, B., Potier, D. and Aerts, S. (2012) 'i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules', *Nucleic Acids Res: Vol. 15*, pp. e114.

Hochschild, A., Douhan, J., 3rd and Ptashne, M. (1986) 'How lambda repressor and lambda Cro distinguish between OR1 and OR3', *Cell,* 47(5), pp. 807-16. doi: 10.1016/0092-8674(86)90523-4.

Hong, J. W., Hendrix, D. A. and Levine, M. S. (2008) 'Shadow enhancers as a source of evolutionary novelty', *Science,* 321(5894), pp. 1314. doi: 10.1126/science.1160631.

Iampietro, C., Cléard, F., Gyurkovics, H., Maeda, R. K. and Karch, F. (2008) 'Boundary swapping in the Drosophila Bithorax complex'. doi: 10.1242/dev.025700.

Imrichová, H., Hulselmans, G., Kalender Atak, Z., Potier, D. and Aerts, S. (2015) 'i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly', *Nucleic Acids Res,* 43(Web Server issue), pp. W57-64. doi: 10.1093/nar/gkv395.

Indjeian, V. B., Kingman, G. A., Jones, F. C., Guenther, C. A., Grimwood, J., Schmutz, J., Myers, R. M. and Kingsley, D. M. (2016) 'Evolving New Skeletal Traits by cis-Regulatory Changes in Bone Morphogenetic Proteins', *Cell,* 164(1-2), pp. 45-56. doi: 10.1016/j.cell.2015.12.007.

Inestrosa, N. C., Sunkel, C. E., Arriagada, J., Garrido, J. and Godoy-Herrera, R. (1996) 'Abnormal development of the locomotor activity in yellow larvae of Drosophila: a cuticular defect?', *Genetica,* 97(2), pp. 205-10.

Irish, V., Lehmann, R. and Akam, M. (1989) 'The Drosophila posterior-group gene nanos functions by repressing hunchback activity', *Nature,* 338(6217), pp. 646-8. doi: 10.1038/338646a0.

Jacobs, J., Atkins, M., Davie, K., Imrichova, H., Romanelli, L., Christiaens, V., Hulselmans, G., Potier, D., Wouters, J., Taskiran, II, Paciello, G., González-Blas, C. B., Koldere, D., Aibar, S., Halder, G. and Aerts, S. (2018) 'The transcription factor Grainyhead primes epithelial enhancers for spatiotemporal activation by displacing nucleosomes', *Nat Genet,* 50(7), pp. 1011-20. doi: 10.1038/s41588-018-0140-x.

Jenett, A., Rubin, G. M., Ngo, T. T. B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B. D., Cavallaro, A., Hall, D., Jeter, J., Iyer, N., Fetter, D., Hausenfluck, J. H., Peng, H., Trautman, E. T., Svirskas, R., Myers, E. W., Iwinski, Z. R., Aso, Y., DePasquale, G. M., Enos, A., Hulamm, P., Lam, S. C. B., Li, H. H., Laverty, T. R., Long, F., Qu, L., Murphy, S. D., Rokicki, K., Safford, T., Shaw, K., Simpson, J. H., Sowell, A., Tae, S., Yu, Y. and Zugates, C. T. (2012) 'A GAL4-Driver Line Resource for Drosophila Neurobiology', *Cell Rep,* 2(4), pp. 991-1001. doi: 10.1016/j.celrep.2012.09.011.

Jeong, S., Rokas, A. and Carroll, S. B. (2006) 'Regulation of body pigmentation by the Abdominal-B Hox protein and its gain and loss in Drosophila evolution', *Cell,* 125(7), pp. 1387-99. doi: 10.1016/j.cell.2006.04.043.

Johnson, L. A., Zhao, Y., Golden, K. and Barolo, S. (2008) 'Reverse-Engineering a Transcriptional Enhancer: A Case Study in Drosophila', *Tissue Eng Part A: Vol. 9*, pp. 1549-59.

Kalay, G., Lachowiec, J., Rosas, U., Dome, M. R. and Wittkopp, P. (2019) 'Redundant and Cryptic Enhancer Activities of the Drosophila yellow Gene', *Genetics,* 212(1), pp. 343-360. doi: 10.1534/genetics.119.301985.

Kannan, R., Berger, C., Myneni, S., Technau, G. M. and Shashidhara, L. S. (2010) 'Abdominal-A mediated repression of Cyclin E expression during cell-fate specification in the Drosophila central nervous system', *Mech Dev,* 127(1-2), pp. 137-45. doi: 10.1016/j.mod.2009.09.008.

Kaplan, N. A., Colosimo, P. F., Liu, X. and Tolwinski, N. S. (2011) 'Complex Interactions between GSK3 and aPKC in Drosophila Embryonic Epithelial Morphogenesis', *PLoS One: Vol. 4*.

Karabacak Calviello, A., Hirsekorn, A., Wurmus, R., Yusuf, D. and Ohler, U. (2019) 'Reproducible inference of transcription factor footprints in ATAC-seq and DNase-seq datasets using protocol-specific bias modeling', *Genome Biol,* 20(1), pp. 42. doi: 10.1186/s13059-019-1654-y.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002) 'MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform', *Nucleic Acids Res,* 30(14), pp. 3059-66. doi: 10.1093/nar/gkf436.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J. A., van der Lee, R., Bessy, A., Cheneby, J., Kulkarni, S. R., Tan, G., Baranasic, D., Arenillas, D. J., Sandelin, A., Vandepoele, K., Lenhard, B., Ballester, B., Wasserman, W. W., Parcy, F. and Mathelier, A. (2018) 'JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework', *Nucleic Acids Res,* 46(D1), pp. D260-d266. doi: 10.1093/nar/gkx1126.

Kittelmann, S. and McGregor, A. P. (2019) 'Modulation and Evolution of Animal Development through microRNA Regulation of Gene Expression', *Genes (Basel),* 10(4). doi: 10.3390/genes10040321.

Kittelmann, S. B., A. D., Franke, F. A., Almudi, I., Yoth, M., Sabaris, G., Couso, J. P., Nunes, M. D. S., Frankel, N., Gomez-Skarmeta, J. L., Pueyo-Marques, J., Arif, S. and McGregor, A. P. (2018) 'Gene regulatory network architecture in different developmental contexts influences the genetic basis of morphological evolution', *PLoS Genet,* 14(5), pp. e1007375. doi: 10.1371/journal.pgen.1007375.

Kleftogiannis, D., Kalnis, P. and Bajic, V. B. (2016) 'Progress and challenges in bioinformatics approaches for enhancer identification', *Brief Bioinform: Vol. 6*, pp. 967-79.

Kojima, T., Sato, M. and Saigo, K. (2000) 'Formation and specification of distal leg segments in Drosophila by dual Bar homeobox genes, BarH1 and BarH2', *Development,* 127(4), pp. 769-78.

Kondo, T., Plaza, S., Zanet, J., Benrabah, E., Valenti, P., Hashimoto, Y., Kobayashi, S., Payre, F. and Kageyama, Y. (2010) 'Small peptides switch the transcriptional activity of Shavenbaby during Drosophila embryogenesis', *Science,* 329(5989), pp. 336-9. doi: 10.1126/science.1188158.

Koshikawa, S., Giorgianni, M. W., Vaccaro, K., Kassner, V. A., Yoder, J. H., Werner, T. and Carroll, S. B. (2015) 'Gain of cis-regulatory activities underlies novel domains of wingless gene expression in Drosophila', *Proc Natl Acad Sci U S A,* 112(24), pp. 7524-9. doi: 10.1073/pnas.1509022112.

Kunarso, G., Chia, N. Y., Jeyakani, J., Hwang, C., Lu, X., Chan, Y. S., Ng, H. H. and Bourque, G. (2010) 'Transposable elements have rewired the core regulatory network of human embryonic stem cells', *Nat Genet,* 42(7), pp. 631-4. doi: 10.1038/ng.600.

Kvon, E. Z. (2015) 'Using transgenic reporter assays to functionally characterize enhancers in animals', *Genomics,* 106(3), pp. 185-192. doi: 10.1016/j.ygeno.2015.06.007.

Kyrchanova, O., Mogila, V., Wolle, D., Magbanua, J. P., White, R., Georgiev, P. and Schedl, P. (2015) 'The boundary paradox in the Bithorax complex', *Mech Dev,* 138 Pt 2, pp. 122-132. doi: 10.1016/j.mod.2015.07.002.

Le, D. D., Shimko, T. C., Aditham, A. K., Keys, A. M., Longwell, S. A., Orenstein, Y. and Fordyce, P. M. (2018) 'Comprehensive, high-resolution binding energy landscapes reveal context dependencies of transcription factor binding', *Proc Natl Acad Sci U S A,* 115(16), pp. E3702-e3711. doi: 10.1073/pnas.1715888115.

Lebrecht, D., Foehr, M., Smith, E., Lopes, F. J., Vanario-Alonso, C. E., Reinitz, J., Burz, D. S. and Hanes, S. D. (2005) 'Bicoid cooperative DNA binding is critical for embryonic patterning in Drosophila', *Proc Natl Acad Sci U S A,* 102(37), pp. 13176-81. doi: 10.1073/pnas.0506462102.

Lee, W., Haslinger, A., Karin, M. and Tjian, R. (1987) 'Activation of transcription by two factors that bind promoter and enhancer sequences of the human metallothionein gene and SV40', *Nature,* 325(6102), pp. 368-72. doi: 10.1038/325368a0.

Lehmann, R. and Nusslein-Volhard, C. (1987) 'hunchback, a gene required for segmentation of an anterior and posterior region of the Drosophila embryo', *Dev Biol,* 119(2), pp. 402-17. doi: 10.1016/0012-1606(87)90045-5.

Letelier, J., de la Calle-Mustienes, E., Pieretti, J., Naranjo, S., Maeso, I., Nakamura, T., Pascual-Anaya, J., Shubin, N. H., Schneider, I., Martinez-Morales, J. R. and Gomez-Skarmeta, J. L. (2018) 'A conserved Shh cis-regulatory module highlights a common developmental origin of unpaired and paired fins', *Nat Genet,* 50(4), pp. 504-509. doi: 10.1038/s41588-018-0080-5.

Lettice, L. A., Heaney, S. J., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A., Goode, D., Elgar, G., Hill, R. E. and de Graaff, E. (2003) 'A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly', *Hum Mol Genet,* 12(14), pp. 1725-35. doi: 10.1093/hmg/ddg180.

Li, X. Y. and Eisen, M. B. (2018) 'Mutation of sequences flanking and separating transcription factor binding sites in a Drosophila enhancer significantly alter its output', *Biorxiv*. doi: 10.1101/379974.

Li, X. Y., MacArthur, S., Bourgon, R., Nix, D., Pollard, D. A., Iyer, V. N., Hechmer, A., Simirenko, L., Stapleton, M., Luengo Hendriks, C. L., Chu, H. C., Ogawa, N., Inwood, W., Sementchenko, V., Beaton, A., Weiszmann, R., Celniker, S. E., Knowles, D. W., Gingeras, T., Speed, T. P., Eisen, M. B. and Biggin, M. D. (2008) 'Transcription factors bind thousands of active and inactive regions in the Drosophila blastoderm', *PLoS Biol,* 6(2), pp. e27. doi: 10.1371/journal.pbio.0060027.

Li, Z., Schulz, M. H., Look, T., Begemann, M., Zenke, M. and Costa, I. G. (2019) 'Identification of transcription factor binding sites using ATAC-seq', *Genome Biol,* 20(1), pp. 45. doi: 10.1186/s13059-019-1642-2.

Liang, H. L., Nien, C. Y., Liu, H. Y., Metzstein, M. M., Kirov, N. and Rushlow, C. (2008) 'The zinc-finger protein Zelda is a key activator of the early zygotic genome in Drosophila', *Nature,* 456(7220), pp. 400-3. doi: 10.1038/nature07388.

Lionnet, T., Czaplinski, K., Darzacq, X., Shav-Tal, Y., Wells, A. L., Chao, J. A., Park, H. Y., de Turris, V., Lopez-Jones, M. and Singer, R. H. (2011) 'A transgenic mouse for in vivo detection of endogenous labeled mRNA', *Nat Methods,* 8(2), pp. 165-70. doi: 10.1038/nmeth.1551.

Liu, Q., Onal, P., Datta, R. R., Rogers, J. M., Schmidt-Ott, U., Bulyk, M. L., Small, S. and Thornton, J. W. (2018) 'Ancient mechanisms for the evolution of the bicoid homeodomain's function in fly development', *eLife*.

Loehlin, D. W. and Werren, J. H. (2012) 'Evolution of shape by multiple regulatory changes to a growth gene', *Science,* 335(6071), pp. 943-7. doi: 10.1126/science.1215193.

Lohmann, I. and McGinnis, W. (2002) 'Hox Genes: it's all a matter of context', *Curr Biol,* 12(15), pp. R514-6.

Long, H. K., Prescott, S. L. and Wysocka, J. (2016) 'Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution', *Cell,* 167(5), pp. 1170-1187. doi: 10.1016/j.cell.2016.09.018.

Lucas, T., Ferraro, T., Roelens, B., De Las Heras Chanes, J., Walczak, A. M., Coppey, M. and Dostatni, N. (2013) 'Live imaging of bicoid-dependent transcription in Drosophila embryos', *Curr Biol,* 23(21), pp. 2135-9. doi: 10.1016/j.cub.2013.08.053.

Ludwig, M. Z., Bergman, C., Patel, N. H. and Kreitman, M. (2000) 'Evidence for stabilizing selection in a eukaryotic enhancer element', *Nature,* 403(6769), pp. 564-7. doi: 10.1038/35000615.

Ludwig, M. Z. and Kreitman, M. (1995) 'Evolutionary dynamics of the enhancer region of even-skipped in Drosophila', *Mol Biol Evol,* 12(6), pp. 1002-11. doi: 10.1093/oxfordjournals.molbev.a040277.

Ludwig, M. Z., Patel, N. H. and Kreitman, M. (1998) 'Functional analysis of eve stripe 2 enhancer evolution in Drosophila: rules governing conservation and change', *Development,* 125(5), pp. 949-58.

Lukowitz, W., Schroder, C., Glaser, G., Hulskamp, M. and Tautz, D. (1994) 'Regulatory and coding regions of the segmentation gene hunchback are functionally conserved between Drosophila virilis and Drosophila melanogaster', *Mech Dev,* 45(2), pp. 105-15.

Lynch, V. J. and Wagner, G. P. (2008) 'Resurrecting the role of transcription factor change in developmental evolution', *Evolution,* 62(9), pp. 2131-54. doi: 10.1111/j.1558-5646.2008.00440.x.

Ma, X., Yuan, D., Diepold, K., Scarborough, T. and Ma, J. (1996) 'The Drosophila morphogenetic protein Bicoid binds DNA cooperatively', *Development,* 122(4), pp. 1195-206.

MacArthur, S. and Brookfield, J. F. (2004) 'Expected rates and modes of evolution of enhancer sequences', *Mol Biol Evol,* 21(6), pp. 1064-73. doi: 10.1093/molbev/msh105.

Mace, K. A., Pearson, J. C. and McGinnis, W. (2005) 'An epidermal barrier wound repair pathway in Drosophila is mediated by grainy head', *Science,* 308(5720), pp. 381-5. doi: 10.1126/science.1107573.

Madeira, F., Park, Y. M., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A. R. N., Potter, S. C., Finn, R. D. and Lopez, R. (2019) 'The EMBL-EBI search and sequence analysis tools APIs in 2019', *Nucleic Acids Res,* 47(W1), pp. W636-w641. doi: 10.1093/nar/gkz268.

Maeda, R. K. and Karch, F. (2006) 'The ABC of the BX-C: the bithorax complex explained', *Development,* 133(8), pp. 1413-22. doi: 10.1242/dev.02323.

Maeda, R. K. and Karch, F. (2009) 'The bithorax complex of Drosophila an exceptional Hox cluster', *Curr Top Dev Biol,* 88, pp. 1-33. doi: 10.1016/s0070-2153(09)88001-0.

Maeso, I., Irimia, M., Tena, J. J., Casares, F. and Gomez-Skarmeta, J. L. (2013) 'Deep conservation of cis-regulatory elements in metazoans', *Philos Trans R Soc Lond B Biol Sci,* 368(1632), pp. 20130020. doi: 10.1098/rstb.2013.0020.

Maeso, I. and Tena, J. J. (2016) 'Favorable genomic environments for cis-regulatory evolution: A novel theoretical framework', *Semin Cell Dev Biol,* 57, pp. 2-10. doi: 10.1016/j.semcdb.2015.12.003.

Mallo, M. and Alonso, C. R. (2013) 'The regulation of Hox gene expression during animal development', *Development,* 140(19), pp. 3951-63. doi: 10.1242/dev.068346.

Mallo, M., Wellik, D. M. and Deschamps, J. (2010) 'Hox genes and regional patterning of the vertebrate body plan', *Dev Biol,* 344(1), pp. 7-15. doi: 10.1016/j.ydbio.2010.04.024.

Mann, R. S., Lelli, K. M. and Joshi, R. (2009) 'Hox specificity unique roles for cofactors and collaborators', *Curr Top Dev Biol,* 88, pp. 63-101. doi: 10.1016/s0070-2153(09)88003-4.

Marcellini, S. and Simpson, P. (2006) 'Two or Four Bristles: Functional Evolution of an Enhancer of scute in Drosophilidae', *PLoS Biol: Vol. 12*.

Markstein, M., Zinzen, R., Markstein, P., Yee, K. P., Erives, A., Stathopoulos, A. and Levine, M. (2004) 'A regulatory code for neurogenic gene expression in the Drosophila embryo', *Development,* 131(10), pp. 2387-94. doi: 10.1242/dev.01124.

Martin, A., McCulloch, K. J., Patel, N. H., Briscoe, A. D., Gilbert, L. E. and Reed, R. D. (2014) 'Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations', *Evodevo,* 5(1), pp. 7. doi: 10.1186/2041-9139-5-7.

Martin, A. and Orgogozo, V. (2013) 'The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation', *Evolution,* 67(5), pp. 1235-50. doi: 10.1111/evo.12081.

Martinez, C., Rest, J. S., Kim, A. R., Ludwig, M., Kreitman, M., White, K. and Reinitz, J. (2014) 'Ancestral resurrection of the Drosophila S2E enhancer reveals accessible evolutionary paths through compensatory change', *Mol Biol Evol,* 31(4), pp. 903-16. doi: 10.1093/molbev/msu042.

McGregor, A. P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D. G., Payre, F. and Stern, D. L. (2007) 'Morphological evolution through multiple cis-regulatory mutations at a single gene', *Nature,* 448(7153), pp. 587-90. doi: 10.1038/nature05988.

McGregor, A. P., Shaw, P. J., Hancock, J. M., Bopp, D., Hediger, M., Wratten, N. S. and Dover, G. A. (2001) 'Rapid restructuring of bicoid-dependent hunchback promoters within and between Dipteran species: implications for molecular coevolution', *Evol Dev,* 3(6), pp. 397-407.

Menoret, D., Santolini, M., Fernandes, I., Spokony, R., Zanet, J., Gonzalez, I., Latapie, Y., Ferrer, P., Rouault, H., White, K. P., Besse, P., Hakim, V., Aerts, S., Payre, F. and Plaza, S. (2013) 'Genome-wide analyses of Shavenbaby target genes reveals distinct features of enhancer organization', *Genome Biol,* 14(8), pp. R86. doi: 10.1186/gb-2013-14-8-r86.

Mevel-Ninio, M., Terracol, R., Salles, C., Vincent, A. and Payre, F. (1995) 'ovo, a Drosophila gene required for ovarian development, is specifically expressed in the germline and shares most of its coding sequences with shavenbaby, a gene involved in embryo patterning', *Mech Dev,* 49(1-2), pp. 83-95.

Moses, A. M., Pollard, D. A., Nix, D. A., Iyer, V. N., Li, X. Y., Biggin, M. D. and Eisen, M. B. (2006) 'Large-scale turnover of functional transcription factor binding sites in Drosophila', *PLoS Comput Biol,* 2(10), pp. e130. doi: 10.1371/journal.pcbi.0020130.

Mundade, R., Ozer, H. G., Wei, H., Prabhu, L. and Lu, T. (2014) 'Role of ChIP-seq in the discovery of transcription factor binding sites, differential gene regulation mechanism, epigenetic marks and beyond', *Cell Cycle: Vol. 18*, pp. 2847-52.

Ng, T., Yu, F. and Roy, S. (2006) 'A homologue of the vertebrate SET domain and zinc finger protein Blimp-1 regulates terminal differentiation of the tracheal system in the Drosophila embryo', *Dev Genes Evol,* 216(5), pp. 243-52. doi: 10.1007/s00427-005-0044-5.

Nien, C. Y., Liang, H. L., Butcher, S., Sun, Y., Fu, S., Gocha, T., Kirov, N., Manak, J. R. and Rushlow, C. (2011) 'Temporal coordination of gene networks by Zelda in the early

Drosophila embryo', *PLoS Genet,* 7(10), pp. e1002339. doi: 10.1371/journal.pgen.1002339.

Nord, A. S., Blow, M. J., Attanasio, C., Akiyama, J. A., Holt, A., Hosseini, R., Phouanenavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V., Rubenstein, J. L., Rubin, E. M., Pennacchio, L. A. and Visel, A. (2013) 'Rapid and pervasive changes in genome-wide enhancer usage during mammalian development', *Cell,* 155(7), pp. 1521-31. doi: 10.1016/j.cell.2013.11.033.

Nourmohammad, A. and Lassig, M. (2011) 'Formation of regulatory modules by local sequence duplication', *PLoS Comput Biol,* 7(10), pp. e1002167. doi: 10.1371/journal.pcbi.1002167.

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(7), pp. 1277-89. doi: 10.1016/j.cell.2008.05.023.

Nunes, M. D., Arif, S., Schlotterer, C. and McGregor, A. P. (2013) 'A perspective on micro-evo-devo: progress and potential', *Genetics,* 195(3), pp. 625-34. doi: 10.1534/genetics.113.156463.

Nusslein-Volhard, C. and Wieschaus, E. (1980) 'Mutations affecting segment number and polarity in Drosophila', *Nature,* 287(5785), pp. 795-801. doi: 10.1038/287795a0.

O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D. and Kingsley, D. M. (2015) 'A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA', *Elife,* 4, pp. e05290. doi: 10.7554/eLife.05290.

Odom, D. T., Dowell, R. D., Jacobsen, E. S., Gordon, W., Danford, T. W., MacIsaac, K. D., Rolfe, P. A., Conboy, C. M., Gifford, D. K. and Fraenkel, E. (2007) 'Tissue-specific transcriptional regulation has diverged significantly between human and mouse', *Nat Genet,* 39(6), pp. 730-2. doi: 10.1038/ng2047.

Okonechnikov, K., Golosova, O. and Fursov, M. (2012) 'Unipro UGENE: a unified bioinformatics toolkit', *Bioinformatics,* 28(8), pp. 1166-7. doi: 10.1093/bioinformatics/bts091.

Osterwalder, M., Barozzi, I., Tissieres, V., Fukuda-Yuzawa, Y., Mannion, B. J., Afzal, S. Y., Lee, E. A., Zhu, Y., Plajzer-Frick, I., Pickle, C. S., Kato, M., Garvin, T. H., Pham, Q. T., Harrington, A. N., Akiyama, J. A., Afzal, V., Lopez-Rios, J., Dickel, D. E., Visel, A. and Pennacchio, L. A. (2018) 'Enhancer redundancy provides phenotypic robustness in mammalian development', *Nature,* 554(7691), pp. 239-243. doi: 10.1038/nature25461.

Ozturk-Colak, A., Stephan-Otto Attolini, C., Casanova, J. and Araujo, S. J. (2018) 'Blimp-1 Mediates Tracheal Lumen Maturation in Drosophila melanogaster', *Genetics,* 210(2), pp. 653-663. doi: 10.1534/genetics.118.301444.

Paris, M., Kaplan, T., Li, X. Y., Villalta, J. E., Lott, S. E. and Eisen, M. B. (2013) 'Extensive divergence of transcription factor binding in Drosophila embryos with highly conserved gene expression', *PLoS Genet,* 9(9), pp. e1003748. doi: 10.1371/journal.pgen.1003748.

Park, J., Estrada, J., Johnson, G., Vincent, B. J., Ricci-Tam, C., Bragdon, M. D., Shulgina, Y., Cha, A., Wunderlich, Z., Gunawardena, J. and DePace, A. H. (2019) 'Dissecting the sharp response of a canonical developmental enhancer reveals multiple sources of cooperativity', *Elife,* 8. doi: 10.7554/eLife.41266.

Pataki, C., Matusek, T., Kurucz, E., Ando, I., Jenny, A. and Mihaly, J. (2010) 'Drosophila Rab23 is involved in the regulation of the number and planar polarization of the adult cuticular hairs', *Genetics,* 184(4), pp. 1051-65. doi: 10.1534/genetics.109.112060.

Payne, J. L. and Wagner, A. (2014) 'The robustness and evolvability of transcription factor binding sites', *Science,* 343(6173), pp. 875-7. doi: 10.1126/science.1249046.

Payne, J. L. and Wagner, A. (2015) 'Mechanisms of mutational robustness in transcriptional regulation', *Front Genet,* 6. doi: 10.3389/fgene.2015.00322.

Payre, F. (2004) 'Genetic control of epidermis differentiation in Drosophila', *Int J Dev Biol,* 48(2-3), pp. 207-15. doi: 10.1387/ijdb.041828fp.

Payre, F., Vincent, A. and Carreno, S. (1999) 'ovo/svb integrates Wingless and DER pathways to control epidermis differentiation', *Nature,* 400(6741), pp. 271-5. doi: 10.1038/22330.

Pearson, J. C., Lemons, D. and McGinnis, W. (2005) 'Modulating Hox gene functions during animal body patterning', *Nat Rev Genet,* 6(12), pp. 893-904. doi: 10.1038/nrg1726.

Pennacchio, L. A., Ahituv, N., Moses, A. M., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A. and Rubin, E. M. (2006) 'In vivo enhancer analysis of human conserved non-coding sequences', *Nature,* 444(7118), pp. 499-502. doi: 10.1038/nature05295.

Pennacchio, L. A., Loots, G. G., Nobrega, M. A. and Ovcharenko, I. (2007) 'Predicting tissue-specific enhancers in the human genome', *Genome Res,* 17(2), pp. 201-11. doi: 10.1101/gr.5972507.

Perry, M. W., Boettiger, A. N., Bothma, J. P. and Levine, M. (2010) 'Shadow enhancers foster robustness of Drosophila gastrulation', *Curr Biol,* 20(17), pp. 1562-7. doi: 10.1016/j.cub.2010.07.043.

Perry, M. W., Boettiger, A. N. and Levine, M. (2011) 'Multiple enhancers ensure precision of gap gene-expression patterns in the Drosophila embryo', *Proc Natl Acad Sci U S A,* 108(33), pp. 13570-5. doi: 10.1073/pnas.1109873108.

Perry, M. W., Bothma, J. P., Luu, R. D. and Levine, M. (2012) 'Precision of Hunchback expression in the Drosophila embryo', *Curr Biol,* 22(23), pp. 2247-52. doi: 10.1016/j.cub.2012.09.051.

Pfeiffer, B. D., Jenett, A., Hammonds, A. S., Ngo, T. T., Misra, S., Murphy, C., Scully, A., Carlson, J. W., Wan, K. H., Laverty, T. R., Mungall, C., Svirskas, R., Kadonaga, J. T., Doe, C. Q., Eisen, M. B., Celniker, S. E. and Rubin, G. M. (2008) 'Tools for neuroanatomy and neurogenetics in Drosophila', *Proc Natl Acad Sci U S A,* 105(28), pp. 9715-20. doi: 10.1073/pnas.0803697105.

Porcelli, D., Fischer, B., Russell, S. and White, R. (2019) 'Chromatin accessibility plays a key role in selective targeting of Hox proteins', *Genome Biology,* 20(1), pp. 115. doi: doi:10.1186/s13059-019-1721-4.

Prabhakar, S., Poulin, F., Shoukry, M., Afzal, V., Rubin, E. M., Couronne, O. and Pennacchio, L. A. (2006) 'Close sequence comparisons are sufficient to identify human cis-regulatory elements', *Genome Res,* 16(7), pp. 855-63. doi: 10.1101/gr.4717506.

Prasad, N., Tarikere, S., Khanale, D., Habib, F. and Shashidhara, L. S. (2016) 'A comparative genomic analysis of targets of Hox protein Ultrabithorax amongst distant insect species', *Sci Rep,* 6, pp. 27885. doi: 10.1038/srep27885.

Preger-Ben Noon, E., Davis, F. P. and Stern, D. L. (2016) 'Evolved Repression Overcomes Enhancer Robustness', *Dev Cell,* 39(5), pp. 572-584. doi: 10.1016/j.devcel.2016.10.010.

Preger-Ben Noon, E., Sabaris, G., Ortiz, D. M., Sager, J., Liebowitz, A., Stern, D. L. and Frankel, N. (2018) 'Comprehensive Analysis of a cis-Regulatory Region Reveals Pleiotropy in Enhancer Function', *Cell Rep,* 22(11), pp. 3021-3031. doi: 10.1016/j.celrep.2018.02.073.

Prud'homme, B., Gompel, N., Rokas, A., Kassner, V. A., Williams, T. M., Yeh, S. D., True, J. R. and Carroll, S. B. (2006) 'Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene', *Nature,* 440(7087), pp. 1050-3. doi: 10.1038/nature04597.

Quillien, A., Abdalla, M., Yu, J., Ou, J., Zhu, L. J. and Lawson, N. D. (2017) 'Robust Identification of Developmentally Active Endothelial Enhancers in Zebrafish Using FANS-Assisted ATAC-Seq', *Cell Rep,* 20(3), pp. 709-720. doi: 10.1016/j.celrep.2017.06.070.

Quinlan, A. R. and Hall, I. M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics,* 26(6), pp. 841-2. doi: 10.1093/bioinformatics/btq033.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A. and Wysocka, J. (2011) 'A unique chromatin signature uncovers early developmental enhancers in humans', *Nature,* 470(7333), pp. 279-83. doi: 10.1038/nature09692.

Randsholt, N. B., Maschat, F. and Santamaria, P. (2000) 'polyhomeotic controls engrailed expression and the hedgehog signaling pathway in imaginal discs', *Mech Dev,* 95(1-2), pp. 89-99. doi: 10.1016/s0925-4773(00)00342-7.

Reed, R. D., Papa, R., Martin, A., Hines, H. M., Counterman, B. A., Pardo-Diaz, C., Jiggins, C. D., Chamberlain, N. L., Kronforst, M. R., Chen, R., Halder, G., Nijhout, H. F. and McMillan, W. O. (2011) 'optix drives the repeated convergent evolution of butterfly wing pattern mimicry', *Science,* 333(6046), pp. 1137-41. doi: 10.1126/science.1208227.

Rickels, R. and Shilatifard, A. (2018) 'Enhancer Logic and Mechanics in Development and Disease', *Trends Cell Biol,* 28(8), pp. 608-630. doi: 10.1016/j.tcb.2018.04.003.

Risca, V. I. and Greenleaf, W. J. (2015) 'Unraveling the 3D genome: genomics tools for multiscale exploration', *Trends Genet,* 31(7), pp. 357-72. doi: 10.1016/j.tig.2015.03.010.

Rister, J. and Desplan, C. (2010) 'Deciphering the genome's regulatory code: the many languages of DNA', *Bioessays,* 32(5), pp. 381-4. doi: 10.1002/bies.200900197.

Rohith, B. N. and Shyamala, B. V. (2017) 'Scalloped a member of the Hippo tumor suppressor pathway controls mushroom body size in Drosophila brain by non-canonical regulation of neuroblast proliferation', *Dev Biol,* 432(2), pp. 203-214. doi: 10.1016/j.ydbio.2017.10.016.

Rowan, S., Siggers, T., Lachke, S. A., Yue, Y., Bulyk, M. L. and Maas, R. L. (2010) 'Precise temporal control of the eye regulatory gene Pax6 via enhancer-binding site affinity', *Genes Dev,* 24(10), pp. 980-5. doi: 10.1101/gad.1890410.

Rozowski, M. and Akam, M. (2002) 'Hox gene control of segment-specific bristle patterns in Drosophila', *Genes Dev,* 16(9), pp. 1150-62. doi: 10.1101/gad.219302.

Ruby, J. G., Stark, A., Johnston, W. K., Kellis, M., Bartel, D. P. and Lai, E. C. (2007) 'Evolution, biogenesis, expression, and target predictions of a substantially expanded set of Drosophila microRNAs', *Genome Res,* 17(12), pp. 1850-64. doi: 10.1101/gr.6597907.

Rux, D. R. and Wellik, D. M. (2017) 'Hox Genes in the Adult Skeleton: Novel Functions Beyond Embryonic Development', *Dev Dyn,* 246(4), pp. 310-7. doi: 10.1002/dvdy.24482.

Savitsky, M., Kim, M., Kravchuk, O. and Schwartz, Y. B. (2016) 'Distinct Roles of Chromatin Insulator Proteins in Control of the Drosophila Bithorax Complex', *Genetics,* 202(2), pp. 601-17. doi: 10.1534/genetics.115.179309.

Schertel, C., Rutishauser, T., Forstemann, K. and Basler, K. (2012) 'Functional characterization of Drosophila microRNAs by a novel in vivo library', *Genetics,* 192(4), pp. 1543-52. doi: 10.1534/genetics.112.145383.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J. Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P. and Cardona, A. (2012) 'Fiji: an open-source platform for biological-image analysis', *Nat Methods,* 9(7), pp. 676-82. doi: 10.1038/nmeth.2019.

Schmidt, D., Wilson, M. D., Ballester, B., Schwalie, P. C., Brown, G. D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C. P., Mackay, S., Talianidis, I., Flicek, P. and Odom, D. T. (2010) 'Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding', *Science,* 328(5981), pp. 1036-40. doi: 10.1126/science.1186176.

Schroder, C., Tautz, D., Seifert, E. and Jackle, H. (1988) 'Differential regulation of the two transcripts from the Drosophila gap segmentation gene hunchback', *Embo j,* 7(9), pp. 2881-7.

Shaw, P. J., Wratten, N. S., McGregor, A. P. and Dover, G. A. (2002) 'Coevolution in bicoid-dependent promoters and the inception of regulatory incompatibilities among species of higher Diptera', *Evol Dev,* 4(4), pp. 265-77.

Shlyueva, D., Stampfel, G. and Stark, A. (2014) 'Transcriptional enhancers: from properties to genome-wide predictions', *Nat Rev Genet,* 15(4), pp. 272-86. doi: 10.1038/nrg3682.

Shroff, S., Joshi, M. and Orenic, T. V. (2007) 'Differential Delta expression underlies the diversity of sensory organ patterns among the legs of the Drosophila adult', *Mech Dev,* 124(1), pp. 43-58. doi: 10.1016/j.mod.2006.09.004.

Simpson, P. and Ayyar, S. (2008) 'Evolution of Cis-Regulatory Sequences in Drosophila', *Advances in Genetics,* 61(1), pp. 67-106. doi: 10.1016/S0065-2660(07)00003-X.

Slattery, M., Negre, N. and White, K. P. (2012) 'Interpreting the regulatory genome: the genomics of transcription factor function in Drosophila melanogaster', *Brief Funct Genomics,* 11(5), pp. 336-46. doi: 10.1093/bfgp/els034.

Slattery, M. and White, K. P. (2013) 'Enhanced dissection of the regulatory genome', *Nat Methods,* 10(8), pp. 710-2. doi: 10.1038/nmeth.2577.

Spitz, F. and Furlong, E. E. (2012) 'Transcription factors: from enhancer binding to developmental control', *Nat Rev Genet,* 13(9), pp. 613-26. doi: 10.1038/nrg3207.

Stefflova, K., Thybert, D., Wilson, M. D., Streeter, I., Aleksic, J., Karagianni, P., Brazma, A., Adams, D. J., Talianidis, I., Marioni, J. C., Flicek, P. and Odom, D. T. (2013) 'Cooperativity and rapid evolution of cobound transcription factors in closely related mammals', *Cell,* 154(3), pp. 530-40. doi: 10.1016/j.cell.2013.07.007.

Stern, D. L. (1998) 'A role of Ultrabithorax in morphological differences between Drosophila species', *Nature,* 396(6710), pp. 463-6. doi: 10.1038/24863.

Stern, D. L. (2003) 'The Hox gene Ultrabithorax modulates the shape and size of the third leg of Drosophila by influencing diverse mechanisms', *Dev Biol,* 256(2), pp. 355-66. doi: 10.1016/s0012-1606(03)00035-6.

Stern, D. L. (2011) *Evolution, Development, & the Predictable Genome.* Greenwood Village: Roberts and Company.

Stern, D. L. and Frankel, N. (2013) 'The structure and evolution of cis-regulatory regions: the shavenbaby story', *Philos Trans R Soc Lond B Biol Sci,* 368(1632), pp. 20130028. doi: 10.1098/rstb.2013.0028.

Stern, D. L. and Orgogozo, V. (2008) 'The loci of evolution: how predictable is genetic evolution?', *Evolution,* 62(9), pp. 2155-77. doi: 10.1111/j.1558-5646.2008.00450.x.

Stern, D. L. and Orgogozo, V. (2009) 'Is genetic evolution predictable?', *Science,* 323(5915), pp. 746-51. doi: 10.1126/science.1158997.

Stewart, A. J., Hannenhalli, S. and Plotkin, J. B. (2012) 'Why transcription factor binding sites are ten nucleotides long', *Genetics,* 192(3), pp. 973-85. doi: 10.1534/genetics.112.143370.

Struhl, G., Struhl, K. and Macdonald, P. M. (1989) 'The gradient morphogen bicoid is a concentration-dependent transcriptional activator', *Cell,* 57(7), pp. 1259-73. doi: 10.1016/0092-8674(89)90062-7.

Strutt, D. and Warrington, S. J. (2008) 'Planar polarity genes in the Drosophila wing regulate the localisation of the FH3-domain protein Multiple Wing Hairs to control the site of hair production', *Development,* 135(18), pp. 3103-11. doi: 10.1242/dev.025205.

Sucena, E., Delon, I., Jones, I., Payre, F. and Stern, D. L. (2003) 'Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism', *Nature,* 424(6951), pp. 935-8. doi: 10.1038/nature01768.

Sucena, E. and Stern, D. L. (2000) 'Divergence of larval morphology between Drosophila sechellia and its sibling species caused by cis-regulatory evolution of ovo/shaven-baby', *Proc Natl Acad Sci U S A,* 97(9), pp. 4530-4. doi: 10.1073/pnas.97.9.4530.

Suryamohan, K. and Halfon, M. S. (2015) 'Identifying transcriptional cis-regulatory modules in animal genomes', *Wiley Interdiscip Rev Dev Biol,* 4(2), pp. 59-84. doi: 10.1002/wdev.168.

Swanson, C. I., Schwimmer, D. B. and Barolo, S. (2011) 'Rapid evolutionary rewiring of a structurally constrained eye enhancer', *Curr Biol,* 21(14), pp. 1186-96. doi: 10.1016/j.cub.2011.05.056.

Tanay, A. (2006) 'Extensive low-affinity transcriptional interactions in the yeast genome', *Genome Res,* 16(8), pp. 962-72. doi: 10.1101/gr.5113606.

Tautz, D. (1988) 'Regulation of the Drosophila segmentation gene hunchback by two maternal morphogenetic centres', *Nature,* 332(6161), pp. 281-4. doi: 10.1038/332281a0.

Tautz, D., Lehmann, R., Schnürch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jäckle, H. (1987) 'Finger protein of novel structure encoded by hunchback , a second member of the gap class of Drosophila segmentation genes', *Nature,* 327(6121), pp. 383-389. doi: doi:10.1038/327383a0.

Tautz, D. and Pfeifle, C. (1989) 'A non-radioactive in situ hybridization method for the localization of specific RNAs in Drosophila embryos reveals translational control of the segmentation gene hunchback', *Chromosoma,* 98(2), pp. 81-5. doi: 10.1007/bf00291041.

Tautz, D., Trick, M. and Dover, G. A. (1986) 'Cryptic simplicity in DNA is a major source of genetic variation', *Nature,* 322(6080), pp. 652-6. doi: 10.1038/322652a0.

ten Bosch, J. R., Benavides, J. A. and Cline, T. W. (2006) 'The TAGteam DNA motif controls the timing of Drosophila pre-blastoderm transcription', *Development,* 133(10), pp. 1967-77. doi: 10.1242/dev.02373.

Treisman, J., Gonczy, P., Vashishtha, M., Harris, E. and Desplan, C. (1989) 'A single amino acid can determine the DNA binding specificity of homeodomain proteins', *Cell,* 59(3), pp. 553-62. doi: 10.1016/0092-8674(89)90038-x.

True, J. R. and Haag, E. S. (2001) 'Developmental system drift and flexibility in evolutionary trajectories', *Evol Dev,* 3(2), pp. 109-19.

Tsompana, M. and Buck, M. J. (2014) 'Chromatin accessibility: a window into the genome', *Epigenetics Chromatin,* 7(1), pp. 33. doi: 10.1186/1756-8935-7-33.

Tugrul, M., Paixao, T., Barton, N. H. and Tkacik, G. (2015) 'Dynamics of Transcription Factor Binding Site Evolution', *PLoS Genet,* 11(11), pp. e1005639. doi: 10.1371/journal.pgen.1005639.

Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y. and Kondoh, H. (2003) 'Functional analysis of chicken Sox2 enhancers highlights an array of diverse regulatory elements that are conserved in mammals', *Dev Cell,* 4(4), pp. 509-19.

Villar, D., Flicek, P. and Odom, D. T. (2014) 'Evolution of transcription factor binding in metazoans - mechanisms and functional implications', *Nat Rev Genet,* 15(4), pp. 221-33. doi: 10.1038/nrg3481.

Vincent, B. J., Estrada, J. and DePace, A. H. (2016) 'The appeasement of Doug: a synthetic approach to enhancer biology', *Integr Biol (Camb),* 8(4), pp. 475-84. doi: 10.1039/c5ib00321k.

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E. M. and Pennacchio, L. A. (2009) 'ChIP-seq accurately predicts tissue-specific activity of enhancers', *Nature,* 457(7231), pp. 854-8. doi: 10.1038/nature07730.

Visel, A., Minovitsky, S., Dubchak, I. and Pennacchio, L. A. (2007) 'VISTA Enhancer Browser--a database of tissue-specific human enhancers', *Nucleic Acids Res,* 35(Database issue), pp. D88-92. doi: 10.1093/nar/gkl822.

Wang, X. and Chamberlin, H. M. (2004) 'Evolutionary innovation of the excretory system in Caenorhabditis elegans', *Nat Genet,* 36(3), pp. 231-2. doi: 10.1038/ng1301.

Wieschaus, E., Nusslein-Volhard, C. and Jurgens, G. (1984) 'Mutations affecting the pattern of the larval cuticle inDrosophila melanogaster : III. Zygotic loci on the X-chromosome and fourth chromosome', *Wilehm Roux Arch Dev Biol,* 193(5), pp. 296-307. doi: 10.1007/bf00848158.

Wittkopp, P. J. and Kalay, G. (2011) 'Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence', *Nat Rev Genet,* 13(1), pp. 59-69. doi: 10.1038/nrg3095.

Wong, L. L. and Adler, P. N. (1993) 'Tissue polarity genes of Drosophila regulate the subcellular location for prehair initiation in pupal wing cells', *J Cell Biol,* 123(1), pp. 209-21. doi: 10.1083/jcb.123.1.209.

Wratten, N. S., McGregor, A. P., Shaw, P. J. and Dover, G. A. (2006) 'Evolutionary and functional analysis of the tailless enhancer in Musca domestica and Drosophila melanogaster', *Evol Dev,* 8(1), pp. 6-15. doi: 10.1111/j.1525-142X.2006.05070.x.

Wray, G. A. (2006) 'The evolution of embryonic gene expression in sea urchins', *Integr Comp Biol,* 46(3), pp. 233-42. doi: 10.1093/icb/icj030.

Wunderlich, Z., Bragdon, M. D., Vincent, B. J., White, J. A., Estrada, J. and DePace, A. H. (2015) 'Krüppel expression levels are maintained through compensatory evolution of shadow enhancers', *Cell Rep,* 12(11), pp. 1740-7. doi: 10.1016/j.celrep.2015.08.021.

Xu, Z., Chen, H., Ling, J., Yu, D., Struffi, P. and Small, S. (2014) 'Impacts of the ubiquitous factor Zelda on Bicoid-dependent DNA binding and transcription in Drosophila', *Genes Dev,* 28(6), pp. 608-21. doi: 10.1101/gad.234534.113.

Yan, J., Huen, D., Morely, T., Johnson, G., Gubb, D., Roote, J. and Adler, P. N. (2008) 'The multiple-wing-hairs gene encodes a novel GBD-FH3 domain-containing protein that functions both prior to and after wing hair initiation', *Genetics,* 180(1), pp. 219-28. doi: 10.1534/genetics.108.091314.

Yuan, D., Ma, X. and Ma, J. (1996) 'Sequences outside the homeodomain of bicoid are required for protein-protein interaction', *J Biol Chem,* 271(35), pp. 21660-5.

Yuan, D., Ma, X. and Ma, J. (1999) 'Recognition of multiple patterns of DNA sites by Drosophila homeodomain protein Bicoid', *J Biochem,* 125(4), pp. 809-17. doi: 10.1093/oxfordjournals.jbchem.a022353.

Yunger, S., Rosenfeld, L., Garini, Y. and Shav-Tal, Y. (2010) 'Single-allele analysis of transcription kinetics in living mammalian cells', *Nat Methods,* 7(8), pp. 631-3. doi: 10.1038/nmeth.1482.

Zhang, Z. and Gerstein, M. (2003) 'Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements', *J Biol,* 2(2), pp. 11. doi: 10.1186/1475-4924-2-11.

Zhao, C., Dave, V., Yang, F., Scarborough, T. and Ma, J. (2000a) 'Target Selectivity of Bicoid Is Dependent on Nonconsensus Site Recognition and Protein-Protein Interaction', *Mol Cell Biol: Vol. 21*, pp. 8112-23.

Zhao, C., Dave, V., Yang, F., Scarborough, T. and Ma, J. (2000b) 'Target selectivity of bicoid is dependent on nonconsensus site recognition and protein-protein interaction', *Mol Cell Biol,* 20(21), pp. 8112-23. doi: 10.1128/mcb.20.21.8112-8123.2000.

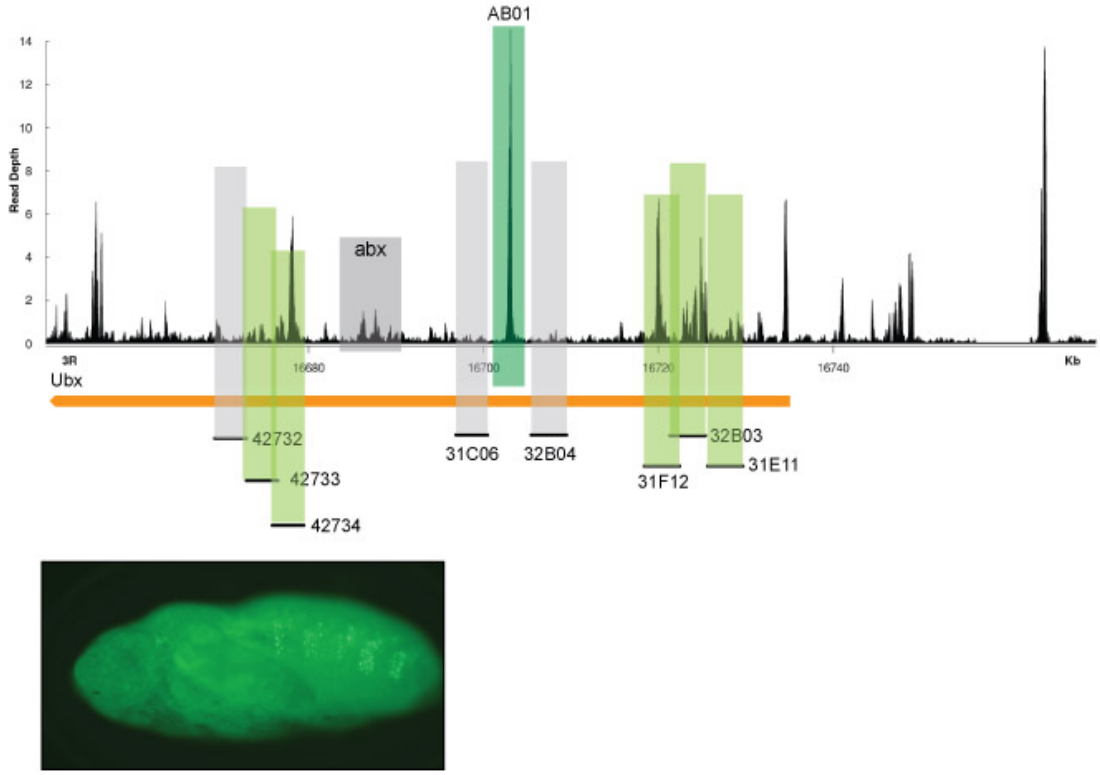
Zhou, J., Barolo, S., Szymanski, P. and Levine, M. (1996) 'The Fab-7 element of the bithorax complex attenuates enhancer-promoter interactions in the Drosophila embryo', *Genes Dev,* 10(24), pp. 3195-201. doi: 10.1101/gad.10.24.3195.

Zhu, Y., Sun, L., Chen, Z., Whitaker, J. W., Wang, T. and Wang, W. (2013) 'Predicting enhancer transcription and activity from chromatin modifications', *Nucleic Acids Res: Vol. 22*, pp. 10032-43.

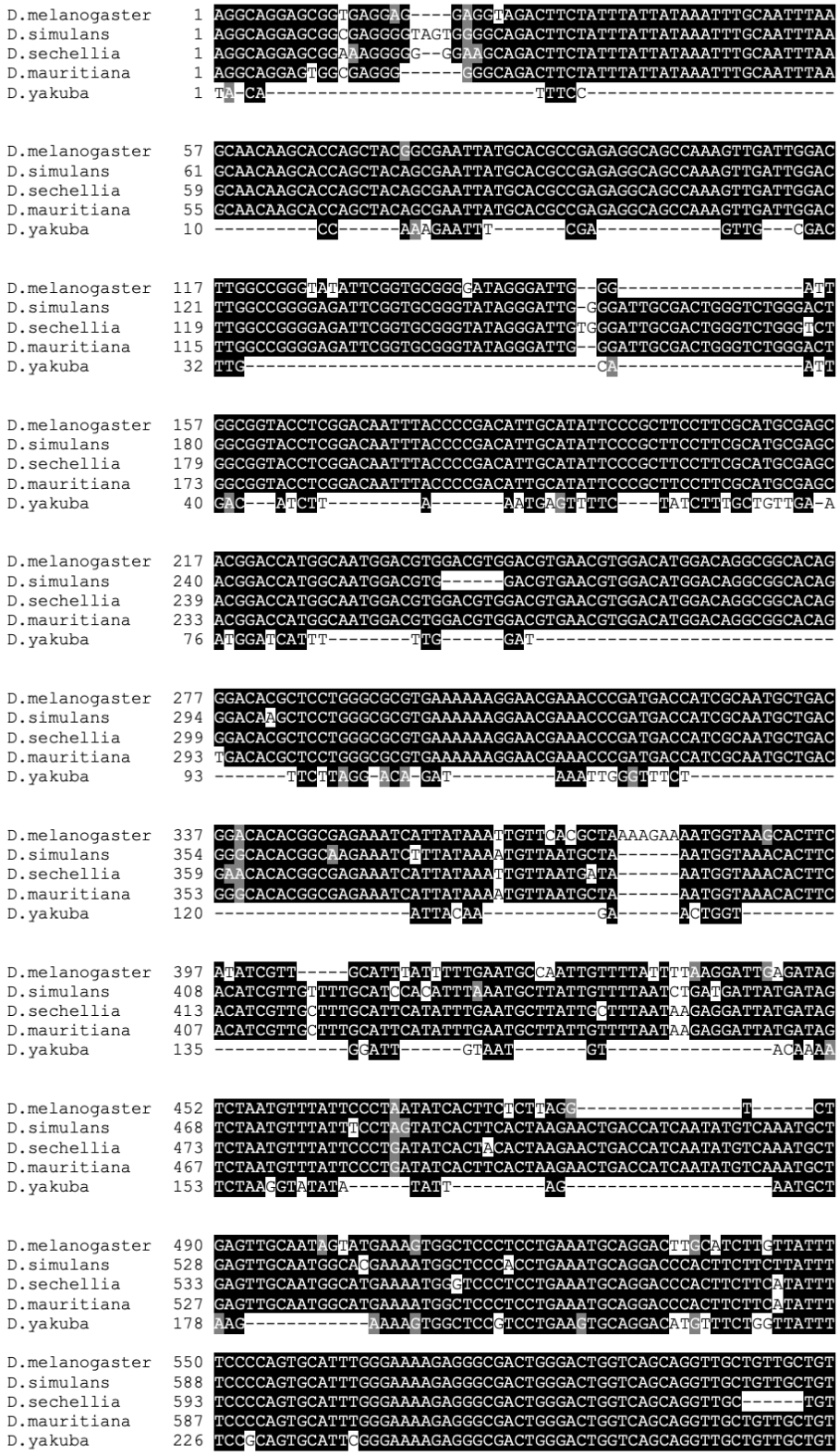
Zinzen, R. P., Girardot, C., Gagneur, J., Braun, M. and Furlong, E. E. (2009) 'Combinatorial binding predicts spatio-temporal cis-regulatory activity', *Nature,* 462(7269), pp. 65-70. doi: 10.1038/nature08531.

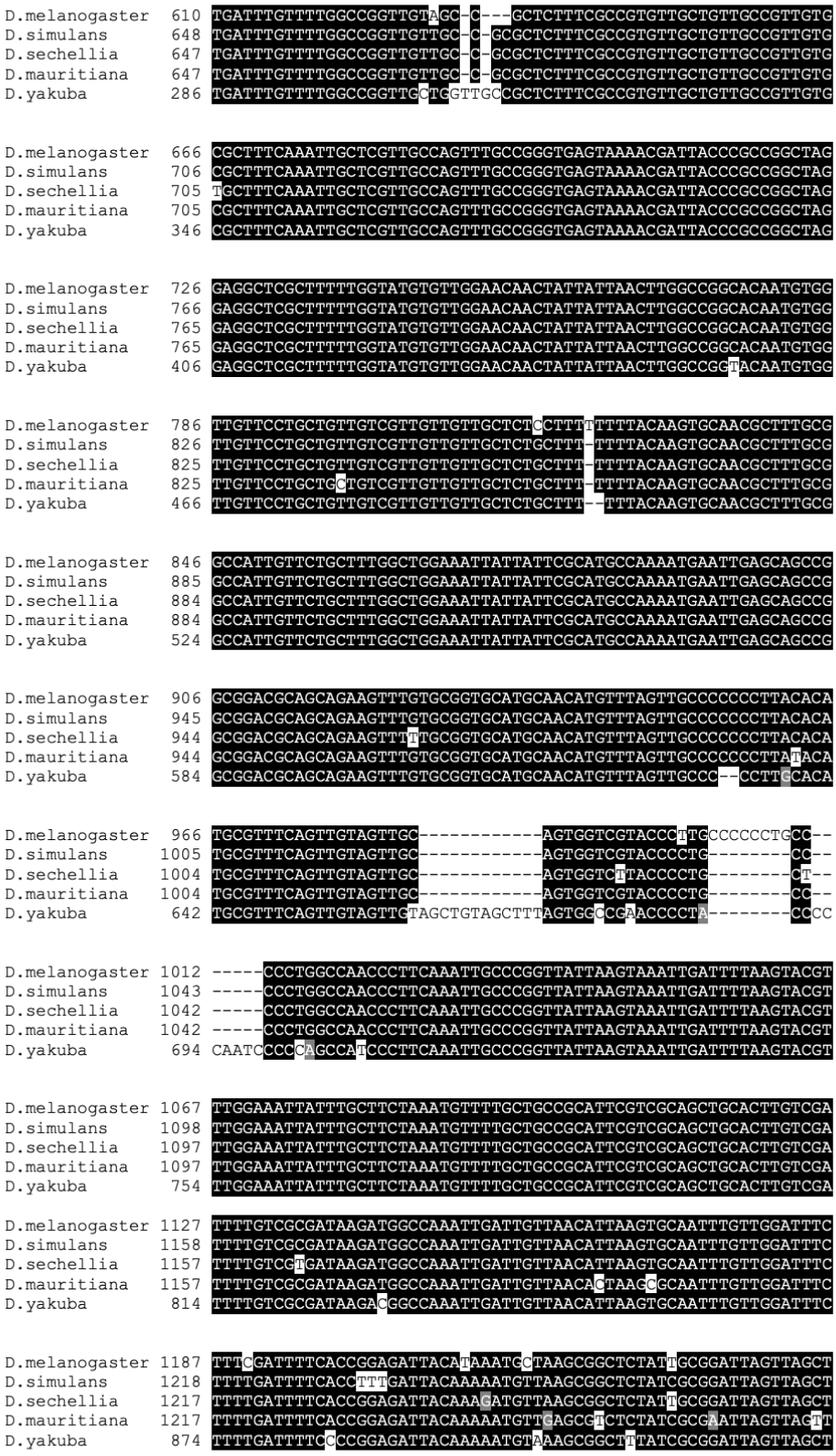
Supplementary materials

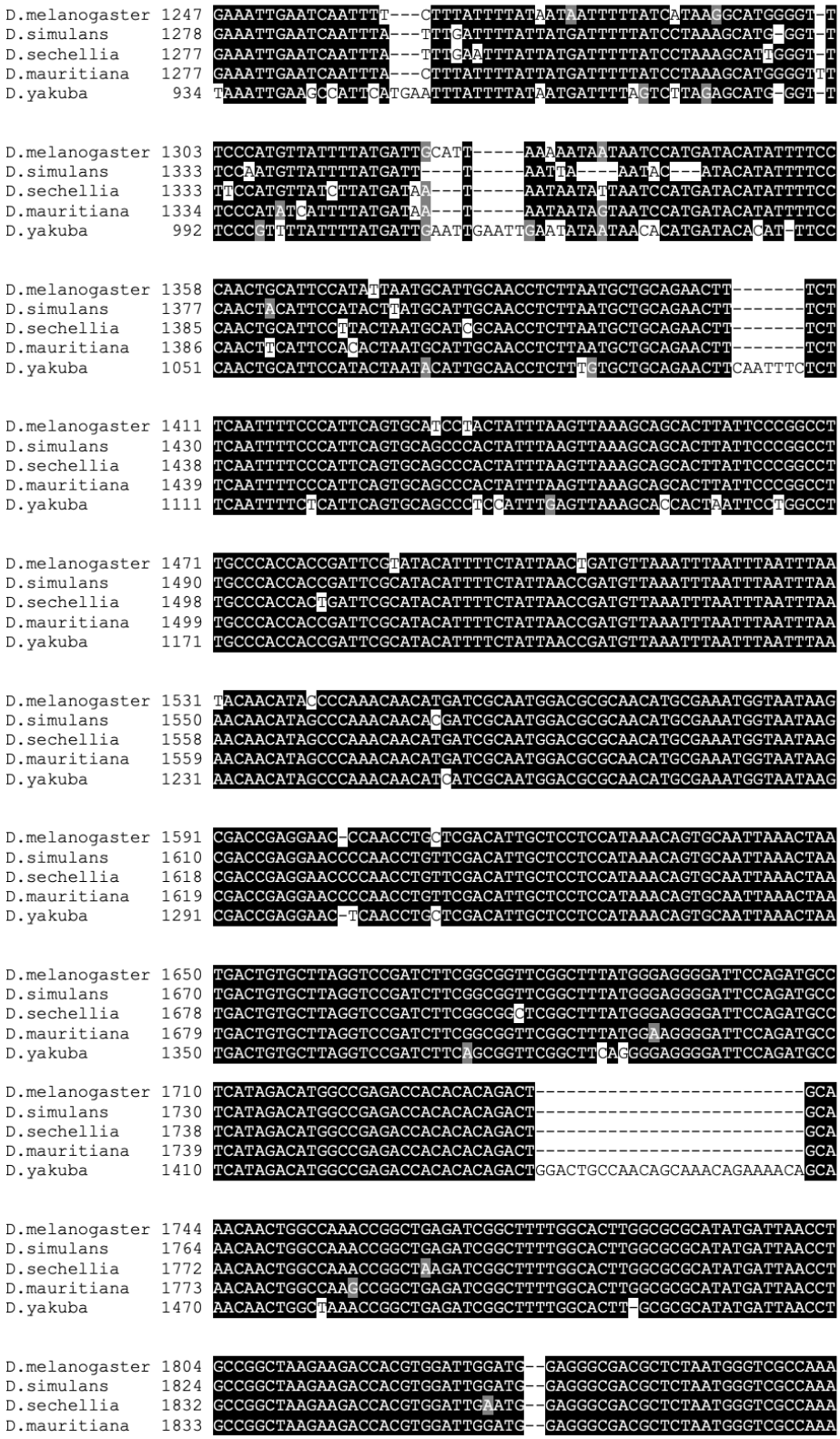
Chapter 2

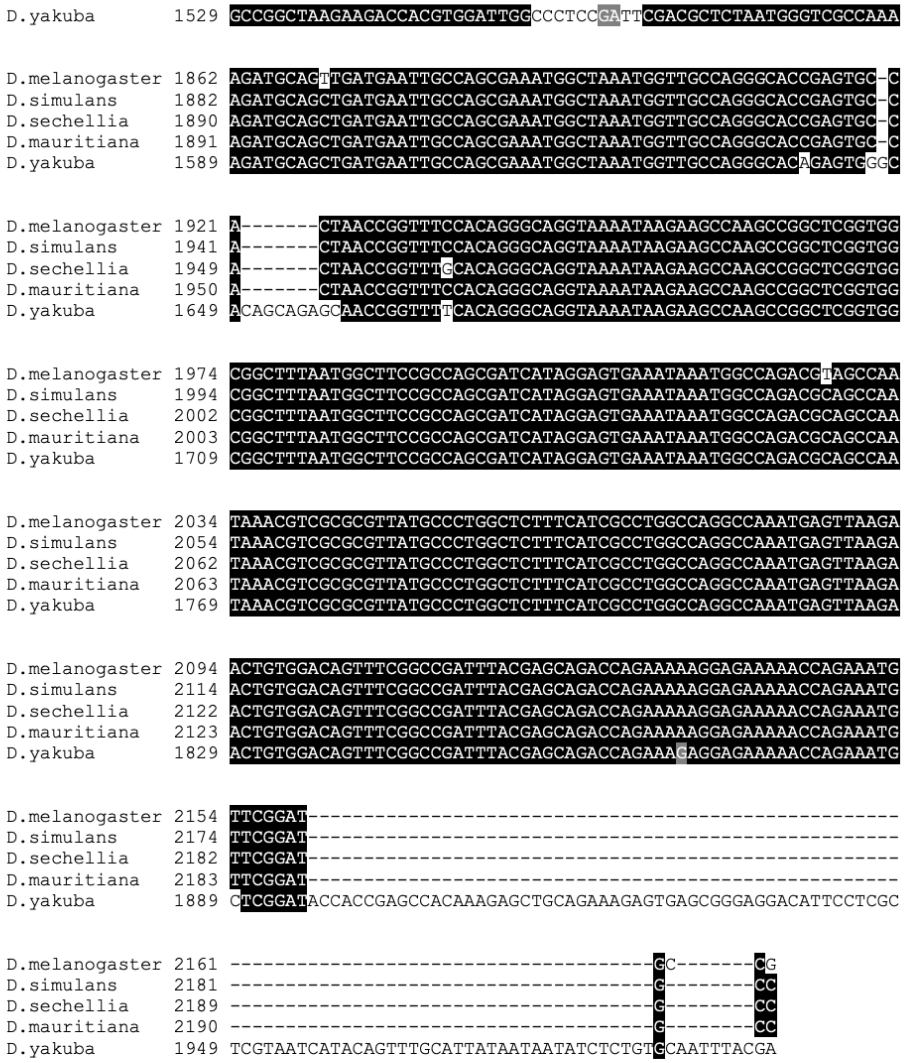


Supplementary Fig. 1. Expression driven by the line AB01. Shown above is the ATAC-seq profile for the Ubx gene with the reporter constructs tested in Chapter 2. Line AB01 is shown in dark green. The expression driven by this line is shown below the ATAC-seq profile. The expression at 24 hAPF is specific to abdominal histoblast cells.


D.melanogaster 1 AGGCAGGAGCGGTGAGGAG----GAGGTAGACTTCTATTTATTATAAATTTGCAATTTAA
D.simulans 1 AGGCAGGAGCGGCGAGGGGTAGTGGGGCAGACTTCTATTTATTATAAATTTGCAATTTAA
D.sechellia 1 AGGCAGGAGCGGAAAGGGG--GGAAGCAGACTTCTATTTATTATAAATTTGCAATTTAA
D.mauritiana 1 AGGCAGGAGTGGCGAGGG------GGGCAGACTTCTATTTATTATAAATTTGCAATTTAA
D.yakuba 1 TA-CA-------------------------TTTCC-----------------------
D.melanogaster 57 GCAACAAGCACCAGCTACGGCGAATTATGCACGCCGAGAGGCAGCCAAAGTTGATTGGAC
D.simulans 61 GCAACAAGCACCAGCTACAGCGAATTATGCACGCCGAGAGGCAGCCAAAGTTGATTGGAC
D.sechellia 59 GCAACAAGCACCAGCTACAGCGAATTATGCACGCCGAGAGGCAGCCAAAGTTGATTGGAC
D.mauritiana 55 GCAACAAGCACCAGCTACAGCGAATTATGCACGCCGAGAGGCAGCCAAAGTTGATTGGAC
D.yakuba 10 ----------CC------AAAGAATTT-------CGA-----------GTTG---CGAC
D.melanogaster 117 TTGGCCGGGTATATTCGGTGCGGGGATAGGGATTG--GG-----------------ATT
D.simulans 121 TTGGCCGGGGAGATTCGGTGCGGGTATAGGGATTG-GGGATTGCGACTGGGTCTGGGACT
D.sechellia 119 TTGGCCGGGGAGATTCGGTGCGGGTATAGGGATTGTGGGATTGCGACTGGGTCTGGGTCT
D.mauritiana 115 TTGGCCGGGGAGATTCGGTGCGGGTATAGGGATTG--GGATTGCGACTGGGTCTGGGACT
D.yakuba 32 TTG-----------------------------------CA-----------------ATT
D.melanogaster 157 GGCGGTACCTCGGACAATTTACCCCGACATTGCATATTCCCGCTTCCTTCGCATGCGAGC
D.simulans 180 GGCGGTACCTCGGACAATTTACCCCGACATTGCATATTCCCGCTTCCTTCGCATGCGAGC
D.sechellia 179 GGCGGTACCTCGGACAATTTACCCCGACATTGCATATTCCCGCTTCCTTCGCATGCGAGC
D.mauritiana 173 GGCGGTACCTCGGACAATTTACCCCGACATTGCATATTCCCGCTTCCTTCGCATGCGAGC
D.yakuba 40 GAC---ATCTT---------A-------AATGAGTTTTC----TATCTTTGCTGTTGA-A
D.melanogaster 217 ACGGACCATGGCAATGGACGTGGACGTGGACGTGAACGTGGACATGGACAGGCGGCACAG
D.simulans 240 ACGGACCATGGCAATGGACGTG------GACGTGAACGTGGACATGGACAGGCGGCACAG
D.sechellia 239 ACGGACCATGGCAATGGACGTGGACGTGGACGTGAACGTGGACATGGACAGGCGGCACAG
D.mauritiana 233 ACGGACCATGGCAATGGACGTGGACGTGGACGTGAACGTGGACATGGACAGGCGGCACAG
D.yakuba 76 ATGGATCATTT--------TTG------GAT----------------------------
D.melanogaster 277 GGACACGCTCCTGGGCGCGTGAAAAAAGGAACGAAACCCGATGACCATCGCAATGCTGAC
D.simulans 294 GGACAAGCTCCTGGGCGCGTGAAAAAAGGAACGAAACCCGATGACCATCGCAATGCTGAC
D.sechellia 299 GGACACGCTCCTGGGCGCGTGAAAAAAGGAACGAAACCCGATGACCATCGCAATGCTGAC
D.mauritiana 293 TGACACGCTCCTGGGCGCGTGAAAAAAGGAACGAAACCCGATGACCATCGCAATGCTGAC
D.yakuba 93 -------TTCTTAGG-ACA-GAT----------AAATTGGGTTTCT--------------
D.melanogaster 337 GGACACACGGCGAGAAATCATTATAAATTGTTCACGCTAAAGAAAATGGTAAGCACTTC
D.simulans 354 GGCACACGGCAAGAAATCTTTATAAAATGTTAATGCTA------AATGGTAAACACTTC
D.sechellia 359 GAACACACGGCGAGAAATCATTATAAATTGTTAATGATA------AATGGTAAACACTTC
D.mauritiana 353 GGGCACACGGCGAGAAATCATTATAAAATGTTAATGCTA------AATGGTAAACACTTC
D.yakuba 120 -------------------ATTACAA----------GA------ACTGGT---------
D.melanogaster 397 ATATCGTT-----GCATTTATTTTTGAATGCCAATTGTTTTATTTTAAGGATTCAGATAG
D.simulans 408 ACATCGTTGTTTTGCATGCACATTTAAATGCTTATTGTTTTAATCTGATGATTATGATAG
D.sechellia 413 ACATCGTTCCTTTGCATTCATATTTGAATGCTTATTGCTTTAATAAGAGGATTATGATAG
D.mauritiana 407 ACATCGTTGCTTTGCATTCATATTTGAATGCTTATTGTTTTAATAAGAGGATTATGATAG
D.yakuba 135 -------------GGATT------GTAAT-------GT----------------ACAAAA
D.melanogaster 452 TCTAATGTTTATTCCCTAATATCACTTCTCTTAGG----------------T------CT
D.simulans 468 TCTAATGTTTATTTCCTAGTATCACTTCACTAAGAACTGACCATCAATATGTCAAATGCT
D.sechellia 473 TCTAATGTTTATTCCCTGATATCACTACACTAAGAACTGACCATCAATATGTCAAATGCT
D.mauritiana 467 TCTAATGTTTATTCCCTGATATCACTTCACTAAGAACTGACCATCAATATGTCAAATGCT
D.yakuba 153 TCTAAGGTATATA------TATT---------AG--------------------AATGCT
D.melanogaster 490 GAGTTGCAATAGTATGAAAGTGGCTCCCTCCTGAAATGCAGGACTTGCATCTTGTTATTT
D.simulans 528 GAGTTGCAATGGCACGAAAATGGCTCCCACCTGAAATGCAGGACCCACTTCTTCTTATTT
D.sechellia 533 GAGTTGCAATGGCATGAAAATGGGTCCCTCCTGAAATGCAGGACCCACTTCTTCATATTT
D.mauritiana 527 GAGTTGCAATGGCATGAAAATGGCTCCCTCCTGAAATGCAGGACCCACTTCTTCATATTT
D.yakuba 178 AAG-------------AAAAGTGGCTCCGTCCTGAAGTGCAGGACATGTTTCTGGTTATTT
D.melanogaster 550 TCCCCAGTGCATTTGGGAAAAGAGGGCGACTGGGACTGGTCAGCAGGTTGCTGTTGCTGT
D.simulans 588 TCCCCAGTGCATTTGGGAAAAGAGGGCGACTGGGACTGGTCAGCAGGTTGCTGTTGCTGT
D.sechellia 593 TCCCCAGTGCATTTGGGAAAAGAGGGCGACTGGGACTGGTCAGCAGGTTGC------TGT
D.mauritiana 587 TCCCCAGTGCATTTGGGAAAAGAGGGCGACTGGGACTGGTCAGCAGGTTGCTGTTGCTGT
D.yakuba 226 TCGCAGTGCATTGGGGAAAAGAGGGCGACTGGGACTGGTCAGCAGGTTGCTGTTGCTGT


D.melanogaster 610 TGATTTGTTTTGGCCGGTTGTAGC-C---GCTCTTTCGCCGTGTTGCTGTTGCCGTTGTG
D.simulans 648 TGATTTGTTTTGGCCGGTTGTTGC-C-GCGCTCTTTCGCCGTGTTGCTGTTGCCGTTGTG
D.sechellia 647 TGATTTGTTTTGGCCGGTTGTTGC-C-GCGCTCTTTCGCCGTGTTGCTGTTGCCGTTGTG
D.mauritiana 647 TGATTTGTTTTGGCCGGTTGTTGC-C-GCGCTCTTTCGCCGTGTTGCTGTTGCCGTTGTG
D.yakuba 286 TGATTTGTTTTGGCCGGTTGCTGGTTGCCGCTCTTTCGCCGTGTTGCTGTTGCCGTTGTG
D.melanogaster 666 CGCTTTCAAATTGCTCGTTGCCAGTTTGCCGGGTGAGTAAAACGATTACCCGCCGGCTAG
D.simulans 706 CGCTTTCAAATTGCTCGTTGCCAGTTTGCCGGGTGAGTAAAACGATTACCCGCCGGCTAG
D.sechellia 705 TGCTTTCAAATTGCTCGTTGCCAGTTTGCCGGGTGAGTAAAACGATTACCCGCCGGCTAG
D.mauritiana 705 CGCTTTCAAATTGCTCGTTGCCAGTTTGCCGGGTGAGTAAAACGATTACCCGCCGGCTAG
D.yakuba 346 CGCTTTCAAATTGCTCGTTGCCAGTTTGCCGGGTGAGTAAAACGATTACCCGCCGGCTAG
D.melanogaster 726 GAGGCTCGCTTTTTGGTATGTGTTGGAACAACTATTATTAACTTGGCCGGCACAATGTGG
D.simulans 766 GAGGCTCGCTTTTTGGTATGTGTTGGAACAACTATTATTAACTTGGCCGGCACAATGTGG
D.sechellia 765 GAGGCTCGCTTTTTGGTATGTGTTGGAACAACTATTATTAACTTGGCCGGCACAATGTGG
D.mauritiana 765 GAGGCTCGCTTTTTGGTATGTGTTGGAACAACTATTATTAACTTGGCCGGCACAATGTGG
D.yakuba 406 GAGGCTCGCTTTTTGGTATGTGTTGGAACAACTATTATTAACTTGGCCGGTACAATGTGG
D.melanogaster 786 TTGTTCCTGCTGTTGTCGTTGTTGTTGCTCTCCTTTTTTTTACAAGTGCAACGCTTTGCG
D.simulans 826 TTGTTCCTGCTGTTGTCGTTGTTGTTGCTCTGCTTT-TTTTACAAGTGCAACGCTTTGCG
D.sechellia 825 TTGTTCCTGCTGTTGTCGTTGTTGTTGCTCTGCTTT-TTTTACAAGTGCAACGCTTTGCG
D.mauritiana 825 TTGTTCCTGCTGCTGTCGTTGTTGTTGCTCTGCTTT-TTTTACAAGTGCAACGCTTTGCG
D.yakuba 466 TTGTTCCTGCTGTTGTCGTTGTTGTTGCTCTGCTTT--TTTACAAGTGCAACGCTTTGCG
D.melanogaster 846 GCCATTGTTCTGCTTTGGCTGGAAATTATTATTCGCATGCCAAAATGAATTGAGCAGCCG
D.simulans 885 GCCATTGTTCTGCTTTGGCTGGAAATTATTATTCGCATGCCAAAATGAATTGAGCAGCCG
D.sechellia 884 GCCATTGTTCTGCTTTGGCTGGAAATTATTATTCGCATGCCAAAATGAATTGAGCAGCCG
D.mauritiana 884 GCCATTGTTCTGCTTTGGCTGGAAATTATTATTCGCATGCCAAAATGAATTGAGCAGCCG
D.yakuba 524 GCCATTGTTCTGCTTTGGCTGGAAATTATTATTCGCATGCCAAAATGAATTGAGCAGCCG
D.melanogaster 906 GCGGACGCAGCAGAAGTTTGTGCGGTGCATGCAACATGTTTAGTTGCCCCCCCTTACACA
D.simulans 945 GCGGACGCAGCAGAAGTTTGTGCGGTGCATGCAACATGTTTAGTTGCCCCCCCTTACACA
D.sechellia 944 GCGGACGCAGCAGAAGTTTTTGCGGTGCATGCAACATGTTTAGTTGCCCCCCCTTACACA
D.mauritiana 944 GCGGACGCAGCAGAAGTTTGTGCGGTGCATGCAACATGTTTAGTTGCCCCCCCTTATACA
D.yakuba 584 GCGGACGCAGCAGAAGTTTGTGCGGTGCATGCAACATGTTTAGTTGCCC--CCTTGCACA
D.melanogaster 966 TGCGTTTCAGTTGTAGTTGC----------AGTGGTCGTACCCTTGCCCCCCTGCC--
D.simulans 1005 TGCGTTTCAGTTGTAGTTGC----------AGTGGTCGTACCCCTG--------CC--
D.sechellia 1004 TGCGTTTCAGTTGTAGTTGC----------AGTGGTCTTACCCCTG--------CT--
D.mauritiana 1004 TGCGTTTCAGTTGTAGTTGC----------AGTGGTCGTACCCCTG--------CC--
D.yakuba 642 TGCGTTTCAGTTGTAGTTGTAGCTGTAGCTTTAGTGGCCGAACCCCTA--------CCCC
D.melanogaster 1012 -----CCCTGGCCAACCCTTCAAATTGCCCGGTTATTAAGTAAATTGATTTTAAGTACGT
D.simulans 1043 -----CCCTGGCCAACCCTTCAAATTGCCCGGTTATTAAGTAAATTGATTTTAAGTACGT
D.sechellia 1042 -----CCCTGGCCAACCCTTCAAATTGCCCGGTTATTAAGTAAATTGATTTTAAGTACGT
D.mauritiana 1042 -----CCCTGGCCAACCCTTCAAATTGCCCGGTTATTAAGTAAATTGATTTTAAGTACGT
D.yakuba 694 CAATCCCCCAGCCATCCCTTCAAATTGCCCGGTTATTAAGTAAATTGATTTTAAGTACGT
D.melanogaster 1067 TTGGAAATTATTTGCTTCTAAATGTTTTGCTGCCGCATTCGTCGCAGCTGCACTTGTCGA
D.simulans 1098 TTGGAAATTATTTGCTTCTAAATGTTTTGCTGCCGCATTCGTCGCAGCTGCACTTGTCGA
D.sechellia 1097 TTGGAAATTATTTGCTTCTAAATGTTTTGCTGCCGCATTCGTCGCAGCTGCACTTGTCGA
D.mauritiana 1097 TTGGAAATTATTTGCTTCTAAATGTTTTGCTGCCGCATTCGTCGCAGCTGCACTTGTCGA
D.yakuba 754 TTGGAAATTATTTGCTTCTAAATGTTTTGCTGCCGCATTCGTCGCAGCTGCACTTGTCGA
D.melanogaster 1127 TTTTGTCGCGATAAGATGGCCAAATTGATTGTTAACATTAAGTGCAATTTGTTGGATTTC
D.simulans 1158 TTTTGTCGCGATAAGATGGCCAAATTGATTGTTAACATTAAGTGCAATTTGTTGGATTTC
D.sechellia 1157 TTTTGTCGTGATAAGATGGCCAAATTGATTGTTAACATTAAGTGCAATTTGTTGGATTTC
D.mauritiana 1157 TTTTGTCGCGATAAGATGGCCAAATTGATTGTTAACACTAAGCGCAATTTGTTGGATTTC
D.yakuba 814 TTTTGTCGCGATAAGACGGCCAAATTGATTGTTAACATTAAGTGCAATTTGTTGGATTTC
D.melanogaster 1187 TTTCGATTTTCACCGGAGATTACATAAATGCTAAGCGGCTCTATTGCGGATTAGTTAGCT
D.simulans 1218 TTTTGATTTTCACCTTTGATTACAAAAATGTTAAGCGGCTCTATCGCGGATTAGTTAGCT
D.sechellia 1217 TTTTGATTTTCACCGGAGATTACAAAGATGTTAAGCGGCTCTATTGCGGATTAGTTAGCT
D.mauritiana 1217 TTTTGATTTTCACCGGAGATTACAAAAATGTTGAGCGTCTCTATCGCGAATTAGTTAGTT
D.yakuba 874 TTTTGATTTTCCCCGGAGATTACAAAAATGTAAAGCGGCTTTATCGCGGATTAGTTAGCT


D.melanogaster 1247 GAAATTGAATCAATTTT---CTTTATTTTATAATAATTTTTATCATAAGGCATGGGGT-T
D.simulans 1278 GAAATTGAATCAATTTA---TTTGATTTTATTATGATTTTTATCCTAAAGCATG-GGT-T
D.sechellia 1277 GAAATTGAATCAATTTA---TTTGAATTTATTATGATTTTTATCCTAAAGCATTGGGT-T
D.mauritiana 1277 GAAATTGAATCAATTTA---CTTTATTTTATTATGATTTTTATCCTAAAGCATGGGGTTT
D.yakuba 934 TAAATTGAACCCATTCATGAATTTATTTTATAATGATTTTAGTCTTAGAGCATG-GGT-T
D.melanogaster 1303 TCCCATGTTATTTTATGATTCCATT-----AAAAATAATAATCCATGATACATATTTTCC
D.simulans 1333 TCCAATGTTATTTTATGATT-----T-----AATTA----AATAC---ATACATATTTTCC
D.sechellia 1333 TTCCATGTTATCTTATGATAA----T-----AATAATATTAATCCATGATACATATTTTCC
D.mauritiana 1334 TCCCATATCATTTTATGATAA----T-----AATAATAGTAATCCATGATACATATTTTCC
D.yakuba 992 TCCCGTTTTATTTTATGATTCAATTGAATTGAATATAATAACACATGATACACAT-TTCC
D.melanogaster 1358 CAACTGCATTCCATATTAATGCATTGCAACCTCTTAATGCTGCAGAACTT-------TCT
D.simulans 1377 CAACTACATTCCATACTTATGCATTGCAACCTCTTAATGCTGCAGAACTT-------TCT
D.sechellia 1385 CAACTGCATTCCTTACTAATGCATCGCAACCTCTTAATGCTGCAGAACTT-------TCT
D.mauritiana 1386 CAACTTCATTCCACACTAATGCATTGCAACCTCTTAATGCTGCAGAACTT-------TCT
D.yakuba 1051 CAACTGCATTCCATACTAATACATTGCAACCTCTTTGTGCTGCAGAACTTCAATTTCTCT
D.melanogaster 1411 TCAATTTTCCCATTCAGTGCATCCTACTATTTAAGTTAAAGCAGCACTTATTCCCGGCCT
D.simulans 1430 TCAATTTTCCCATTCAGTGCAGCCCACTATTTAAGTTAAAGCAGCACTTATTCCCGGCCT
D.sechellia 1438 TCAATTTTCCCATTCAGTGCAGCCCACTATTTAAGTTAAAGCAGCACTTATTCCCGGCCT
D.mauritiana 1439 TCAATTTTCCCATTCAGTGCAGCCCACTATTTAAGTTAAAGCAGCACTTATTCCCGGCCT
D.yakuba 1111 TCAATTTTCTCATTCAGTGCAGCCCTGCATTTGAGTTAAAGCACCACTAATTCCTGGCCT
D.melanogaster 1471 TGCCCACCACCGATTCGTATACATTTTCTATTAACTGATGTTAAATTTAATTTAATTTAA
D.simulans 1490 TGCCCACCACCGATTCGCATACATTTTCTATTAACCGATGTTAAATTTAATTTAATTTAA
D.sechellia 1498 TGCCCACCACTGATTCGCATACATTTTCTATTAACCGATGTTAAATTTAATTTAATTTAA
D.mauritiana 1499 TGCCCACCACCGATTCGCATACATTTTCTATTAACCGATGTTAAATTTAATTTAATTTAA
D.yakuba 1171 TGCCCACCACCGATTCGCATACATTTTCTATTAACCGATGTTAAATTTAATTTAATTTAA
D.melanogaster 1531 TACAACATACCCCAAACAACATGATCGCAATGGACGCGCAACATGCGAAATGGTAATAAG
D.simulans 1550 AACAACATAGCCCAAACAACACGATCGCAATGGACGCGCAACATGCGAAATGGTAATAAG
D.sechellia 1558 AACAACATAGCCCAAACAACATGATCGCAATGGACGCGCAACATGCGAAATGGTAATAAG
D.mauritiana 1559 AACAACATAGCCCAAACAACATGATCGCAATGGACGCGCAACATGCGAAATGGTAATAAG
D.yakuba 1231 AACAACATAGCCCAAACAACATCATCGCAATGGACGCGCAACATGCGAAATGGTAATAAG
D.melanogaster 1591 CGACCGAGGAAC-CCAACCTGCTCGACATTGCTCCTCCATAAACAGTGCAATTAAACTAA
D.simulans 1610 CGACCGAGGAACCCCAACCTGTTCGACATTGCTCCTCCATAAACAGTGCAATTAAACTAA
D.sechellia 1618 CGACCGAGGAACCCCAACCTGTTCGACATTGCTCCTCCATAAACAGTGCAATTAAACTAA
D.mauritiana 1619 CGACCGAGGAACCCCAACCTGTTCGACATTGCTCCTCCATAAACAGTGCAATTAAACTAA
D.yakuba 1291 CGACCGAGGAAC-TCAACCTGCTCGACATTGCTCCTCCATAAACAGTGCAATTAAACTAA
D.melanogaster 1650 TGACTGTGCTTAGGTCCGATCTTCGGCGGTTCGGCTTTATGGGAGGGGATTCCAGATGCC
D.simulans 1670 TGACTGTGCTTAGGTCCGATCTTCGGCGGTTCGGCTTTATGGGAGGGGATTCCAGATGCC
D.sechellia 1678 TGACTGTGCTTAGGTCCGATCTTCGGCGGCTCGGCTTTATGGGAGGGGATTCCAGATGCC
D.mauritiana 1679 TGACTGTGCTTAGGTCCGATCTTCGGCGGTTCGGCTTTATGGAAGGGGATTCCAGATGCC
D.yakuba 1350 TGACTGTGCTTAGGTCCGATCTTCAGCGGTTCGGCTTCAGGGGAGGGGATTCCAGATGCC
D.melanogaster 1710 TCATAGACATGGCCGAGACCACACACAGACT--------------------------GCA
D.simulans 1730 TCATAGACATGGCCGAGACCACACACAGACT--------------------------GCA
D.sechellia 1738 TCATAGACATGGCCGAGACCACACACAGACT--------------------------GCA
D.mauritiana 1739 TCATAGACATGGCCGAGACCACACACAGACT--------------------------GCA
D.yakuba 1410 TCATAGACATGGCCGAGACCACACACAGACTGGACTGCCAACAGCAAACAGAAAACAGCA
D.melanogaster 1744 AACAACTGGCCAAACCGGCTGAGATCGGCTTTTGGCACTTGGCGCGCATATGATTAACCT
D.simulans 1764 AACAACTGGCCAAACCGGCTGAGATCGGCTTTTGGCACTTGGCGCGCATATGATTAACCT
D.sechellia 1772 AACAACTGGCCAAACCGGCTAAGATCGGCTTTTGGCACTTGGCGCGCATATGATTAACCT
D.mauritiana 1773 AACAACTGGCCAAGCCGGCTGAGATCGGCTTTTGGCACTTGGCGCGCATATGATTAACCT
D.yakuba 1470 AACAACTGGGTAAACCGGCTGAGATCGGCTTTTGGCACTT-GCGCGCATATGATTAACCT
D.melanogaster 1804 GCCGGCTAAGAAGACCACGTGGATTGGATG--GAGGGCGACGCTCTAATGGGTCGCCAAA
D.simulans 1824 GCCGGCTAAGAAGACCACGTGGATTGGATG--GAGGGCGACGCTCTAATGGGTCGCCAAA
D.sechellia 1832 GCCGGCTAAGAAGACCACGTGGATTGAATG--GAGGGCGACGCTCTAATGGGTCGCCAAA
D.mauritiana 1833 GCCGGCTAAGAAGACCACGTGGATTGGATG--GAGGGCGACGCTCTAATGGGTCGCCAAA



Supplementary Fig. 2. Multiple species alignment of *Ubx* VT42733 for *D. melanogaste*r, *D. simulans*, *D. sechellia, D. mauritiana* and *D. yakuba*. This alignment was submitted to the MEME suite software to find motifs that occur in this enhancer.

Chapter 3

Supplementary Table. 1. The list of genes produced by TargetScan that are predicted to be targeted by miR-92a. Both the number of conserved and poorly conserved sites are shown

				Conserved sites				Poorly conserved sites			
Target gene	**Transcript**	**Symbol**	**ID**	**total**	**8mer**	**7mer-m8**	**7mer-1A**	**total**	**8mer**	**7mer-m8**	**7mer-1A**
FBgn0003382	FBtr0088120	sha	CG13209	**5**	2	3	0	**0**	0	0	0
FBgn0019968	FBtr0087356	Khc-73	CG8183	**3**	1	0	2	**2**	0	2	0
FBgn0004396	FBtr0075557	CrebA	CG7450	**4**	0	4	0	**0**	0	0	0
FBgn0031258	FBtr0078085	CG4297	CG4297	**2**	1	1	0	**1**	0	1	0
FBgn0039530	FBtr0085190	Tusp	CG5586	**2**	2	0	0	**0**	0	0	0
FBgn0038073	FBtr0082650	CG14395	CG14395	**2**	2	0	0	**0**	0	0	0
FBgn0038815	FBtr0083913	CG5466	CG5466	**2**	1	1	0	**1**	0	0	1
FBgn0033867	FBtr0087629	Cpr50Ca	CG13338	**1**	0	1	0	**2**	1	0	1
FBgn0038439	FBtr0083306	Cad89D	CG14900	**2**	1	1	0	**0**	0	0	0
FBgn0011656	FBtr0088447	Mef2	CG1429	**2**	1	1	0	**0**	0	0	0
FBgn0035283	FBtr0072832	CG12024	CG12024	**2**	1	1	0	**0**	0	0	0
FBgn0010350	FBtr0076688	CdsA	CG7962	**1**	0	1	0	**1**	1	0	0

FBgn0032001	FBtr0089935	CG8360	CG8360	**1**	1	0	0	**1**	0	1	0
FBgn0004914	FBtr0079747	Hnf4	CG9310	**2**	0	1	1	**2**	0	0	2
FBgn0015772	FBtr0081189	Nak	CG10637	**2**	0	1	1	**1**	0	1	0
FBgn0032023	FBtr0079604	CG14274	CG14274	**1**	0	1	0	**2**	0	1	1
FBgn0020309	FBtr0080260	crol	CG14938	**2**	1	0	1	**0**	0	0	0
FBgn0004370	FBtr0073524	Ptp10D	CG1817	**2**	1	0	1	**0**	0	0	0
FBgn0029835	FBtr0070869	CG5921	CG5921	**2**	1	0	1	**0**	0	0	0
FBgn0043362	FBtr0079144	bchs	CG14001	**1**	0	0	1	**1**	1	0	0
FBgn0034906	FBtr0072243	CG13561	CG13561	**1**	1	0	0	**1**	0	0	1
FBgn0030581	FBtr0073990	CG14408	CG14408	**1**	1	0	0	**1**	0	0	1
FBgn0039810	FBtr0085752	CG15549	CG15549	**1**	1	0	0	**1**	0	0	1
FBgn0039431	FBtr0085023	CG6490	CG6490	**1**	1	0	0	**1**	0	0	1
FBgn0086372	FBtr0081572	lap	CG2520	**3**	0	1	2	**0**	0	0	0
FBgn0023441	FBtr0087343	fus	CG8205	**1**	0	1	0	**2**	0	0	2
FBgn0030532	FBtr0073829	CG11071	CG11071	**1**	0	1	0	**2**	0	0	2
FBgn0039500	FBtr0085181	CG5984	CG5984	**1**	0	1	0	**2**	0	0	2
FBgn0035229	FBtr0072737	CG7852	CG7852	**1**	0	1	0	**2**	0	0	2
FBgn0013948	FBtr0084169	Eip93F	CG18389	**2**	0	2	0	**0**	0	0	0
FBgn0031457	FBtr0077736	CG3077	CG3077	**2**	0	2	0	**0**	0	0	0
FBgn0003870	FBtr0085826	ttk	CG1856	**1**	0	1	0	**1**	0	1	0

FBgn0031966	FBtr0079555	CG14532	CG14532	**1**	0	1	0	**1**	0	1	0
FBgn0054000	FBtr0100055	CG34000	CG34000	**1**	0	1	0	**1**	0	1	0
FBgn0034166	FBtr0087046	CG6472	CG6472	**1**	0	1	0	**1**	0	1	0
FBgn0030668	FBtr0074053	CG8128	CG8128	**1**	0	1	0	**1**	0	1	0
FBgn0000011	FBtr0080203	ab	CG4807	**1**	1	0	0	**0**	0	0	0
FBgn0015567	FBtr0089488	alpha-Adaptin	CG4260	**1**	1	0	0	**0**	0	0	0
FBgn0035625	FBtr0077089	Blimp-1	CG5249	**1**	1	0	0	**0**	0	0	0
FBgn0000157	FBtr0072379	Dll	CG3629	**1**	1	0	0	**0**	0	0	0
FBgn0000577	FBtr0088095	en	CG9015	**1**	1	0	0	**0**	0	0	0
FBgn0028642	FBtr0089038	esn	CG12833	**1**	1	0	0	**0**	0	0	0
FBgn0024555	FBtr0082817	flfl	CG9351	**1**	1	0	0	**0**	0	0	0
FBgn0005633	FBtr0074910	fln	CG7445	**1**	1	0	0	**0**	0	0	0
FBgn0004652	FBtr0083640	fru	CG14307	**1**	1	0	0	**0**	0	0	0
FBgn0004656	FBtr0071118	fs(1)h	CG2252	**1**	1	0	0	**0**	0	0	0
FBgn0030930	FBtr0074625	GalNAc-T2	CG6394	**1**	1	0	0	**0**	0	0	0
FBgn0004380	FBtr0077157	Klp64D	CG10642	**1**	1	0	0	**0**	0	0	0
FBgn0001319	FBtr0112809	kn	CG10197	**1**	1	0	0	**0**	0	0	0
FBgn0039688	FBtr0085460	Kul	CG1964	**1**	1	0	0	**0**	0	0	0
FBgn0037416	FBtr0078599	Osi9	CG15592	**1**	1	0	0	**0**	0	0	0
FBgn0038418	FBtr0083282	pad	CG10309	**1**	1	0	0	**0**	0	0	0

FBgn0017549	FBtr0087254	Ric	CG8418	**1**	1	0	0	**0**	0	0	0
FBgn0004579	FBtr0089913	salm	CG6464	**1**	1	0	0	**0**	0	0	0
FBgn0086475	FBtr0075231	sec3	CG3885	**1**	1	0	0	**0**	0	0	0
FBgn0051632	FBtr0079341	sens-2	CG31632	**1**	1	0	0	**0**	0	0	0
FBgn0040280	FBtr0079004	SP555	CG14041	**1**	1	0	0	**0**	0	0	0
FBgn0003733	FBtr0088938	tor	CG1389	**1**	1	0	0	**0**	0	0	0
FBgn0004397	FBtr0070420	Vinc	CG3299	**1**	1	0	0	**0**	0	0	0
FBgn0036360	FBtr0075855	CG10713	CG10713	**1**	1	0	0	**0**	0	0	0
FBgn0034420	FBtr0100282	CG10737	CG10737	**1**	1	0	0	**0**	0	0	0
FBgn0030408	FBtr0073656	CG11085	CG11085	**1**	1	0	0	**0**	0	0	0
FBgn0033996	FBtr0087363	CG11807	CG11807	**1**	1	0	0	**0**	0	0	0
FBgn0039831	FBtr0085758	CG12054	CG12054	**1**	1	0	0	**0**	0	0	0
FBgn0032186	FBtr0079965	CG13136	CG13136	**1**	1	0	0	**0**	0	0	0
FBgn0039288	FBtr0084804	CG13653	CG13653	**1**	1	0	0	**0**	0	0	0
FBgn0035190	FBtr0072632	CG13913	CG13913	**1**	1	0	0	**0**	0	0	0
FBgn0038438	FBtr0083307	CG14899	CG14899	**1**	1	0	0	**0**	0	0	0
FBgn0035451	FBtr0073130	CG14973	CG14973	**1**	1	0	0	**0**	0	0	0
FBgn0029692	FBtr0070623	CG15376	CG15376	**1**	1	0	0	**0**	0	0	0
FBgn0029748	FBtr0070743	CG15464	CG15464	**1**	1	0	0	**0**	0	0	0
FBgn0035144	FBtr0072563	CG17181	CG17181	**1**	1	0	0	**0**	0	0	0

FBgn0050463	FBtr0087112	CG30463	CG30463	**1**	1	0	0	**0**	0	0	0
FBgn0051122	FBtr0083653	CG31122	CG31122	**1**	1	0	0	**0**	0	0	0
FBgn0051224	FBtr0083691	CG31224	CG31224	**1**	1	0	0	**0**	0	0	0
FBgn0051637	FBtr0079287	CG31637	CG31637	**1**	1	0	0	**0**	0	0	0
FBgn0031375	FBtr0077870	CG31670	CG31670	**1**	1	0	0	**0**	0	0	0
FBgn0051949	FBtr0077757	CG31949	CG31949	**1**	1	0	0	**0**	0	0	0
FBgn0052416	FBtr0077110	CG32416	CG32416	**1**	1	0	0	**0**	0	0	0
FBgn0085227	FBtr0112391	CG34198	CG34198	**1**	1	0	0	**0**	0	0	0
FBgn0085250	FBtr0112414	CG34221	CG34221	**1**	1	0	0	**0**	0	0	0
FBgn0085383	FBtr0112564	CG34354	CG34354	**1**	1	0	0	**0**	0	0	0
FBgn0037017	FBtr0078280	CG4074	CG4074	**1**	1	0	0	**0**	0	0	0
FBgn0038787	FBtr0083899	CG4360	CG4360	**1**	1	0	0	**0**	0	0	0
FBgn0038740	FBtr0083859	CG4562	CG4562	**1**	1	0	0	**0**	0	0	0
FBgn0038744	FBtr0083820	CG4733	CG4733	**1**	1	0	0	**0**	0	0	0
FBgn0036555	FBtr0075518	CG6017	CG6017	**1**	1	0	0	**0**	0	0	0
FBgn0032513	FBtr0080468	CG6565	CG6565	**1**	1	0	0	**0**	0	0	0
FBgn0036741	FBtr0075182	CG7510	CG7510	**1**	1	0	0	**0**	0	0	0
FBgn0034143	FBtr0087119	CG8303	CG8303	**1**	1	0	0	**0**	0	0	0
FBgn0033903	FBtr0087521	CG8323	CG8323	**1**	1	0	0	**0**	0	0	0
FBgn0037744	FBtr0082131	CG8417	CG8417	**1**	1	0	0	**0**	0	0	0

FBgn0034808	FBtr0071988	CG9896	CG9896	**1**	1	0	0	**0**	0	0	0
FBgn0010229	FBtr0081479	Hr39	CG8676	**2**	0	1	1	**0**	0	0	0
FBgn0000567	FBtr0075202	Eip74EF	CG32180	**1**	0	1	0	**1**	0	0	1
FBgn0042645	FBtr0084226	epsin-like	CG31170	**1**	0	0	1	**1**	0	1	0
FBgn0030252	FBtr0073388	Myo10A	CG2174	**1**	0	1	0	**1**	0	0	1
FBgn0031737	FBtr0079164	obst-E	CG11142	**1**	0	0	1	**1**	0	1	0
FBgn0037422	FBtr0078605	Osi13	CG15595	**1**	0	0	1	**1**	0	1	0
FBgn0003162	FBtr0071578	Pu	CG9441	**1**	0	0	1	**1**	0	1	0
FBgn0039257	FBtr0084754	tnc	CG13648	**1**	0	1	0	**1**	0	0	1
FBgn0037848	FBtr0082288	Tsp86D	CG4591	**1**	0	0	1	**1**	0	1	0
FBgn0030249	FBtr0112978	CG11203	CG11203	**1**	0	1	0	**1**	0	0	1
FBgn0036193	FBtr0076111	CG14135	CG14135	**1**	0	0	1	**1**	0	1	0
FBgn0033448	FBtr0088486	CG1623	CG1623	**1**	0	1	0	**1**	0	0	1
FBgn0051191	FBtr0083984	CG31191	CG31191	**1**	0	0	1	**1**	0	1	0
FBgn0052774	FBtr0070662	CG32774	CG32774	**1**	0	1	0	**1**	0	0	1
FBgn0085369	FBtr0112543	CG34340	CG34340	**1**	0	1	0	**1**	0	0	1
FBgn0034775	FBtr0071916	CG3536	CG3536	**1**	0	0	1	**1**	0	1	0
FBgn0039730	FBtr0085608	CG7903	CG7903	**1**	0	0	1	**1**	0	1	0
FBgn0033677	FBtr0087975	CG8321	CG8321	**1**	0	1	0	**1**	0	0	1
FBgn0000562	FBtr0114586	egl	CG4051	**2**	0	0	2	**0**	0	0	0

FBgn0021895	FBtr0072234	ytr	CG18426	**2**	0	0	2	**0**	0	0	0
FBgn0033609	FBtr0088179	CG13213	CG13213	**2**	0	0	2	**0**	0	0	0
FBgn0031698	FBtr0113022	CG14023	CG14023	**2**	0	0	2	**0**	0	0	0
FBgn0023407	FBtr0080513	B4	CG9239	**1**	0	0	1	**1**	0	0	1
FBgn0039528	FBtr0085201	dsd	CG5634	**1**	0	0	1	**1**	0	0	1
FBgn0004873	FBtr0086546	hts	CG9325	**1**	0	0	1	**1**	0	0	1
FBgn0023184	FBtr0100213	Nop60B	CG3333	**1**	0	0	1	**1**	0	0	1
FBgn0004449	FBtr0078509	Ten-m	CG5723	**1**	0	0	1	**1**	0	0	1
FBgn0051206	FBtr0083875	CG31206	CG31206	**1**	0	0	1	**1**	0	0	1
FBgn0031981	FBtr0079541	CG7466	CG7466	**1**	0	0	1	**1**	0	0	1
FBgn0032021	FBtr0079609	CG7781	CG7781	**1**	0	0	1	**1**	0	0	1
FBgn0000054	FBtr0086111	Adf1	CG15845	**1**	0	1	0	**0**	0	0	0
FBgn0038535	FBtr0089412	alt	CG18212	**1**	0	1	0	**0**	0	0	0
FBgn0015591	FBtr0084744	Ast	CG13633	**1**	0	1	0	**0**	0	0	0
FBgn0000244	FBtr0082119	by	CG9379	**1**	0	1	0	**0**	0	0	0
FBgn0029846	FBtr0089393	Ca-alpha1T	CG15899	**1**	0	1	0	**0**	0	0	0
FBgn0051536	FBtr0110958	Cdep	CG31536	**1**	0	1	0	**0**	0	0	0
FBgn0032409	FBtr0080330	Ced-12	CG5336	**1**	0	1	0	**0**	0	0	0
FBgn0015623	FBtr0079250	Cpr	CG11567	**1**	0	1	0	**0**	0	0	0
FBgn0005677	FBtr0080875	dac	CG4952	**1**	0	1	0	**0**	0	0	0

FBgn0016794	FBtr0072926	dos	CG1044	**1**	0	1	0	**0**	0	0	0
FBgn0027835	FBtr0086678	Dp1	CG5170	**1**	0	1	0	**0**	0	0	0
FBgn0000524	FBtr0070914	dx	CG3929	**1**	0	1	0	**0**	0	0	0
FBgn0039411	FBtr0113293	dys	CG32474	**1**	0	1	0	**0**	0	0	0
FBgn0000546	FBtr0086008	EcR	CG1765	**1**	0	1	0	**0**	0	0	0
FBgn0005640	FBtr0073131	Eip63E	CG10579	**1**	0	1	0	**0**	0	0	0
FBgn0034433	FBtr0086558	endoB	CG9834	**1**	0	1	0	**0**	0	0	0
FBgn0028736	FBtr0073216	Ero1L	CG1333	**1**	0	1	0	**0**	0	0	0
FBgn0000630	FBtr0110991	f	CG5424	**1**	0	1	0	**0**	0	0	0
FBgn0024238	FBtr0074441	Fim	CG8649	**1**	0	1	0	**0**	0	0	0
FBgn0086698	FBtr0077911	frtz	CG17657	**1**	0	1	0	**0**	0	0	0
FBgn0039580	FBtr0085297	Gfat2	CG1345	**1**	0	1	0	**0**	0	0	0
FBgn0001138	FBtr0081808	grn	CG9656	**1**	0	1	0	**0**	0	0	0
FBgn0017397	FBtr0084177	how	CG10293	**1**	0	1	0	**0**	0	0	0
FBgn0038736	FBtr0113248	ire-1	CG4583	**1**	0	1	0	**0**	0	0	0
FBgn0032129	FBtr0079861	jp	CG4405	**1**	0	1	0	**0**	0	0	0
FBgn0001297	FBtr0099988	kay	CG33956	**1**	0	1	0	**0**	0	0	0
FBgn0004381	FBtr0076030	Klp68D	CG7293	**1**	0	1	0	**0**	0	0	0
FBgn0033984	FBtr0087407	Lap1	CG10255	**1**	0	1	0	**0**	0	0	0
FBgn0000464	FBtr0081260	Lar	CG10443	**1**	0	1	0	**0**	0	0	0

FBgn0014343	FBtr0075911	mirr	CG10601	**1**	0	1	0	**0**	0	0	0
FBgn0086705	FBtr0079268	mmy	CG9535	**1**	0	1	0	**0**	0	0	0
FBgn0053545	FBtr0091511	nab	CG33545	**1**	0	1	0	**0**	0	0	0
FBgn0029970	FBtr0071142	Nek2	CG17256	**1**	0	1	0	**0**	0	0	0
FBgn0015776	FBtr0079384	nrv1	CG9258	**1**	0	1	0	**0**	0	0	0
FBgn0085424	FBtr0112642	nub	CG34395	**1**	0	1	0	**0**	0	0	0
FBgn0020389	FBtr0074885	Paps	CG8363	**1**	0	1	0	**0**	0	0	0
FBgn0031681	FBtr0079026	pgant5	CG31651	**1**	0	1	0	**0**	0	0	0
FBgn0037364	FBtr0078727	Rab23	CG2108	**1**	0	1	0	**0**	0	0	0
FBgn0003205	FBtr0082122	Ras85D	CG9375	**1**	0	1	0	**0**	0	0	0
FBgn0003209	FBtr0079745	raw	CG12437	**1**	0	1	0	**0**	0	0	0
FBgn0030362	FBtr0073575	regucalcin	CG1803	**1**	0	1	0	**0**	0	0	0
FBgn0026369	FBtr0071641	Sara	CG15667	**1**	0	1	0	**0**	0	0	0
FBgn0015295	FBtr0087244	shark	CG18247	**1**	0	1	0	**0**	0	0	0
FBgn0052423	FBtr0077173	shep	CG32423	**1**	0	1	0	**0**	0	0	0
FBgn0038788	FBtr0083882	Sirt2	CG5085	**1**	0	1	0	**0**	0	0	0
FBgn0042630	FBtr0075747	Sox21b	CG32139	**1**	0	1	0	**0**	0	0	0
FBgn0044823	FBtr0078739	Spec2	CG14672	**1**	0	1	0	**0**	0	0	0
FBgn0015014	FBtr0082005	tgo	CG11987	**1**	0	1	0	**0**	0	0	0
FBgn0036285	FBtr0075980	toe	CG10704	**1**	0	1	0	**0**	0	0	0

FBgn0004924	FBtr0074005	Top1	CG6146	**1**	0	1	0	**0**	0	0	0
FBgn0037917	FBtr0082413	wkd	CG5344	**1**	0	1	0	**0**	0	0	0
FBgn0038420	FBtr0083324	CG10311	CG10311	**1**	0	1	0	**0**	0	0	0
FBgn0035490	FBtr0073203	CG1136	CG1136	**1**	0	1	0	**0**	0	0	0
FBgn0031232	FBtr0078146	CG11617	CG11617	**1**	0	1	0	**0**	0	0	0
FBgn0035429	FBtr0073052	CG12017	CG12017	**1**	0	1	0	**0**	0	0	0
FBgn0038304	FBtr0083051	CG12241	CG12241	**1**	0	1	0	**0**	0	0	0
FBgn0033477	FBtr0088433	CG12918	CG12918	**1**	0	1	0	**0**	0	0	0
FBgn0036593	FBtr0075475	CG13048	CG13048	**1**	0	1	0	**0**	0	0	0
FBgn0035643	FBtr0077027	CG13287	CG13287	**1**	0	1	0	**0**	0	0	0
FBgn0034809	FBtr0071987	CG13542	CG13542	**1**	0	1	0	**0**	0	0	0
FBgn0039161	FBtr0084553	CG13606	CG13606	**1**	0	1	0	**0**	0	0	0
FBgn0033363	FBtr0088666	CG13744	CG13744	**1**	0	1	0	**0**	0	0	0
FBgn0035270	FBtr0072792	CG13933	CG13933	**1**	0	1	0	**0**	0	0	0
FBgn0033212	FBtr0088939	CG1399	CG1399	**1**	0	1	0	**0**	0	0	0
FBgn0031349	FBtr0077875	CG14351	CG14351	**1**	0	1	0	**0**	0	0	0
FBgn0038162	FBtr0082859	CG14368	CG14368	**1**	0	1	0	**0**	0	0	0
FBgn0029880	FBtr0070949	CG14443	CG14443	**1**	0	1	0	**0**	0	0	0
FBgn0038341	FBtr0083160	CG14869	CG14869	**1**	0	1	0	**0**	0	0	0
FBgn0035543	FBtr0073298	CG15020	CG15020	**1**	0	1	0	**0**	0	0	0

FBgn0035795	FBtr0076829	CG16998	CG16998	**1**	0	1	0	**0**	0	0	0
FBgn0037884	FBtr0082363	CG17184	CG17184	**1**	0	1	0	**0**	0	0	0
FBgn0038829	FBtr0083993	CG17271	CG17271	**1**	0	1	0	**0**	0	0	0
FBgn0038462	FBtr0083375	CG17556	CG17556	**1**	0	1	0	**0**	0	0	0
FBgn0027597	FBtr0077903	CG17712	CG17712	**1**	0	1	0	**0**	0	0	0
FBgn0034378	FBtr0086606	CG18604	CG18604	**1**	0	1	0	**0**	0	0	0
FBgn0050083	FBtr0087318	CG30083	CG30083	**1**	0	1	0	**0**	0	0	0
FBgn0050360	FBtr0088750	CG30360	CG30360	**1**	0	1	0	**0**	0	0	0
FBgn0051048	FBtr0085327	CG31048	CG31048	**1**	0	1	0	**0**	0	0	0
FBgn0051140	FBtr0084497	CG31140	CG31140	**1**	0	1	0	**0**	0	0	0
FBgn0052062	FBtr0076289	CG32062	CG32062	**1**	0	1	0	**0**	0	0	0
FBgn0052105	FBtr0075953	CG32105	CG32105	**1**	0	1	0	**0**	0	0	0
FBgn0052119	FBtr0075867	CG32119	CG32119	**1**	0	1	0	**0**	0	0	0
FBgn0052195	FBtr0075125	CG32195	CG32195	**1**	0	1	0	**0**	0	0	0
FBgn0052364	FBtr0076720	CG32364	CG32364	**1**	0	1	0	**0**	0	0	0
FBgn0052767	FBtr0070760	CG32767	CG32767	**1**	0	1	0	**0**	0	0	0
FBgn0053279	FBtr0076858	CG33279	CG33279	**1**	0	1	0	**0**	0	0	0
FBgn0053543	FBtr0091509	CG33543	CG33543	**1**	0	1	0	**0**	0	0	0
FBgn0034997	FBtr0072265	CG3376	CG3376	**1**	0	1	0	**0**	0	0	0
FBgn0036008	FBtr0076426	CG3408	CG3408	**1**	0	1	0	**0**	0	0	0

FBgn0083981	FBtr0110992	CG34145	CG34145	**1**	0	1	0	**0**	0	0	0
FBgn0034854	FBtr0072036	CG3493	CG3493	**1**	0	1	0	**0**	0	0	0
FBgn0038463	FBtr0083370	CG3534	CG3534	**1**	0	1	0	**0**	0	0	0
FBgn0038461	FBtr0083376	CG3678	CG3678	**1**	0	1	0	**0**	0	0	0
FBgn0031304	FBtr0077956	CG4552	CG4552	**1**	0	1	0	**0**	0	0	0
FBgn0031306	FBtr0077957	CG4577	CG4577	**1**	0	1	0	**0**	0	0	0
FBgn0030778	FBtr0100450	CG4678	CG4678	**1**	0	1	0	**0**	0	0	0
FBgn0034304	FBtr0086780	CG5742	CG5742	**1**	0	1	0	**0**	0	0	0
FBgn0033853	FBtr0087685	CG6145	CG6145	**1**	0	1	0	**0**	0	0	0
FBgn0036533	FBtr0075542	CG6151	CG6151	**1**	0	1	0	**0**	0	0	0
FBgn0036819	FBtr0075062	CG6856	CG6856	**1**	0	1	0	**0**	0	0	0
FBgn0037074	FBtr0078404	CG7324	CG7324	**1**	0	1	0	**0**	0	0	0
FBgn0039704	FBtr0085485	CG7802	CG7802	**1**	0	1	0	**0**	0	0	0
FBgn0034057	FBtr0087268	CG8314	CG8314	**1**	0	1	0	**0**	0	0	0
FBgn0036424	FBtr0075727	CG9598	CG9598	**1**	0	1	0	**0**	0	0	0
FBgn0000028	FBtr0074015	acj6	CG9151	**1**	0	0	1	**0**	0	0	0
FBgn0000097	FBtr0077850	aop	CG3166	**1**	0	0	1	**0**	0	0	0
FBgn0025186	FBtr0071774	ari-2	CG5709	**1**	0	0	1	**0**	0	0	0
FBgn0000117	FBtr0089988	arm	CG11579	**1**	0	0	1	**0**	0	0	0
FBgn0011744	FBtr0076737	Arp66B	CG7558	**1**	0	0	1	**0**	0	0	0

FBgn0031298	FBtr0077953	Atg4	CG4428	**1**	0	0	1	**0**	0	0	0
FBgn0026597	FBtr0085553	Axn	CG7926	**1**	0	0	1	**0**	0	0	0
FBgn0039447	FBtr0085050	beat-VII	CG14249	**1**	0	0	1	**0**	0	0	0
FBgn0025724	FBtr0080469	beta'Cop	CG6699	**1**	0	0	1	**0**	0	0	0
FBgn0013759	FBtr0084161	Caki	CG6703	**1**	0	0	1	**0**	0	0	0
FBgn0015919	FBtr0075909	caup	CG10605	**1**	0	0	1	**0**	0	0	0
FBgn0029092	FBtr0088548	ced-6	CG11804	**1**	0	0	1	**0**	0	0	0
FBgn0000303	FBtr0089367	Cha	CG12345	**1**	0	0	1	**0**	0	0	0
FBgn0000317	FBtr0080723	ck	CG7595	**1**	0	0	1	**0**	0	0	0
FBgn0050045	FBtr0113359	Cpr49Aa	CG30045	**1**	0	0	1	**0**	0	0	0
FBgn0031432	FBtr0077749	Cyp309a1	CG9964	**1**	0	0	1	**0**	0	0	0
FBgn0023388	FBtr0081503	Dap160	CG1099	**1**	0	0	1	**0**	0	0	0
FBgn0027594	FBtr0072798	drpr	CG2086	**1**	0	0	1	**0**	0	0	0
FBgn0000541	FBtr0072523	E(bx)	CG32346	**1**	0	0	1	**0**	0	0	0
FBgn0000570	FBtr0070091	elav	CG4262	**1**	0	0	1	**0**	0	0	0
FBgn0050147	FBtr0086255	Hil	CG30147	**1**	0	0	1	**0**	0	0	0
FBgn0000448	FBtr0088366	Hr46	CG33183	**1**	0	0	1	**0**	0	0	0
FBgn0031305	FBtr0077986	Iris	CG4715	**1**	0	0	1	**0**	0	0	0
FBgn0002283	FBtr0075311	l(3)73Ah	CG4195	**1**	0	0	1	**0**	0	0	0
FBgn0041203	FBtr0073659	LIMK1	CG1848	**1**	0	0	1	**0**	0	0	0

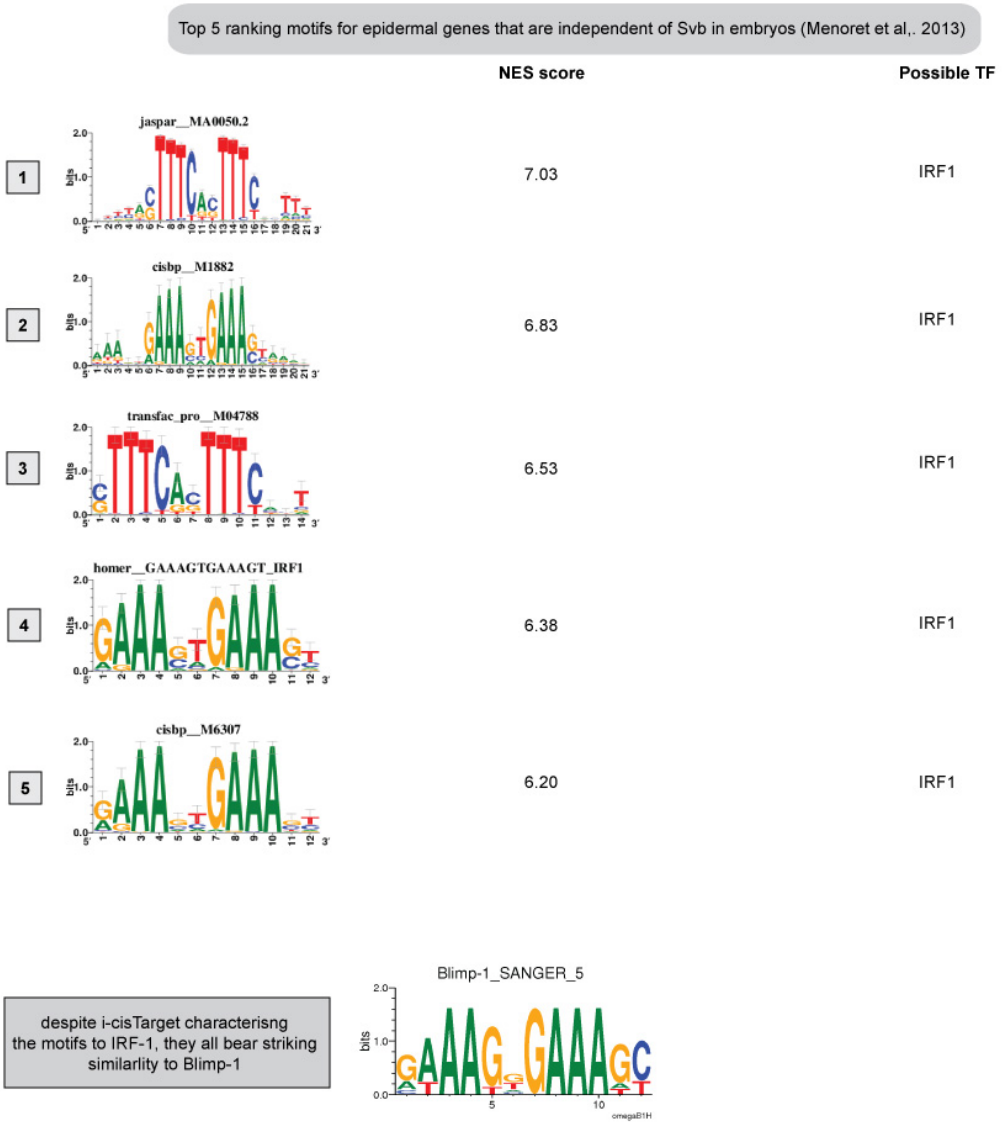
FBgn0046704	FBtr0079300	Liprin-alpha	CG11199	**1**	0	0	1	**0**	0	0	0
FBgn0028582	FBtr0076794	lqf	CG8532	**1**	0	0	1	**0**	0	0	0
FBgn0000037	FBtr0072367	mAcR-60C	CG4356	**1**	0	0	1	**0**	0	0	0
FBgn0034521	FBtr0086259	Mgat1	CG13431	**1**	0	0	1	**0**	0	0	0
FBgn0038818	FBtr0083936	Nep4	CG4058	**1**	0	0	1	**0**	0	0	0
FBgn0004102	FBtr0089650	oc	CG12154	**1**	0	0	1	**0**	0	0	0
FBgn0028996	FBtr0089104	onecut	CG1922	**1**	0	0	1	**0**	0	0	0
FBgn0025360	FBtr0088865	Optix	CG18455	**1**	0	0	1	**0**	0	0	0
FBgn0037428	FBtr0078611	Osi18	CG1169	**1**	0	0	1	**0**	0	0	0
FBgn0037415	FBtr0078598	Osi8	CG15591	**1**	0	0	1	**0**	0	0	0
FBgn0003016	FBtr0114465	osp	CG3479	**1**	0	0	1	**0**	0	0	0
FBgn0044826	FBtr0083312	Pak3	CG14895	**1**	0	0	1	**0**	0	0	0
FBgn0038237	FBtr0082985	Pde6	CG8279	**1**	0	0	1	**0**	0	0	0
FBgn0017558	FBtr0112892	Pdk	CG8808	**1**	0	0	1	**0**	0	0	0
FBgn0001970	FBtr0080629	Pgant35A	CG7480	**1**	0	0	1	**0**	0	0	0
FBgn0011826	FBtr0074294	Pp2B-14D	CG9842	**1**	0	0	1	**0**	0	0	0
FBgn0003149	FBtr0076593	Prm	CG5939	**1**	0	0	1	**0**	0	0	0
FBgn0040752	FBtr0087601	Prosap	CG30483	**1**	0	0	1	**0**	0	0	0
FBgn0003254	FBtr0086512	rib	CG7230	**1**	0	0	1	**0**	0	0	0
FBgn0015778	FBtr0082796	rin	CG9412	**1**	0	0	1	**0**	0	0	0

FBgn0003339	FBtr0081657	Scr	CG1030	**1**	0	0	1	**0**	0	0	0
FBgn0041094	FBtr0076168	scyl	CG7590	**1**	0	0	1	**0**	0	0	0
FBgn0010851	FBtr0076938	sgl	CG10072	**1**	0	0	1	**0**	0	0	0
FBgn0029761	FBtr0070766	SK	CG10706	**1**	0	0	1	**0**	0	0	0
FBgn0023423	FBtr0084032	slmb	CG3412	**1**	0	0	1	**0**	0	0	0
FBgn0003507	FBtr0083215	srp	CG3992	**1**	0	0	1	**0**	0	0	0
FBgn0033876	FBtr0087645	synaptogyrin	CG10808	**1**	0	0	1	**0**	0	0	0
FBgn0010280	FBtr0075448	Taf4	CG5444	**1**	0	0	1	**0**	0	0	0
FBgn0031030	FBtr0074773	Tao-1	CG14217	**1**	0	0	1	**0**	0	0	0
FBgn0040697	FBtr0073257	Teh3	CG18676	**1**	0	0	1	**0**	0	0	0
FBgn0003715	FBtr0072456	Tkr	CG16778	**1**	0	0	1	**0**	0	0	0
FBgn0003717	FBtr0085059	Tl	CG5490	**1**	0	0	1	**0**	0	0	0
FBgn0003721	FBtr0089960	Tm1	CG4898	**1**	0	0	1	**0**	0	0	0
FBgn0029508	FBtr0086172	Tsp42Ea	CG18817	**1**	0	0	1	**0**	0	0	0
FBgn0004436	FBtr0078849	UbcD6	CG2013	**1**	0	0	1	**0**	0	0	0
FBgn0005671	FBtr0082670	Vha55	CG17369	**1**	0	0	1	**0**	0	0	0
FBgn0016076	FBtr0079070	vri	CG14029	**1**	0	0	1	**0**	0	0	0
FBgn0034876	FBtr0072114	wmd	CG3957	**1**	0	0	1	**0**	0	0	0
FBgn0039104	FBtr0084475	CG10252	CG10252	**1**	0	0	1	**0**	0	0	0
FBgn0031395	FBtr0077848	CG10874	CG10874	**1**	0	0	1	**0**	0	0	0

FBgn0036334	FBtr0075875	CG11267	CG11267	**1**	0	0	1	**0**	0	0	0
FBgn0024362	FBtr0070208	CG11412	CG11412	**1**	0	0	1	**0**	0	0	0
FBgn0037585	FBtr0081852	CG11718	CG11718	**1**	0	0	1	**0**	0	0	0
FBgn0033958	FBtr0087476	CG12858	CG12858	**1**	0	0	1	**0**	0	0	0
FBgn0036676	FBtr0075261	CG13028	CG13028	**1**	0	0	1	**0**	0	0	0
FBgn0035699	FBtr0077007	CG13300	CG13300	**1**	0	0	1	**0**	0	0	0
FBgn0035246	FBtr0072822	CG13928	CG13928	**1**	0	0	1	**0**	0	0	0
FBgn0038658	FBtr0083725	CG14292	CG14292	**1**	0	0	1	**0**	0	0	0
FBgn0038124	FBtr0082733	CG14380	CG14380	**1**	0	0	1	**0**	0	0	0
FBgn0037449	FBtr0081675	CG15185	CG15185	**1**	0	0	1	**0**	0	0	0
FBgn0028844	FBtr0080632	CG15283	CG15283	**1**	0	0	1	**0**	0	0	0
FBgn0037698	FBtr0082125	CG16779	CG16779	**1**	0	0	1	**0**	0	0	0
FBgn0036267	FBtr0076015	CG17144	CG17144	**1**	0	0	1	**0**	0	0	0
FBgn0036958	FBtr0074850	CG17233	CG17233	**1**	0	0	1	**0**	0	0	0
FBgn0039887	FBtr0085873	CG2053	CG2053	**1**	0	0	1	**0**	0	0	0
FBgn0035383	FBtr0072978	CG2107	CG2107	**1**	0	0	1	**0**	0	0	0
FBgn0051342	FBtr0113404	CG31342	CG31342	**1**	0	0	1	**0**	0	0	0
FBgn0051522	FBtr0078939	CG31522	CG31522	**1**	0	0	1	**0**	0	0	0
FBgn0051523	FBtr0078935	CG31523	CG31523	**1**	0	0	1	**0**	0	0	0
FBgn0051687	FBtr0081361	CG31687	CG31687	**1**	0	0	1	**0**	0	0	0

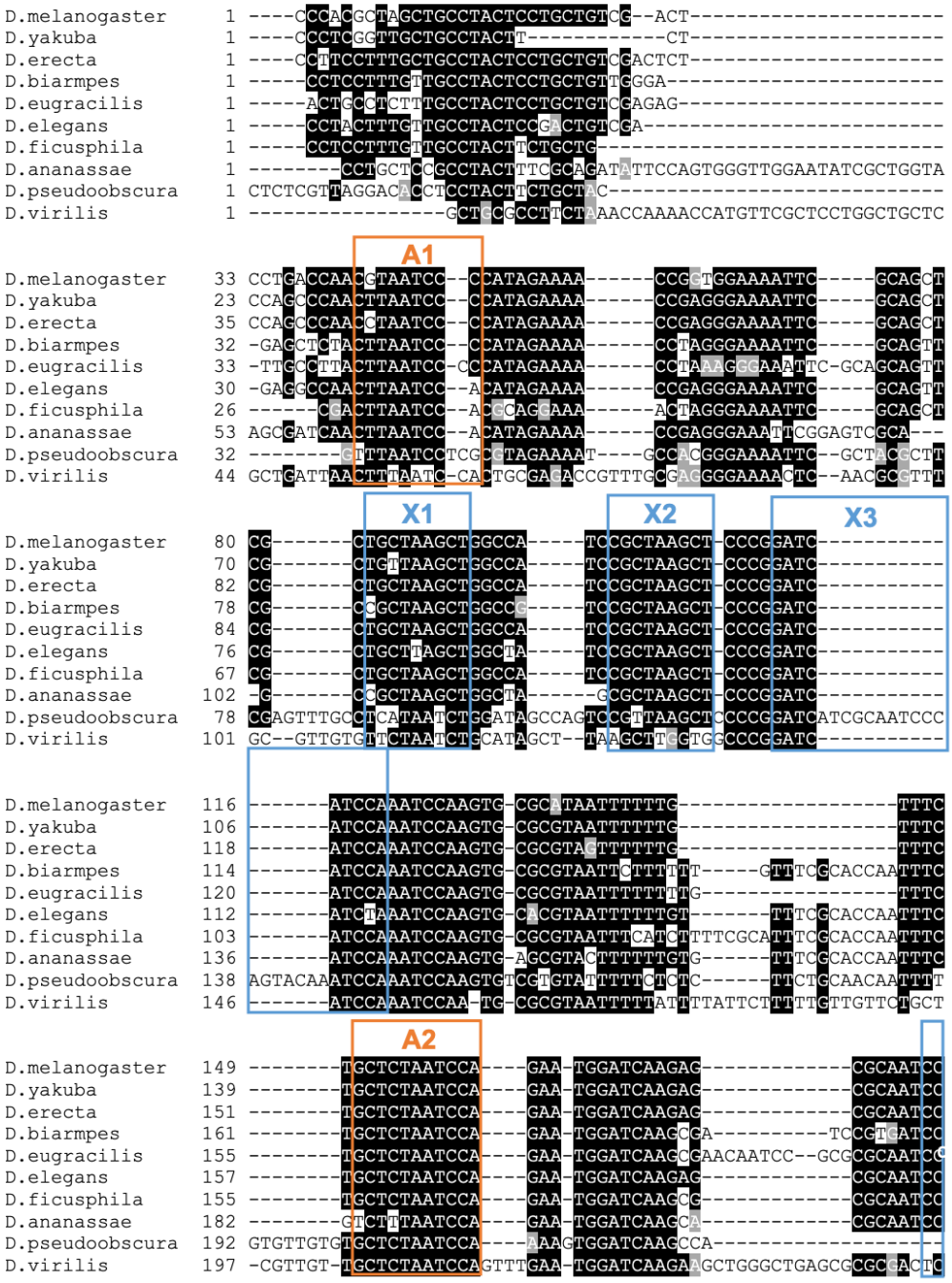
FBgn0051700	FBtr0085933	CG31700	CG31700	**1**	0	0	1	**0**	0	0	0
FBgn0051869	FBtr0080189	CG31869	CG31869	**1**	0	0	1	**0**	0	0	0
FBgn0052085	FBtr0076133	CG32085	CG32085	**1**	0	0	1	**0**	0	0	0
FBgn0052138	FBtr0114573	CG32138	CG32138	**1**	0	0	1	**0**	0	0	0
FBgn0047135	FBtr0073106	CG32276	CG32276	**1**	0	0	1	**0**	0	0	0
FBgn0052574	FBtr0074329	CG32574	CG32574	**1**	0	0	1	**0**	0	0	0
FBgn0052651	FBtr0073672	CG32651	CG32651	**1**	0	0	1	**0**	0	0	0
FBgn0053174	FBtr0073825	CG33174	CG33174	**1**	0	0	1	**0**	0	0	0
FBgn0033094	FBtr0086141	CG3403	CG3403	**1**	0	0	1	**0**	0	0	0
FBgn0054030	FBtr0100085	CG34030	CG34030	**1**	0	0	1	**0**	0	0	0
FBgn0033095	FBtr0086140	CG3409	CG3409	**1**	0	0	1	**0**	0	0	0
FBgn0085196	FBtr0112358	CG34167	CG34167	**1**	0	0	1	**0**	0	0	0
FBgn0085198	FBtr0112360	CG34169	CG34169	**1**	0	0	1	**0**	0	0	0
FBgn0031257	FBtr0078061	CG4133	CG4133	**1**	0	0	1	**0**	0	0	0
FBgn0030745	FBtr0074268	CG4239	CG4239	**1**	0	0	1	**0**	0	0	0
FBgn0036623	FBtr0075417	CG4729	CG4729	**1**	0	0	1	**0**	0	0	0
FBgn0039352	FBtr0084898	CG5053	CG5053	**1**	0	0	1	**0**	0	0	0
FBgn0034873	FBtr0072058	CG5360	CG5360	**1**	0	0	1	**0**	0	0	0
FBgn0036556	FBtr0075521	CG5830	CG5830	**1**	0	0	1	**0**	0	0	0
FBgn0046247	FBtr0085160	CG5938	CG5938	**1**	0	0	1	**0**	0	0	0

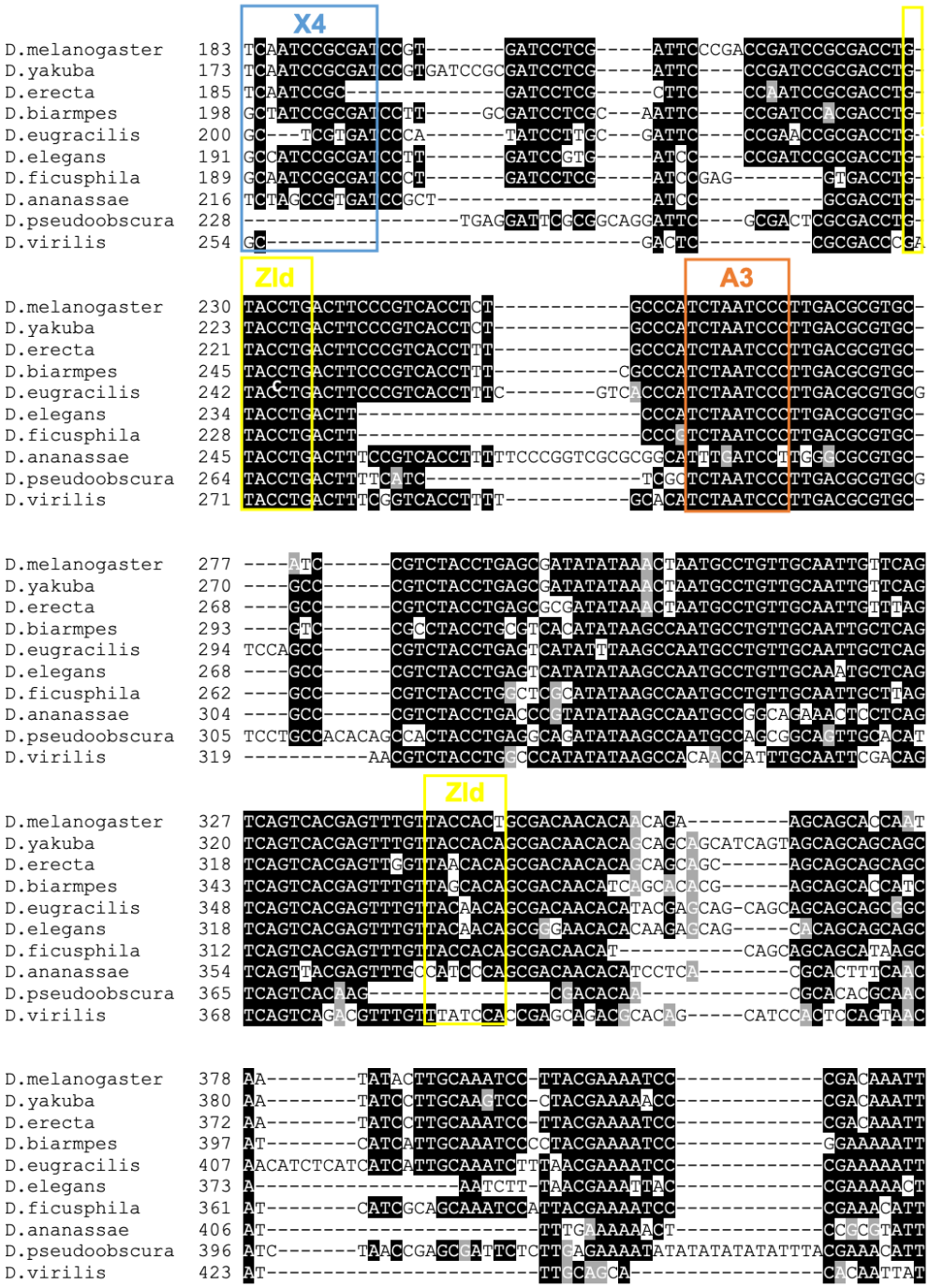
FBgn0036202	FBtr0113165	CG6024	CG6024	**1**	0	0	1	**0**	0	0	0
FBgn0038339	FBtr0113239	CG6118	CG6118	**1**	0	0	1	**0**	0	0	0
FBgn0036154	FBtr0076182	CG6168	CG6168	**1**	0	0	1	**0**	0	0	0
FBgn0032340	FBtr0080231	CG6181	CG6181	**1**	0	0	1	**0**	0	0	0
FBgn0037807	FBtr0082240	CG6293	CG6293	**1**	0	0	1	**0**	0	0	0
FBgn0034685	FBtr0071750	CG6393	CG6393	**1**	0	0	1	**0**	0	0	0
FBgn0039187	FBtr0084608	CG6454	CG6454	**1**	0	0	1	**0**	0	0	0
FBgn0032388	FBtr0089572	CG6686	CG6686	**1**	0	0	1	**0**	0	0	0
FBgn0036800	FBtr0075108	CG6897	CG6897	**1**	0	0	1	**0**	0	0	0
FBgn0028541	FBtr0080532	CG7364	CG7364	**1**	0	0	1	**0**	0	0	0
FBgn0035815	FBtr0110885	CG7422	CG7422	**1**	0	0	1	**0**	0	0	0
FBgn0038608	FBtr0083586	CG7670	CG7670	**1**	0	0	1	**0**	0	0	0
FBgn0038122	FBtr0082732	CG8138	CG8138	**1**	0	0	1	**0**	0	0	0
FBgn0031478	FBtr0077629	CG8814	CG8814	**1**	0	0	1	**0**	0	0	0
FBgn0028500	FBtr0078512	CG9063	CG9063	**1**	0	0	1	**0**	0	0	0
FBgn0030617	FBtr0074026	CG9095	CG9095	**1**	0	0	1	**0**	0	0	0
FBgn0034810	FBtr0071986	CG9895	CG9895	**1**	0	0	1	**0**	0	0	0

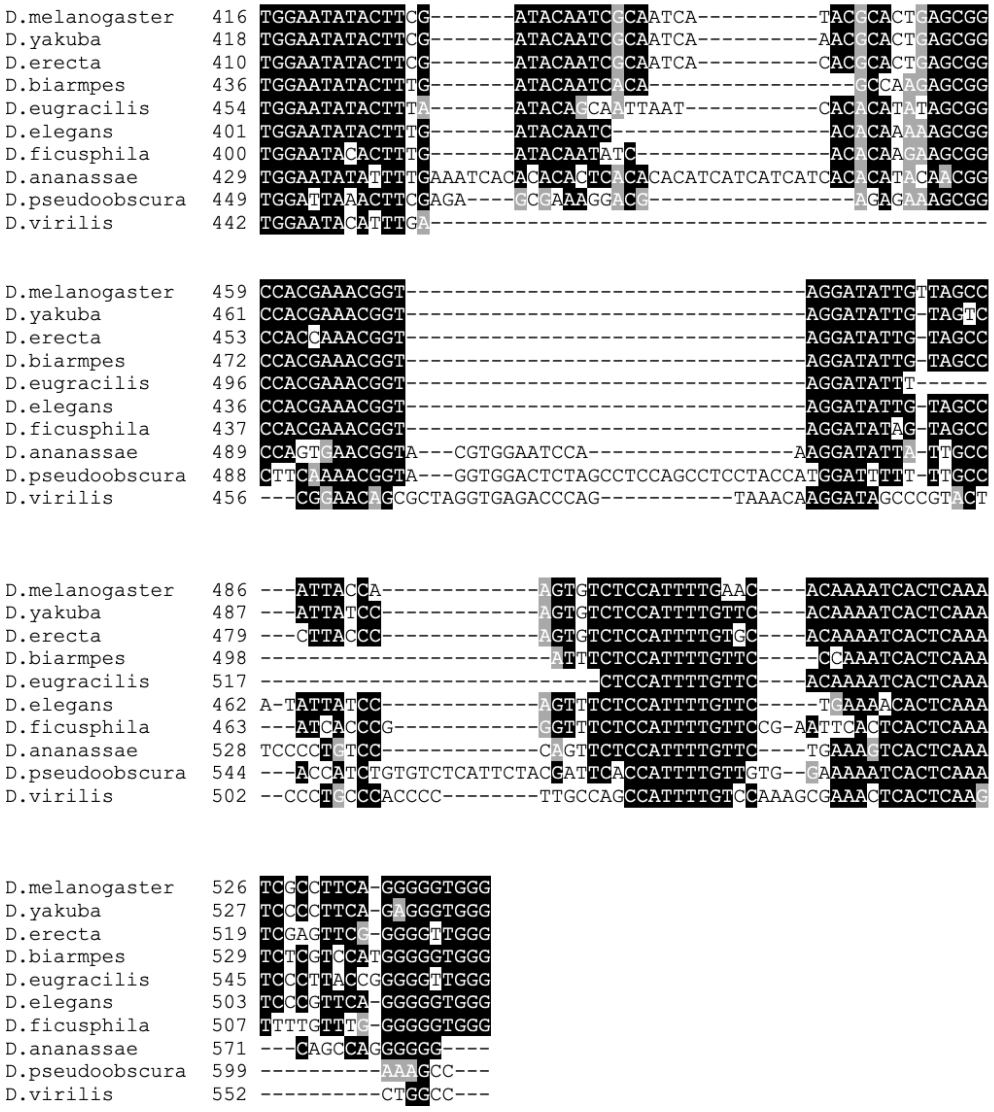


Supplementary Fig. 3. Results from i-cisTarget for motif enrichment in epidermal genes that are not dependent on Svb. According to i-cisTarget the top 5 motifs all correspond to IRF genes, however all the motifs also bear resemblance to the Blimp-1 motif (shown below).

Chapter 4




X4
Zld
A3
Zld



Supplementary Fig. 4. Multiple species alignment of the hb P2 enhancer for several *Drosophila* species. Binding sites are annotated on the alignment with high affinity sites shown in orange, low affinity sites shown in blue and Zelda sites shown in yellow. The alignment reveals that as well as turnover in the Bcd binding sites themselves, there has been extensive turnover in the inter-nucleotide spacing between these sites. On the other hand, there is also the present of conserved blocks of DNA that are not validated TF sites for Bcd or any other TF.